

# Persona-based Assessment of Software Engineering Student Research Projects: An Experience Report

Chetan Arora\*<sup>†</sup>, Laura Tubino\*, Andrew Cain\*, Kevin Lee\*, Vasudha Malhotra\*

\*Deakin University, Geelong, Australia

<sup>†</sup>Monash University, Victoria, Australia

Email: chetan.arora@monash.edu, {laura.t, andrew.cain, kevin.lee, v.malhotra}@deakin.edu.au

**Abstract**—Students enrolled in software engineering degrees are generally required to undertake a research project in their final year through which they demonstrate the ability to conduct research, communicate outcomes, and build in-depth expertise in an area. Assessment in these projects typically involves evaluating the product of their research via a thesis or a similar artifact. However, this misses a range of other factors that go into producing successful software engineers and researchers. Incorporating aspects such as process, attitudes, project complexity, and supervision support into the assessment can provide a more holistic evaluation of the performance likely to better align with the intended learning outcomes. In this paper, we present on our experience of adopting an innovative assessment approach to enhance learning outcomes and research performance in our software engineering research projects. Our approach adopted a task-oriented approach to portfolio assessment that incorporates student personas, frequent formative feedback, delayed summative grading, and standards-aligned outcomes-based assessment. We report upon our continuous improvement journey in adapting tasks and criteria to address the challenges of assessing student research projects. Our lessons learnt demonstrate the value of personas to guide the development of holistic rubrics, giving meaning to grades and focusing staff and student attention on attitudes and skills rather than a product only.

**Index Terms**—Software Engineering Education, Research Projects, Portfolio-based Assessment

## I. INTRODUCTION

Providing students with research experiences improves critical thinking and analytical skills, develops their self-confidence, allows them to feel part of the software engineering community [1], and strengthens their employability profile [2]. Therefore, Software Engineering (SE) degrees commonly include a research component which results in a thesis [3], [4]. Including research projects in SE degrees, however, has certain challenges which need resolution. Providing students with an authentic and equitable experience is difficult due to the variability of projects and the quality of supervisors. Students also present high variability in skills and understanding, with some being able to contribute intellectually to a research project idea and design and others needing clear instructions every step of the way. In addition, SE research projects commonly involve a practical problem in need of a solution. Doing high-quality empirical work within the time frame of a typical coursework research project is difficult [5], and therefore requires a balance between academic values and practical relevance [3]. These factors are among those that

make assessing students' outcomes of research projects such a complex endeavor [6].

The implications of research project design and assessment not accurately reflecting the skills and achievements of students in these research projects are many, including student disengagement, attrition, and complaints. In addition, in our institution, the grade a student achieves in these research projects is considered when awarding PhD scholarships, making it a high-stakes issue. Along similar lines, the assessment in SE research projects reflects the ability of students to apply their overall learning in the SE degree in an ideally self-managed project. This has implications for the employment prospects of SE graduates and the reputational outcomes for the institution.

To address these issues, we redesigned the assessment in SE research projects, wherein students work on a SE research topic with one or more supervisors, in a one-to-one supervision model. We worked on shifting the assessment focus, from the final thesis to the research process, through the submission of small regular tasks resulting in a thesis and a project management report. Scaffolding the process also ensures every student gets formative feedback throughout the teaching period. This process-oriented model is enhanced by SE research student personas that highlight to students the expectations at different achievement levels, and to research project supervisors the scaffolding needed by students at different levels. This model aims to address the variability of projects, students, and supervisors, resulting in a more accurate evaluation of a student's skills and achievements.

In this paper, we describe and reflect on how we applied a design thinking approach [7], [8] to research project assessment design. Each iteration is described, emphasising the evaluation and lessons learned. The main contributions of this paper are (i) using a design thinking approach to redesigning SE research projects' assessment and (ii) exemplar assessment rubric and persona model, and our experiences deploying this exemplar assessment rubric and model; and (iii) our reflections and lessons learnt for developing a formative assessment model for improving the learning outcomes and research quality in SE research projects.

**Structure.** Section II provides context on Software Engineering research projects and describes our experiences. Section III describes our assessment redesign approach using design thinking. Sections IV, V and VI present the redesign of the

assessment model over three iterations. Section VII discusses the lessons learnt and our reflections on the three iterations. Section VIII addresses related work. Finally, Section IX presents conclusions and future work.

## II. CONTEXT

In this section, we provide context to our university's teaching and learning environment, SE research projects, and motivate the case for redevelopment of SE research projects based on the issues in the existing model.

### A. Bachelor of Software Engineering

Bachelor of Software Engineering (BSE) students complete a four-year degree, which prepares them for a career leading software development projects. In the first year, they complete technical units in traditional computer science, programming, software engineering, and project management. We note that a course, e.g., Java Programming 101, is termed as *unit* in our context. In their second year, this foundation is expanded with more specialized SE skills such as AI and full-stack development. To utilize and provide evidence of their learning, students complete a technical capstone (industry-problem-based development team project) program in year three, as described in [9]. To develop and demonstrate an ability to use their skills for advanced research and development projects and contribute to the wider SE community, in their final year, students complete a SE research project unit (*SIT723*).

### B. SE Research Project (*SIT723*) Unit

As mentioned above, BSE students complete *SIT723* in their final year. In *SIT723*, students work with one (or more) academic supervisors to undertake an ideally self-driven research project. This project is generally in an area of interest to the student and focused on demonstrating their learning outcomes of focused technical areas of SE research. The unit aims to help them develop research skills necessary to complete a SE research project, and be familiar with SE research process. The unit runs over all semesters at our university, with  $\approx 100$  students enrolled in it each semester. The unit includes research training in the form of workshops (e.g., literature search and review, academic writing and research project management) for all students by the unit coordination team, while in parallel individual students work with the academic supervisors on their specific research projects.

**Project Allocation.** The allocation of students to the research projects begins with compiling a handbook of all research projects on offer for BSE students. All staff members in the department of Information Technology (IT) at our university propose research projects (collated in the handbook) for *SIT723* students. The projects in the handbook range from exploratory (blue-sky) research to scoped projects, industry-style R&D projects, or research projects with industry partners. The topics of the projects are diverse, such as requirements engineering (RE), applied SE in blockchain, cybersecurity, Internet of Things (IoT), Artificial Intelligence (AI) and other

domains. Each project's description would describe the overall scope of the project and the background/interest required to carry out the project, e.g., a project on "RE for AI systems" would require prior experience in addition to RE in machine learning (ML) and an interest in RE for AI systems [10], [11].

The students enrolled in *SIT723* provide their top- $N$  preferences from the handbook. The students are subsequently matched with the research projects based on their preferences and their background in the research project topics. The allocation process also takes students' overall average grade into consideration to avoid 'over-booking' specific projects. For example, if students  $S_x$  and  $S_y$  are interested in the same project and have matching background, the one with higher average grade will be allocated to the project, assuming there is another matching project available for the other student.

**Assessment Process.** At our university, the students outcomes are graded on a five level scale - below 50% Fail ( $F$ ), 50%-59% Pass ( $P$ ), 60%-69% Credit ( $C$ ), 70%-79% Distinction ( $D$ ), and above 80% High-Distinction ( $HD$ ). The actual numerical mark is calculated based on the quality of the assessed artefact. In BSE units, this has traditionally meant 2-3 main assessment artefacts, and possible examination, which are all assessed with defined marking schemes.

In the past, in *SIT723*, the assessment was isolated to an academic's (research project supervisor's) review of the final thesis submitted by a student, based on a coarse marking grid. The marking scheme for *SIT723* contained criteria for different sections of the submitted thesis, e.g., *overview, related work, research questions, methodology, evaluation; findings, interpretation, contributions, future work, references* and the overall *thesis presentation*. The final mark for each criterion, based on the quality produced, was awarded on a scale of 1-10, and totaling to 100. The final grade was then calculated based on the scale mentioned above.

**Assessment Issues in *SIT723*.** The assessment model discussed above had been used in *SIT723* for a long time and until the end of year 2020. The influence of assessment in *SIT723* is not limited to only a final grade of the students, there is a far-reaching impact of the grade. The final grade of *SIT723* has a significant weighting in the PhD scholarship allocation process at our university, which is highly competitive. Therefore, a recurring phenomenon with *SIT723* is supervisors tendency to inflate the grades [12]. There were a number of other issues in the existing assessment model of *SIT723*. The model unintentionally rewarded straightforward projects, to the detriment of more complex or exploratory ones. There were several relevant factors that were not accounted for, e.g., the variety of other documentation required for completing a research project, such as, work plans, work diaries and presentations, and the student's learning experience and skill development, such as, the extent of support provided to students, student's initiative in the project, and reflections on their learning [13]. We were motivated for improving the overall quality of the *SIT723* students and BSE graduates, and addressing issues related to *SIT723* including the ones mentioned above.

### III. REDESIGNING ASSESSMENT FOR THE RESEARCH PROJECT UNITS USING A DESIGN THINKING APPROACH

In this section, we describe how we applied the design thinking approach in the evolution of the SE research project unit (*SIT723*) at our university. We applied the five stages of the design thinking process [7], [8] to improve the *SIT723* design iteratively. Fig. 1 (a) shows an overview of the design thinking process. The first step **Empathize** focuses on understanding the key stakeholders and their concerns in the problem context. The next step is to **Define**, i.e., fully establish the stakeholders' concerns, synthesize these concerns and requirements, and correctly frame the problem at hand. The third step of **Ideate** focuses on brainstorming alternate solutions constructs, followed by the **Prototyping** of one of these alternate solutions. The last step is **Evaluation**, where the prototype solutions are evaluated against the stakeholder requirements and concerns identified in the Empathize and the Design steps. We instantiated the steps iteratively (see Fig. 1 (b)) for redesigning *SIT723*. Below we report on all the steps in a sequence and the iterative improvement of the unit while redesigning, and the lessons learnt.

#### A. Empathize Step

In the first step (empathize), we focused on identifying the key issues with the stakeholders involved in *SIT723* research projects. The key stakeholders are (i) Bachelor of Software Engineering (BSE) students, (ii) project supervisors, (iii) the unit coordination team, and (iv) the BSE management team. We identified the issues by reviewing the current unit content and structure (at the time), and having conversations with relevant stakeholders. There were numerous key issues (KIs) highlighted at this stage.

**KI1 - Assessment focus.** The existing assessment model's analysis helped us identify the excessive focus on the end-product, i.e., "research thesis", in contrast to improving students' research skills. The assessment was conducted at the end of the semester, and the thesis was reviewed by the supervisors for a final grade. From a student's perspective, this gave them little opportunity to hone their research skills, with the research thesis being the main and the only assessed artefact. This also prevented students from getting intermediary feedback during their research projects.

**KI2 - Varied level of guidance required by students.** During our conversations with the supervisors, we determined that the level of support is a key variable factor in *SIT723*. While some students are able to grasp the higher-level requirements of research projects and work independently, others expect a clearly laid out plan and are not able to have an abstract view of the research project as a whole. The overall unit structure, especially the assessment model, did not capture this factor to allow students to work at different levels of agency [14].

**KI3 - Assessment equitability.** SE is a broad field of research with numerous sub-disciplines and applications. The diversity

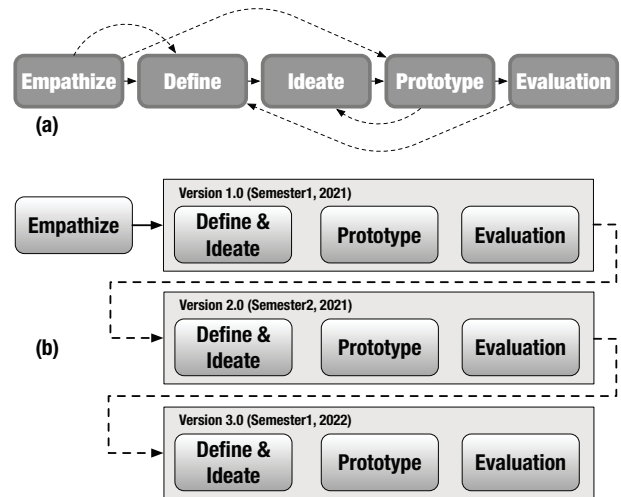


Fig. 1. (a) Design-thinking Approach [7]; (b) *SIT723* Redesign Process and Timeline based on the Design-thinking Approach.

in the research projects and the lack of a common understanding of expected learning outcomes resulted in inequitable assessment. The issue was further exacerbated by the vested interest of supervisors in higher grades for students. For instance, in most editions of *SIT723* prior to redesign, the fraction of students getting a  $\mathcal{H}\mathcal{D}$  was more than 60%. As noted in Section II, the final mark in *SIT723* had high influence on the allocation of otherwise highly-competitive PhD scholarships. This is a common issue in assessment fairness [12], and we identified clear patterns of 'inflated grades' in the unit. From BSE management team's perspective, this was a reputational concern, as the students with high grades in *SIT723* were being considered research ready and were entering SE PhD programme, when in reality they were not.

### IV. *SIT723* REDESIGN VERSION 1.0

Below, we present the first version (1.0) *SIT723*, redesigned based on the KIs identified above.

#### A. Define and Ideate Steps

In the design step, we analyzed all the information collated from different stakeholders and the unit material, e.g., the unit assessment model, and grade distributions in the past. We further discussed ideas on "how might we" address the issues identified in the empathize step [15]. For the ideate part of the process, we discussed several alternative solutions to address each of the key issues established in the empathize step. Next, we discuss the results of our design and ideate steps.

**Address KI1.** In response to KI1, we discussed numerous research skills that we deemed necessary for SE research. We had further discussions on structuring the unit assessment, and alleviating the issues of summative assessment at the end of the semester, based solely on the research thesis. It became clear that a more holistic approach to the management

TABLE I  
ROBERT KEGAN'S DEVELOPMENTAL STAGES

Developmental stage	Characteristics
<b>Instrumental mind</b>	Students commonly make decisions based on the benefits or consequences that may result from these, thus needing authority figures or rules to direct them.
<b>Socialized mind</b>	Students are "good citizens", taking others into account and considering long term consequences. They understand others' points of view, even if different from their own.
<b>Self-authoring mind</b>	Students can evaluate arguments (their own and others') using evidence, take contextual factors into account, acknowledge multiple ways of framing arguments, and reevaluate the basis of a decision as evidence changes.

and assessment in *SIT723* was required. One proposition for addressing this issue was to use Task-Oriented Portfolio Assessment (TOPA) [16]–[18]. TOPA is a process-oriented assessment model. It addresses a few of our issues, such as the development of skills through its iterative feedback loops, holistic development through outcome-based assessment, and supporting the needs of our varied cohort through differentiated tasks. We discuss TOPA implementation in the next step.

**Address KI2.** In response to KI2, we discussed introducing in our assessment model the concept of 'agency'. We defined agency through a social cognitive lens as the ability to exert control over and direct ones own learning [14], [19]. An important aspect of agency as viewed through a social cognitive perspective is its developmental nature [20]. Kegan's constructive developmental theory analyzes developmental changes through adulthood in terms of evolvment and exercise of human agency [21]. According to Kegan's stages of adult development, students in higher education present: instrumental, socialized and self-authoring minds (see Table I). In our assessment design (discussed later in this section), we used Kegan's stages to understand how to better support and challenge our students holistically [21], [22].

**Addressing KI3.** In response to KI3, we wanted to introduce an assessment rubric, as part of the TOPA (discussed above). The assessment rubric should enable equitable assessment of students by focusing on research skills rather than research output only. Another aspect impacting the equitability of assessment was the issue of supervisor's control over their students' final grades, impacted by their vested interest in awarding high grades to secure PhD students and scholarships. To alleviate this issue, we discussed several alternatives, such as a SE conference program committee-style assessment process with no involvement of the supervisors, a moderation process involving a panel to moderate the final assessment conducted by supervisors, and a peer-review process.

### B. Prototype Step

In this step, we implemented the *SIT723* assessment based on student personas and TOPA. Below we present the student personas that guided our assessment model, the assessment rubric, TOPA design, and the grading process.



Fig. 2. Research Personas of Students in *SIT723*.

**Personas.** Based on the stages of adult development identified in Kegan's constructive-developmental theory [21] and present in higher education students [22] (Table I), we created the *SIT723* personas (see Fig. 2). These personas are used in the rubric design to guide us in identifying the different levels of achievement in a meaningful way, rather than arbitrarily. These personas are also used to convey expectations to both, students and supervisors.

- *Research-aware professional* describes a student at an instrumental mind stage. She requires an authority figure, in this case, a supervisor, to guide and scope the project for her. However, she is expected to be aware of all the stages of conducting research and should be able to communicate the project idea by the end of the semester. This student will achieve a Pass ( $\mathcal{P}$ ) level according to our assessment design.

- *Research-engaged professional* describes a student at a socialized mind stage. She relies on the supervisor for guidance and confirming the scope of the project. She has an understanding of all the stages of conducting research and is able to communicate the project idea. This student will achieve a Credit ( $\mathcal{C}$ ) level in our assessment design.

- *Research-capable professional* describes a student at a socialized mind stage. Once provided with an idea and clear direction, she can complete the project with limited involvement of the supervisor. She can conduct all the stages of research at an appropriate level and has a good understanding of the project idea from the start of the semester. This student will achieve a Distinction ( $\mathcal{D}$ ) level.

- *Research-ready professional* describes a student at a self-authoring mind stage. Alone, or in conjunction with the supervisor, she can define a research idea and a clear direction. She can complete the project with the supervisor as a 'critical friend'. She can conduct all the stages of research and is able to communicate and discuss the project idea. This student will achieve a High-Distinction ( $\mathcal{HD}$ ) level.

**Assessment Rubric.** Based on our discussions in the design and ideate steps, we first developed the assessment rubric aligned with the student personas and the research skills deemed relevant for *SIT723*. The assessment rubric and the criteria was developed based on the existing work on research skills development [23]–[25], SE research and evaluation guidelines [26]–[29], Computer Science and SE teaching and learning guidelines [30], [31]. Table II shows

TABLE II  
SIT723 ASSESSMENT RUBRIC.

	PASS	CREDIT	DISTINCTION	HIGH-DISTINCTION
	<i>"Research-aware professional who can follow directions in highly structured environment and a well scoped project led by the supervisor"</i>	<i>"Research-engaged professional who can perform research with clear boundaries set and with limited directions by the supervisor"</i>	<i>"Research-capable professional who can lead a research project with scaffolds provided by the supervisor"</i>	<i>"Research-ready professional who can independently lead a research project with guidance of the supervisor"</i>
Literature Review	<ul style="list-style-type: none"> <li>* Able to identify relevant literature from several sources with a limited search strategy and a limited set of relevant concepts;</li> <li>* Able to summarise literature following appropriate academic integrity guidelines.</li> </ul>	<ul style="list-style-type: none"> <li>* Able to identify relevant literature from several sources with a definite search strategy, with an analysis of relevant concepts;</li> <li>* Able to summarise literature following appropriate academic integrity guidelines, and identify gaps in the literature;</li> </ul>	<ul style="list-style-type: none"> <li>* Able to extensively identify literature from (almost) all relevant sources with a well-defined (and explained) search strategy, with integration of relevant concepts;</li> <li>* Able to synthesize literature following appropriate academic integrity guidelines, identify gaps in the literature and suggest directions for addressing the gaps;</li> <li>* Able to clearly present the innovative/novel part of the research against substantial sources reviewed;</li> </ul>	<ul style="list-style-type: none"> <li>* Able to extensively identify literature from (almost) all relevant sources with a well-defined (and explained) search strategy, with integration of relevant concepts;</li> <li>* Able to synthesize literature following appropriate academic integrity guidelines, identify gaps in the literature and suggest directions for addressing the gaps;</li> <li>* Able to clearly present the innovative/novel part of the research against substantial sources reviewed;</li> </ul>
Acad. Writing	<ul style="list-style-type: none"> <li>* Able to write in a professional manner, using academic language, on topic, with an appropriate document structure, and applying appropriate referencing standards</li> </ul>	<ul style="list-style-type: none"> <li>* Able to write using discipline-specific language, on topic, with clear document structure, and applying appropriate referencing standards;</li> <li>* Able to present the relevant arguments, maintain focus in writing, and demonstrate domain understanding;</li> </ul>	<ul style="list-style-type: none"> <li>* Able to write concisely and succinctly using discipline-specific language, on topic, with clear document structure, and applying appropriate referencing standards;</li> <li>* Able to present mostly well-reasoned arguments, maintain a consistent storyline in writing for a specific audience and demonstrate a clear understanding of the domain;</li> </ul>	<ul style="list-style-type: none"> <li>* Able to write concisely and succinctly using discipline-specific language, on topic, with clear document structure, and applying appropriate referencing standards;</li> <li>* Able to articulate well-crafted arguments, substantiate all claims, maintain a consistent storyline in writing for a specific audience, and demonstrate a clear understanding of the domain;</li> </ul>
Research Design	<ul style="list-style-type: none"> <li>* Able to adequately communicate the research design, and its component parts in context of his/her own research objectives;</li> <li>* Able to define RQs with support from the supervisors;</li> <li>* Able to list any ethics considerations of the project;</li> </ul>	<ul style="list-style-type: none"> <li>* Able to clearly communicate the research design, its component parts, and the rationale behind most decisions in the research design in context of his/her own research objectives;</li> <li>* Able to define RQs based on the gaps in the literature, with little guidance from the supervisor;</li> <li>* Able to justify any ethics considerations of the project;</li> </ul>	<ul style="list-style-type: none"> <li>* Able to clearly communicate the research design, its component parts in context of his/her own research objectives;</li> <li>* Able to interpret and assess different methodologies of research design in the context of his/her own research objectives;</li> <li>* Able to formulate RQs based on the research directions and the gaps identified in literature, and present testable hypothesis;</li> <li>* Able to analyse any ethics considerations of the project;</li> </ul>	<ul style="list-style-type: none"> <li>* Able to clearly communicate the research design, its component parts in context of the research objectives and in line with the domain publication standards;</li> <li>* Able to interpret and assess different methodologies of research design in the context of his/her own research objectives;</li> <li>* Able to formulate RQs based on the research directions and the gaps identified in literature, critically analyse the relevance and thoroughness of RQs, and present testable hypothesis;</li> <li>* Able to analyse any ethics considerations of the project;</li> </ul>
Project Management	<ul style="list-style-type: none"> <li>* Able to communicate project milestones on an appropriate timeline;</li> <li>* Able to report project progress and adherence to milestones;</li> </ul>	<ul style="list-style-type: none"> <li>* Able to draft project milestones on an appropriate timeline based on project goals;</li> <li>* Able to report project progress and adherence to milestones;</li> <li>* Able to collate and manage relevant literature, and data;</li> </ul>	<ul style="list-style-type: none"> <li>* Able to draft project plan with project milestones, timelines based on project goals and task priorities, and a risk management plan;</li> <li>* Able to report project progress and adherence to milestones;</li> <li>* Able to collate and manage the relevant literature, data, versions in a version-controlled repository;</li> <li>* Able to maintain detailed project logbook with meeting notes and project details;</li> </ul>	<ul style="list-style-type: none"> <li>* Able to independently plan and continuously manage the project with project milestones, timelines based on project goals, task priorities and a risk management plan;</li> <li>* Able to adhere to the plan, and/or proactively align project priorities and milestones based on progress, with consent of supervisors;</li> <li>* Able to collate and manage the relevant literature, data, versions in a version-controlled repository;</li> <li>* Able to maintain detailed project logbook with meeting notes, project details and evidence;</li> </ul>
Artefact Development	<ul style="list-style-type: none"> <li>* Able to exhibit some progress towards developing artefact(s), required for preliminary feasibility analysis of the research idea;</li> <li>* Able to demonstrate IT knowledge and skills relevant to their discipline;</li> </ul>	<ul style="list-style-type: none"> <li>* Able to exhibit substantial progress towards developing artefact(s), required for preliminary feasibility analysis of the research idea;</li> <li>* Able to demonstrate proficiency in IT knowledge and skills relevant to their discipline;</li> </ul>	<ul style="list-style-type: none"> <li>* Able to exhibit substantial progress towards developing artefact(s), required for preliminary (feasibility) analysis of the research idea;</li> <li>* Able to perform preliminary analysis of the artefact(s) based on research design;</li> <li>* Able to demonstrate outstanding IT knowledge and skills relevant to their discipline;</li> </ul>	<ul style="list-style-type: none"> <li>* Able to exhibit substantial progress towards developing artefact(s), required for preliminary (feasibility) analysis of the research idea;</li> <li>* Able to perform preliminary analysis of the artefact(s) based on research design;</li> <li>* Able to demonstrate outstanding IT knowledge and skills relevant to their discipline;</li> <li>* Able to communicate the technical and functional details of artefact(s) development;</li> </ul>
Evaluation	<ul style="list-style-type: none"> <li>* Able to collect and present some data required for artefact evaluation;</li> <li>* Able to use quantitative or qualitative analysis methods, with guidance from the supervisors;</li> <li>* Able to use collected data to address RQs;</li> </ul>	<ul style="list-style-type: none"> <li>* Able to collect and present some data required for artefact evaluation;</li> <li>* Able to provide a rationale for the data collected for artefact evaluation;</li> <li>* Able to explain the choice of quantitative and/or qualitative analysis methods used, in relation to the RQs;</li> <li>* Able to analyse collected data to address RQs;</li> </ul>	<ul style="list-style-type: none"> <li>* Able to collect and present sufficient data required for an effective artefact evaluation;</li> <li>* Able to provide a compelling rationale for the data selected for artefact evaluation;</li> <li>* Able to justify the choice of quantitative and/or qualitative analysis methods used, in relation to the RQs;</li> <li>* Able to analyse literature review together with collected data to address RQs, with findings and conclusions;</li> </ul>	<ul style="list-style-type: none"> <li>* Able to collect and present sufficient data required for a comprehensive and thorough artefact evaluation;</li> <li>* Able to provide a compelling rationale for the data selected for artefact evaluation;</li> <li>* Able to justify the choice of quantitative and/or qualitative analysis methods used, in relation to the RQs;</li> <li>* Able to analyse and synthesise literature review together with collected data to address RQs, with clearly expressed findings, interpretations, limitations and conclusions;</li> </ul>
Dissemination	<ul style="list-style-type: none"> <li>* Able to present a research report of 3000-8000 or 8000 words with appropriate layout;</li> <li>* Demonstrates knowledge of the individual contributions to the research project and the results achieved;</li> <li>* Able to make use of LaTeX and BibTex for building documents</li> </ul>	<ul style="list-style-type: none"> <li>* Able to present a research report of 3000-8000 words with appropriate layout, formatting, and illustrations;</li> <li>* Demonstrates knowledge of the research project, results achieved, and relevance for the research domain;</li> <li>* Able to make use of LaTeX and BibTex for building documents, with appropriate illustrations as figures and tables;</li> </ul>	<ul style="list-style-type: none"> <li>* Able to present a research report of 3000-8000 words with a well-structured layout, formatting, and illustrations;</li> <li>* Demonstrates knowledge of all facets of the research project;</li> <li>* Able to make use of LaTeX and BibTex for building documents, with appropriate illustrations as figures and tables;</li> </ul>	<ul style="list-style-type: none"> <li>* Able to present a research report of 3000-8000 words professional-level layout, formatting, and illustrations;</li> <li>* Demonstrates knowledge of all facets of the research project;</li> <li>* Able to make use of LaTeX and BibTex for building documents, with appropriate illustrations as figures and tables;</li> </ul>

TABLE III  
TOPA v1.0.

#### Phase1 – Literature Review

- T1.1 - Project Agreement & Research Integrity Training
- T1.2 - Literature Search and Ethics Plan
- T1.3 - Literature Review (Draft)

#### Phase 2 – Research Design

- T2.1 - Research Design (Draft)
- T2.2 - Project Progress Report 1
- T2.3 – Report Midway Draft

#### Phase 3 – Artefact Development & Evaluation

- T3.1 - Artefact & Evaluation Plan
- T3.2 - Artefact Development and Evaluation Report
- T3.3 - Project Progress Report 2

#### Phase 4 - Dissemination

- T4.1 - Research Report Draft
- T4.2 - Research Presentation Video
- T4.3 - Research Report
- T4.4 - Learning Summary Report (LSR)

the assessment rubric we designed and implemented in *SIT723*. The assessment rubric covers seven criteria (hereafter ARC#1–7), namely, *ARC#1 - Literature Review*, *ARC#2 - Academic (& Technical) Writing*, *ARC#3 - Research Design*, *ARC#4 - Project Management*, *ARC#5 - Artefact Development*, *ARC#6 - Evaluation*, and *ARC#7 - Research Dissemination*. The rubric presents descriptors for each criteria at all four grade levels. For a student to achieve a certain grade, they were required to match the grade for all criteria, i.e., to get a *HD*, the student had to perform at the level across all criteria from the rubric. **Task-oriented Portfolio Assessment.** Based on the discussions in the empathize, design and ideate stages, we introduced TOPA\_v1 (see Table III) where instead of students working towards a single block outcome, i.e., the research thesis, students would work on granular intermediate tasks throughout the semester – which not only added up as their research portfolio, but also ensured all criteria in the rubric were taught, developed and assessed. As shown in Table III, we split the project in four phases, discussed below.

- **Phase 1: Literature Review** focused on project scoping and literature review, and contributed to criteria (ARC#1, #2, #4, and #7). In Phase 1, students had workshops to learn more about Research Processes, Literature Review and Project Management, and the workshops further offered them opportunities to discuss with their peers and the unit team, specifics of their own project. The students were asked to submit following tasks on *OnTrack* – our web-based outcome assessment system [18], [32]. All the tasks on *OnTrack* were reviewed by the students’ supervisors. *OnTrack* offers an option to mark a given task as complete, assign a grade for the graded tasks, or ask the student to resubmit the tasks after taking in the feedback provided for each task. The tasks students worked on and submitted in this phase are:

- *T1.1 - Project Agreement and Training* - The students submitted a task on their project agreement with their

supervisors, with the details of what needs to be done in the project based on an initial meeting between the student and the supervisor. The students were further required to complete a training module on research ethics and research integrity, and students uploaded the evidence of completion on *OnTrack* in T1.1.

- *T1.2 - Literature Search and Ethics Plan* - The students submitted their literature search strategy and results, e.g., the search terms and queries. The students were further asked to submit their ethics plan if their project required contact with any human participants.
- *T1.3 - Literature Review (Draft)* - The students were then asked to submit the draft of their literature review, for feedback from their supervisors at the end of Phase 1.
- **Phase 2: Research Design** focused on developing the research design, i.e., planning the development of a relevant research artefact, formulating research questions (RQs), and planning the research evaluation. The students also addressed the supervisors’ feedback on T1.3 from Phase 1. Phase 2 contributed primarily to criteria (ARC#1, #2, #3, #4, and #7). In Phase 2, students had workshops on Formulating Research Questions and Artefacts and Evaluation in SE research. The tasks students worked on and submitted in this phase are:

- *T2.1 - Research Design (Draft)* - The students submitted the first draft of their RQs and the artefact development and evaluation plan to their supervisors on *OnTrack*.
- *T2.2 - Project Progress Report 1* - The students submitted the overall progress report using a template provided to them. The students reflected on their progress, by positioning their progress to the assessment rubric criteria. As a part of the reflection process, the students self-assessed by grading themselves and providing evidence supporting their self-assessment. This was a graded task, so the supervisors assigned a grade and provided detailed feedback on students’ reflections against the assessment rubric, helping them develop evaluative judgment skills [33].
- *T2.3 - Report Mid-Way Draft* - This task focused on a first draft of the research report formatted in  $\LaTeX$ . The students had access to an official template from the university on Overleaf<sup>1</sup>. The first draft was expected to include Introduction, Literature Review and Research Design, with supervisors’ feedback from previous tasks addressed by the students.

- **Phase 3: Artefact** focused on the development of a research artefact, e.g., a software prototype, algorithm, user interface design and survey design. Phase 3 contributed primarily to criteria (ARC#3, #4, #5 and #7). In Phase 3, students had workshops on academic writing and research dissemination. The tasks students worked on and submitted in this phase are:

- *T3.1 - Artefact & Evaluation Plan* - The students submitted (early on in Phase 3) a detailed plan for artefact development, evidence of the required (programming) environment setup, and an artefact evaluation plan.

<sup>1</sup><https://www.overleaf.com/latex/templates/deakin-sit-thesis-template/tmqzqtctmdw>

- *T3.2 - Artefact Development and Evaluation Report* - The students submitted evidence of their artefact development completion and the artefact evaluation results.
- *T3.3 - Project Progress Report 2* - This task consisted of a progress report, similar to T2.2.
- **Phase 4: Dissemination** focused on the writing and dissemination phase, wherein the students completed their final research report. Phase 4 contributed to all criteria, as students revised all sections of their research report. No workshops were planned for this phase, except on-demand contact sessions with the unit team for clarifying doubts. The tasks students worked on and submitted in this phase are:
  - *T4.1 - Research Report Draft* - The students submitted (early on in Phase 4) the completed draft of their research report in L<sup>A</sup>T<sub>E</sub>X to get feedback from their supervisors.
  - *T4.2 - Research Presentation Video* - In this task, the students recorded a three minute video describing their research project and their research contributions. The task was targeted at helping students convey their research outputs in a short time span.
  - *T4.3 - Research Report* - The students submitted their final research report after addressing the feedback from the supervisors in T4.1.
  - *T4.4 - Learning Summary Report (LSR)* - The students submitted their LSR, wherein they reflect on their overall learning and progress in *SIT723*, including their target grade with an explanation of self-assessment against the assessment rubric [34]. Once the LSR was submitted, *On-Track* enabled the students to automatically generate their ‘portfolio’, that included all the previous task submissions and their supervisor’s feedback.

**Grading Process.** At the end of the semester, the student portfolios were assessed to assign a final grade. From the three prospective solutions discussed for addressing KI3 in Section IV-A, we implemented an adaptation of peer-review process with supervisor as one of the participants. Intermediary feedback to students on individual tasks in different phases (discussed above) during the teaching period was provided by the supervisor. The final grading on the portfolio was performed by three academics - the supervisor and two independent assessors. A large number of academics were involved in the process, as we attempted to engage as many people as possible. The three assessors for each portfolio did not know of each other’s identity, and anonymously reviewed the portfolios. The distribution of the portfolios, random selection of independent assessors, and finalization of grades was managed by the unit team. In case of a (major) discrepancy among the three academics, the unit team moderated the assessor’s responses and assigned the final grade.

### C. Evaluation Step

We ran the redesigned *SIT723* for the first time in the first semester of 2021. We collected verbal feedback from several supervisors on the unit design, both during the semester and at the end of the semester. Most supervisors appreciated the

structure of the unit and the clarity provided by the personas and assessment rubric in guiding students. The *SIT723* design definitely helped in addressing issues across KIs1–3. Anecdotally, in KI1, the focus on different criteria helped students focus on different aspects. For instance, most supervisors reported improved project management by students. For KI2, the personas and assessment rubric captured the varied level of guidance and helped aligning the student expectations, and several students self-assessed to adjust their target grades. For addressing KI3, we did receive positive feedback overall, but we were not sure of the extent to which the issue had been addressed. This was due to the fact that the unit ran during COVID-19 lockdowns (in an online mode only), and only for a single iteration. The *HD* rate was  $\approx 35\%$  in this iteration. Having said that, there were a few further key issues (KIs) raised by several supervisors and the unit team.

**KI4 - Flexibility in TOPA\_v1.** While the TOPA\_v1 task structure helped organize the intermediate tasks for students, students and supervisors felt restricted. The research process and progress in individual projects was not always aligned with the sequence of tasks in the TOPA\_v1 design. For example, some exploratory projects focused longer on literature review and lasted much longer across first three phases of Fig. III.

🔗 **Flexibility in milestone sequencing** - Each SE research project is unique. The assessment model should focus on skills and intermediate milestones rather than the sequence or time-frame in which the milestones need to be achieved.

**KI5 - Flexibility in the assessment rubric.** The supervisors reported on the difficulty for students to perform well across all criteria, given the relatively short time span, the issues with the diversity of SE research projects in general, and project requirements. For example, a student working on the topic of “practitioners perspective on AI-based software for regulatory compliance”, performed an extensive literature review and developed and conducted a survey with practitioners as her research project. While the project did not develop an artefact, typically considered in a SE project, e.g., a software prototype, the project had clear scientific contributions. Therefore, the assessment rubric required flexibility in its application.

🔗 **Flexibility in the assessment rubric** - The flexibility of SE research projects and the limited time span of undergraduate research projects should be reflected in the assessment rubric. This is encouraged by the holistic assessment of the student’s performance against the persona.

## V. *SIT723* REDESIGN VERSION 2.0

### A. Design and Ideate Steps

After the first run of the unit in early 2021, we worked on addressing KI4 and KI5 from version 1.0.

**Address KI4.** While most supervisors and the unit team observed the benefits of the task-oriented model with intermediary feedback, the issues related to the sequencing of the milestones and flexibility in the assessment model required further changes. After rounds of discussions, we proposed a project-reporting based assessment model (TOPA\_v2), which used the TOPA\_v1 design in the background while offering flexibility to the supervisors and students. TOPA\_v2 required regular reporting of project progress, but the students and supervisors were free to sequence the milestones as they wished, as long as the final milestones aligned with the assessment rubric. The students were also provided information from TOPA\_v1 tasks as guiding material.

**Address KI5.** Similar to TOPA\_v1, all stakeholders saw value in the assessment rubric, however some adjustments were required to offer flexibility to students to cater to the diverse SE research projects. We did not change the assessment rubric much, except a single change across all four grade levels for the Dissemination criterion. We made the use of  $\LaTeX$  as optional in the assessment rubric, as some students and supervisors were not comfortable with report writing in  $\LaTeX$ .

### B. Prototyping Step

In this version of *SIT723*, we introduced the changes to the TOPA model (TOPA\_v2), and the rubric's implementation in assessment. The personas and the grading process remained unchanged. Below we present the changes.

**Task-oriented Portfolio Assessment (TOPA\_v2).** Unlike, TOPA\_v1 that focused on sequencing intermediate milestones and tasks for SE research projects, TOPA\_v2 focused on project reporting on a weekly basis. Students submitted a weekly status reports, focused on reporting their weekly progress based on supervision meetings, summary of progress against the overall milestones, issues faced, and assistance required, e.g., issues related to accessing the university IT infrastructure. Students were encouraged to follow the four phases in TOPA\_v2, however, the tasks' submission did not enforce the phases. We retained TOPA\_v1 tasks T1.1, T2.2, T3.3, and T4.1-4.4 in TOPA\_v2. The project progress report tasks (T2.2 and T3.3) were formatively graded, i.e., students received an indicative grade from their supervisors along with the feedback on their progress against the assessment rubric.

**Assessment Rubric Implementation.** As discussed in Section V-A, we did not change much in the assessment rubric criteria, however, we did change the implementation of the rubric and the related expectations. While some supervisors assessed based on the persona profiles, most took the rubric literally. Based on supervisors' feedback (KI5), we lowered the expectation of achieving a matching grade across all criteria to be eligible for that overall grade. This was particularly relevant for the Academic Writing criterion, as we have hundreds of international students from different nationalities, with English as their second or the third language, who were struggling to perform at higher levels in academic writing.

We marked Project Management and Dissemination (ARC#4 and ARC#7) as core criteria for which the students

had to match the grade, e.g., for an overall  $\mathcal{D}$ , they had to get at least the Distinction in Project Management and Dissemination. For the remaining five criteria (ARC#1, ARC#2, ARC#3, ARC#5 and ARC#6), we had following conditions for achieving different overall grade levels:

- **Pass.** All five criteria at  $\mathcal{P}$ ;
- **Credit.** At least three of these five criteria at  $\mathcal{C}$ . The remaining two criteria at  $\mathcal{P}$  or higher;
- **Distinction.** At least two of the five criteria at  $\mathcal{D}$ . The remaining two criteria at  $\mathcal{C}$  or higher;
- **High Distinction.** At least one of these criteria at  $\mathcal{HD}$ . From the remaining criteria, at least two need to be at  $\mathcal{D}$  or higher. All remaining criteria need to be at  $\mathcal{C}$  or higher.

If a student ( $S_x$ ) had the following grades, ARC#1 ( $\mathcal{C}$ ), ARC#2 ( $\mathcal{C}$ ), ARC#3 ( $\mathcal{D}$ ), ARC#4 ( $\mathcal{HD}$ ), ARC#5 ( $\mathcal{HD}$ ), ARC#6 ( $\mathcal{D}$ ), and ARC#7 ( $\mathcal{HD}$ ), they would get an overall  $\mathcal{HD}$  based on the conditions mentioned above. Let's say another student ( $S_y$ ) had the same grades as  $S_x$  for ARCs#1–5 and ARC#7, except a  $\mathcal{HD}$  in ARC#6 (instead of  $\mathcal{D}$ ), they would also get a  $\mathcal{HD}$  grade but a higher mark (out of 100) than  $S_x$ . So,  $S_x$  would receive a mark of 80 ( $\mathcal{HD}$ ), and  $S_y$  would get a mark of 85 ( $\mathcal{HD}$ ).

### C. Evaluation Step

We implemented the next version of *SIT723* in the second half of 2021. We again collected the verbal feedback over the semester from the unit team and supervisors on the unit design and the new changes. All stakeholders reacted extremely positively to the changes introduced in TOPA\_v2 and the flexibility introduced in the implementation of the assessment rubric, i.e., the KI4 and KI5 had been addressed to a large extent. At the end of the semester, we further observed the grading process (KI3) and discovered that issues related to assessor bias [12], [35] had not been mitigated. The situation was better than before the redesign (version 1.0), i.e., supervisors' solely grading the students but there were still remaining issues.

**KI6 - Administrative Issues and Bias in the Grading Process.** There were numerous issues in the previous grading process. Firstly, from an administrative perspective managing the process described in Section IV-B for hundreds of students in a semester was taxing and cumbersome for the unit team. Secondly, the involvement of 40-50 assessors and anonymization attempts in the grading process, further exacerbated the administrative issues. Finally, supervisor's bias still played a substantial role in the final grade as the supervisor was an active part of the grading process. Due to these issues, the overall implementation of the units redesign suffered as the unit team struggled to effectively moderate all the portfolios in this process, and the supervisors' bias was still influential.

💡 **Getting the grading process right is difficult** - The grading process needs to be carefully designed, not only to be equitable but also to be feasible within the university's contextual constraints such as the staff workload issues.



## VI. SIT723 REDESIGN VERSION 3.0

### A. Design and Ideate Steps

In version 3.0, we primarily worked on addressing KI6.

**Address KI6.** We discussed several alternatives for improving the grading process and addressing the issues with supervisors' bias, within our organizational context and constraints, e.g., the workload available for the staff members for the grading process and the duration of assessment for a given solution. Our alternatives included, among others, (i) continuous assessment by an external assessor, however, this was deemed infeasible due to the workload (scalability) issues with hundreds of students in *SIT723* in a semester; and (ii) a grading system similar to a conference peer-review system [36]. We implemented a customized version of the (ii) alternative.

### B. Prototype Step

In this version of *SIT723*, we introduced the changes to the grading process. All other aspects of the unit, e.g., the TOPA\_v2, personas and the assessment rubric seemed to work in a stable mode – without any major issues. The new grading moderation process works as follows:

- At the end of the semester, the supervisor assigns a tentative grade and provides detailed comments for each criterion based on the rubric in *OnTrack*. We note that the end of the semester assessment from the supervisor, in addition to the weekly and milestones-based feedback.
- The student's research portfolio, self-assessment of the student against the assessment rubric in LSR, and the supervisor's comments against the assessment rubric are passed onto a "moderation panel".
- The panel consists of  $\approx 10$  staff members, and functions very similar to the conference program committees. The panel convenes over a working day, and each member receives workload for  $\approx 1.5$  working days for participation.
- The primary task of panel members is to *moderate* the supervisors' tentative grade based on detailed comments and student's self-assessment and the portfolio. The panel members do not *mark* the portfolio to address workload issues, i.e., the assessment of hundreds of students by multiple staff members is not scalable. Each portfolio is first moderated by a pair of panel members. Thereafter, the grade is finalized based on:
  - If the pair agrees with each other on the final grade and the mark is the same as the supervisor's grade  $\rightarrow$  the grade is finalized with no further discussion;
  - If the pair agrees on the final grade, but the grade is not the same as supervisor's grade  $\rightarrow$  the case will be briefly discussed among all panel members, and the pair might be asked to justify the grade change;
  - If the pair disagrees with each other on the final grade, then a third panelist is asked to review the grade and reach consensus independently  $\rightarrow$  the case is also be discussed with the entire panel;

### C. Evaluation Step

In the first semester of 2022, we implemented version 3.0. The grading process seemed to work much better than the previous version, as supervisor was not an active part of the process. The overall impact of the supervisors' influence had decreased, and the overall *HD* rate had significantly decreased as compared to the version before version 1.0. We note that, due to the limited availability of staff members and workload-related issues, we only ran the moderation process for portfolios assessed by supervisors as *F*, *D* or *HD*. For the portfolios graded by supervisors as *P* or *C*, the supervisors' grade was final. The rationale behind this is that only students with *D* or *HD* are eligible for a PhD entrance, and *F* grades required further review to ensure the fairness in the process. While this version is not perfect, it is a significant improvement over the previous versions in addressing supervisor grading bias, according to the unit team and the moderation panel.

At the end of version 3.0, there are still some issues, e.g., supervisors still being responsible for providing intermediate grades on project progress reports (T2.2 and T3.3 in Table III) and the final portfolio, which need addressing and we will attempt to address them in the future semesters.

🔗 **Supervisor's should focus on providing feedback rather than grading** - Supervisors should be responsible for mentoring and providing critical feedback to students at each stage of the process and for facilitating the moderation process. Grading can take into account supervisors' views but they should not be an active part of the grading process.

🔗 **Fidelity of process implementation is challenging due to variability of supervisors** - The fidelity of implementation of the assessment model remains a challenge with the variability of supervisors. Due to this, it is difficult to maintain assessment consistency and equitability.

## VII. LESSONS LEARNT

We started this project to improve the reliability of our assessment of final-year SE research projects, which demonstrated challenges in terms of variability in students' skills, nature of projects, and quality of supervision. Through the process of developing and refining the assessment model for *SIT723* we have largely been able to develop an assessment model that largely addresses the challenges associated with students and projects, and we are working on helping support staff to ensure the quality of supervision improves.

Personas are helping clarify what is expected of students, with grade-level descriptors assisting in communicating the standard of work required from the student and the support required from the supervisor. Rubric descriptors encourage students to seek greater agency and take more responsibility for their learning and research project. This helps demonstrate the benefits of the development rubric, which links assessment results with desired outcomes. Example outcomes of

*SIT723* for a student with high levels of agency are [37]–[40], which were accepted for publication in peer-reviewed venues.

The flexibility provided by TOPA and enacted by adjusting the task requirements in v2.0 has allowed the model to cater to a wide variety of project types associated with SE research, from highly practical to fully theoretical. In assessing the research performance holistically, the challenge is to provide a sufficient structure within the assessment to have defined feedback points without overly constraining the kinds of projects that students can engage with. By shifting to a flexible task structure, we have been able to support students with formative feedback while allowing a wide variety of projects and an overall assessment that takes into consideration the wide range of factors that make up successful SE research projects. Furthermore, the TOPA\_v2 has been great in providing space for academics to make adjustments to assessment without requiring adjustments to published curriculum documentation.

While the assessment rubric has been useful in capturing and representing criteria in a traditional two-dimensional table structure, it requires adjustments to encourage the holistic assessment aimed for. Reactively adjusting the criteria needed to achieve specific grades helped to mitigate some of the issues we experienced. However, this highlights the need for a rethink of the presentation of the rubric, which may assist in achieving greater alignment with intended grade-level outcomes.

Variability in the quality of supervision is a remaining challenge. Implementing this model at scale has involved working with a large number of supervisors, and highlighted the difficulties of designing assessment to be delivered by academics who are not involved in the design [41]. Moderation processes can aid in identifying and addressing some of the issues that arise from this, mitigating its impact on final results. However, we need to rethink aspects of the model to ensure that we are able to identify and support students earlier in the process - thereby ensuring they have the support they require.

More generally, this continuous improvement process has demonstrated the importance of flexibility in the assessment model. Assessment within the unit has retained the same assessment model across all iterations, consisting of frequent formative feedback, delayed summative grading, and standards-aligned outcomes-based assessment. By making iterative changes to the assessment strategy, we have been able to improve the fidelity of the implementation, progressively adjusting the unit to address challenges that have arisen.

### VIII. RELATED WORK

A large body of work exists on students research projects in general [42]–[45], students group-based SE capstone projects [17], [46], [47], however only a few research strands have covered student research projects design or research training in SE. Bernat et al. [48] present a model on SE student research training and engagement, with the close collaboration of SE research groups and the students. In this model, each student define activities and timelines for a task assigned to them. The assigned tasks are deliverables (artefacts), such as

literature review, software products or documentation. Similar to our assessment rubric, this model also focuses on helping students develop dissemination and writing skills.

On similar lines, the structure of a SE Master Thesis program has been discussed in detail for a Swedish university [26], [49]–[51]. In these papers, the authors present the idea of using rubric-based assessment, wherein four rubrics cover the research proposal, research project process, research thesis, and oral presentation. Similar to our assessment rubric of Table II, their rubrics state quality criteria and different levels (superior, good, fair and minimal) of quality for each criterion. While the authors mention these rubrics in their papers, to the best of our knowledge, the complete versions of the rubrics are not available publicly. A subset of the rubrics on timeliness (project), balance (project), and ethical issues (thesis) is presented [49]. Seyed-Abbassi [52] reports on the implementation of SE research projects in a database unit, wherein the students worked in groups on literature review, software artefacts, including software requirements, implementation, academic writing and dissemination via a report, poster demonstrations and peer presentations. This paper also presents the assessment of the research project by assigning points to each intermediate task, and the points are aggregated at the end of the semester by the unit team.

Similar to the papers discussed above, a few strands exist on (i) collaboration with the research groups and industry for improving students' research skills [53], [54], and (ii) research training and skills on specific SE tasks, such as automated code generation [55], and software testing [56], but none of these papers present details on the actual design of the unit and the assessment, which is the Achilles's heel of implementing a research unit or program, like *SIT723*.

### IX. CONCLUSIONS

In this paper, we reflect on the evolution of our assessment model for SE research project units through a design thinking approach. The main challenges this model aims to address consist of catering to a wide range of projects, students' skills, and supervisor quality. Key lessons indicate the importance of designing a flexible and holistic approach for assessing research projects. Our focus on student agency (personas) and on having continuous, process-focused tasks that allow students to report on different areas at regular intervals, provided both the flexibility and structure required by the range of research projects present in SE. Due to the impact that students' grades in this units have on their opportunities to pursue a PhD, the grading process became a strong focus of this redesign. Our moderation processes were able to address the challenges of providing reliable grades, however, we still need to find a timely way to identify and address the challenges of differences in quality of supervision. This is essential for ensuring equitability of opportunity in such a high stake unit.

Going forward we will continue using this design thinking approach. In the next iteration we will go through the empathize stage, this time focusing on empathizing with the supervisors' experience of the assessment.

## REFERENCES

- [1] A.-B. Hunter, S. L. Laursen, and E. Seymour, "Becoming a scientist: The role of undergraduate research in students' cognitive, personal, and professional development," *Science education*, vol. 91, no. 1, pp. 36–74, 2007.
- [2] T. Hatzia Apostolou, D. Dranidis, A. Sotiriadou, P. Kefalas, and I. Niko- lakopoulos, "An authentic student research experience: fostering research skills and boosting the employability profile of students," in *Proceedings of the 23rd Annual ACM Conference on Innovation and Technology in Computer Science Education*, 2018, pp. 254–259.
- [3] E. Knauss, "Constructive master's thesis work in industry: guidelines for applying design science research," in *2021 IEEE/ACM 43rd International Conference on Software Engineering: Software Engineering Education and Training (ICSE-SEET)*. IEEE, 2021, pp. 110–121.
- [4] M. Malachowski, J. M. Osborn, K. K. Karukstis, J. Kinzie, and E. L. Ambos, "Institutionalizing undergraduate research and scaffolding undergraduate research experiences in the stem curriculum," in *Best Practices for Supporting and Expanding Undergraduate Research in Chemistry*. ACS Publications, 2018, pp. 259–269.
- [5] S. Chakraborty, L. Deng, and J. Dehlinger, "Towards authentic undergraduate research experiences in software engineering and machine learning," in *Proceedings of the 3rd International Workshop on Education through Advanced Software Engineering and Artificial Intelligence*, 2021, pp. 54–57.
- [6] H.-P. Bischof, J. D. Furst, D. S. Raicu, and S. D. Ruban, "Top issues in providing successful undergraduate research experiences," in *Proceedings of the 42nd ACM technical symposium on Computer science education*, 2011, pp. 239–240.
- [7] T. A. KELLEY, *The art of innovation: Lessons in creativity from IDEO, America's leading design firm*. Broadway Business, 2001, vol. 10.
- [8] H. Plattner, "Bootcamp bootleg," *Design School Stanford, Palo Alto*, 2010.
- [9] J.-G. Schneider, P. W. Eklund, K. Lee, F. Chen, A. Cain, and M. Abdelrazek, "Adopting industry agile practices in large-scale capstone education," in *2020 IEEE/ACM 42nd International Conference on Software Engineering: Software Engineering Education and Training (ICSE-SEET)*. IEEE, 2020, pp. 119–129.
- [10] K. Ahmad, M. Bano, M. Abdelrazek, C. Arora, and J. Grundy, "What's up with requirements engineering for artificial intelligence systems?" in *2021 IEEE 29th International Requirements Engineering Conference (RE)*. IEEE, 2021, pp. 1–12.
- [11] K. Ahmad, M. Abdelrazek, C. Arora, M. Bano, and J. Grundy, "Requirements engineering for artificial intelligence systems: A systematic mapping study," 2022. [Online]. Available: <https://arxiv.org/abs/2212.10693>
- [12] S. K. Green and R. L. Johnson, *Assessment is essential*. McGraw-Hill Higher Education, 2010.
- [13] O. Hazzan and J. E. Tomayko, "Reflection processes in the teaching and learning of human aspects of software engineering," in *17th Conference on Software Engineering Education and Training, 2004. Proceedings*. IEEE, 2004, pp. 32–38.
- [14] G. Biesta and M. Tedder, "Agency and learning in the lifecourse: Towards an ecological perspective," *Studies in the Education of Adults*, vol. 39, no. 2, pp. 132–149, 2007.
- [15] G. O'Callaghan and C. Connolly, "Developing creativity in computer science initial teacher education through design thinking," in *United Kingdom & Ireland Computing Education Research conference.*, 2020, pp. 45–50.
- [16] T. Love and T. Cooper, "Designing online information systems for portfolio-based assessment: Design criteria and heuristics," *Journal of Information Technology Education: Research*, vol. 3, no. 1, pp. 65–81, 2004.
- [17] L. Tubino, A. Cain, J.-G. Schneider, D. Thiruvady, and N. Fernando, "Authentic individual assessment for team-based software engineering projects," in *2020 IEEE/ACM 42nd International Conference on Software Engineering: Software Engineering Education and Training (ICSE-SEET)*. IEEE, 2020, pp. 71–81.
- [18] A. Cain, L. Tubino, and S. Krishnan, "Using technology to enable a shift from marks to outcomes-based assessment," in *Re-imagining University Assessment in a Digital World*. Springer, 2020, pp. 229–245.
- [19] A. Eteläpelto, K. Vähäsantanen, P. Hökkä, and S. Paloniemi, "What is agency? conceptualizing professional agency at work," *Educational research review*, vol. 10, pp. 45–65, 2013.
- [20] A. Bandura, "Adolescent development from an agentic perspective," *Self-efficacy beliefs of adolescents*, vol. 5, pp. 1–43, 2006.
- [21] R. Kegan, *In over our heads: The mental demands of modern life*. Harvard University Press, 1994.
- [22] M. B. Baxter Magolda, *Making their own way: Narratives for transforming higher education to promote self-development*. Stylus Publishing, LLC., 2004.
- [23] J. Willison and K. O'Regan, "The researcher skill development framework," <https://www.adelaide.edu.au/melt/ua/media/51/rsd-framework.pdf>, 2008.
- [24] J. W. Willison, "When academics integrate research skill development in the curriculum," *Higher Education Research & Development*, vol. 31, no. 6, pp. 905–919, 2012.
- [25] J. Willison and F. Buisman-Pijlman, "Phd prepared: research skill development across the undergraduate years," *International Journal for Researcher Development*, vol. 7, no. 1, pp. 63–83, 2016.
- [26] R. Feldt, M. Höst, and F. Lüders, "Generic skills in software engineering master thesis projects: Towards rubric-based evaluation," in *2009 22nd Conference on Software Engineering Education and Training*, 2009, pp. 12–15.
- [27] B. A. Kitchenham, S. L. Pfleeger, L. M. Pickard, P. W. Jones, D. C. Hoaglin, K. El Emam, and J. Rosenberg, "Preliminary guidelines for empirical research in software engineering," *IEEE Transactions on software engineering*, vol. 28, no. 8, pp. 721–734, 2002.
- [28] C. Wohlin, P. Runeson, M. Höst, M. C. Ohlsson, B. Regnell, and A. Wesslén, *Experimentation in software engineering*. Springer Science & Business Media, 2012.
- [29] K.-J. Stol and B. Fitzgerald, "Guidelines for conducting software engineering research," in *Contemporary Empirical Methods in Software Engineering*. Springer, 2020, pp. 27–62.
- [30] G. Bavota, A. De Lucia, F. Fasano, R. Oliveto, and C. Zottoli, "Teaching software engineering and software project management: An integrated and practical approach," in *2012 34th international conference on software engineering (ICSE)*. IEEE, 2012, pp. 1155–1164.
- [31] O. Hazzan, T. Lapidot, and N. Ragonis, *Guide to teaching computer science: An activity-based approach*. Springer, 2015.
- [32] A. Cain, C. J. Woodward, and S. Pace, "Examining student progress in portfolio assessed introductory programming," in *Proceedings of 2013 IEEE International Conference on Teaching, Assessment and Learning for Engineering (TALE)*. IEEE, 2013, pp. 67–72.
- [33] J. Guest and R. Riegler, "Knowing he standards: how good are students at evaluating academic work?" *Higher Education Research & Development*, vol. 41, no. 3, pp. 714–728, 2022.
- [34] L. Tubino, A. Cain, J.-G. Schneider, D. Thiruvady, and N. Fernando, "Authentic individual assessment for team-based software engineering projects," in *2020 IEEE/ACM 42nd International Conference on Software Engineering: Software Engineering Education and Training (ICSE-SEET)*. IEEE, 2020, pp. 71–81.
- [35] P. Yeates, P. O'Neill, K. Mann, and K. W. Eva, "'you're certainly relatively competent': assessor bias due to recent experiences," *Medical education*, vol. 47, no. 9, pp. 910–922, 2013.
- [36] N. A. Ernst, J. C. Carver, D. Mendez, and M. Torchiano, "Understanding peer review of software engineering papers," *Empirical Software Engineering*, vol. 26, no. 5, pp. 1–29, 2021.
- [37] K. Olorunnife, K. Lee, and J. Kua, "Automatic failure recovery for container-based iot edge applications," *Electronics*, vol. 10, no. 23, p. 3047, 2021.
- [38] O. Wang, B. Cheng, T. Hoang, C. Arora, and X. Liu, "Virtual reality enabled human-centric requirements engineering," in *2021 36th IEEE/ACM International Conference on Automated Software Engineering Workshops (ASEW)*. IEEE, 2021, pp. 159–164.
- [39] K. Lefevre, C. Arora, K. Lee, A. Zaslavsky, M. R. Bouadjeneq, A. Hassani, and I. Razzak, "Modelops for enhanced decision-making and governance in emergency control rooms," *Environment Systems and Decisions*, pp. 1–15, 2022.
- [40] A. Cobb, J.-G. Schneider, and K. Lee, "Towards higher-level abstractions for quantum computing," in *Australasian Computer Science Week 2022*, ser. ACSW 2022. New York, NY, USA: Association for Computing Machinery, 2022, p. 115–124. [Online]. Available: <https://doi.org/10.1145/3511616.3513106>
- [41] L. Tubino, J.-G. Schneider, A. Cain, D. Thiruvady, and C. Ranaweera, "Reforming assessment: challenges beyond design," in *2021 IEEE/ACM 43rd International Conference on Software Engineering: Software Engineering Education and Training (ICSE-SEET)*. IEEE, 2021, pp. 78–88.

- [42] M. Healey and A. Jenkins, *Developing undergraduate research and inquiry*. Higher Education Academy York, 2009.
- [43] R. Matzen and R. Alrifai, "Defining undergraduate research in computer science: a survey of computer science faculty," *Journal of Computing Sciences in Colleges*, vol. 27, no. 3, pp. 31–37, 2012.
- [44] M. C. Linn, E. Palmer, A. Baranger, E. Gerard, and E. Stone, "Undergraduate research experiences: Impacts and opportunities," *Science*, vol. 347, no. 6222, p. 1261757, 2015.
- [45] J. W. McManus and P. J. Costello, "Project based learning in computer science: a student and research advisor's perspective," *Journal of Computing Sciences in Colleges*, vol. 34, no. 3, pp. 38–46, 2019.
- [46] M. Paasivaara, D. Vodă, V. T. Heikkilä, J. Vanhanen, and C. Lassenius, "How does participating in a capstone project with industrial customers affect student attitudes?" in *Proceedings of the 40th International Conference on Software Engineering: Software Engineering Education and Training*, 2018, pp. 49–57.
- [47] M. Paasivaara, J. Vanhanen, and C. Lassenius, "Collaborating with industrial customers in a capstone project course: the customers' perspective," in *2019 IEEE/ACM 41st International Conference on Software Engineering: Software Engineering Education and Training (ICSE-SEET)*. IEEE, 2019, pp. 12–22.
- [48] A. Bernat, P. J. Teller, A. Gates, and N. Delgado, "Structuring the student research experience," in *Proceedings of the 5th annual SIGCSE/SIGCUE ITiCSEconference on Innovation and technology in computer science education*, 2000, pp. 17–20.
- [49] G. Dodig-Crnkovic and R. Feldt, "Professional and ethical issues of software engineering curricula," 2010, 2009.
- [50] G. Dodig-Crnkovic, F. Lüders, M. Höst, and R. Feldt, "Improved support for master's thesis projects in software engineering," *Rapport nr.: Rapport från NSHU*, 2010.
- [51] M. Höst, R. Feldt, and F. Lüders, "Support for different roles in software engineering master's thesis projects," *IEEE Transactions on Education*, vol. 53, no. 2, pp. 288–296, 2010.
- [52] B. Seyed-Abbassi, "Practical aspects of promoting research in a graduate course," *Director*, p. 07, 2006.
- [53] M. L. Shaw and B. R. Gaines, "A research-based masters program in the workplace," *Proceedings of WCCCE*, vol. 98, p. 3rd, 1998.
- [54] M. S. Bargh, A. van Rooij-Peiman, L. Remijn, and S. Choenni, "Research skills for software engineering undergraduates in dutch universities of applied sciences," in *Proceedings of the 15th Annual Conference on Information technology education*, 2014, pp. 101–108.
- [55] Y. Khmelevsky, G. Hains, and C. Li, "Automatic code generation within student's software engineering projects," in *Proceedings of the Seventeenth Western Canadian Conference on Computing Education*, 2012, pp. 29–33.
- [56] A. S. Oyeniran, T. Ademilua, M. Kruus, and R. Ubar, "Environment for innovative university research training in the field of digital test," in *2021 30th Annual Conference of the European Association for Education in Electrical and Information Engineering (EAEEIE)*. IEEE, 2021, pp. 1–6.