I betrayed my ethical principles: Investigating master's dissertation marking practices in CS

Mireilla Bikanga Ada
School of Computing Science
The University of Glasgow
Glasgow, United Kingdom
https://orcid.org/0000-0001-5406-6935

Abstract— **The project or dissertation is the most significant element in a master's program. In the UK, this assessment is worth a third or more credits out of 180 and contributes to student degrees, which is seen as a benchmark for employability. Despite the increased regulation and accountability regarding academic standards, there is very little debate on grading practices, particularly dissertation marking. The limited research on the assessment of master's dissertations indicates the need for more research in that area. This paper presents a study that investigated markers' (*n*=31) master's project dissertations marking practices in Computing Science at a university. Kruskal-Wallis tests revealed that participants' marking practices were not statistically different despite having different marking loads and master's dissertation marking experiences. Findings also show that many assessors do not typically use marking schemes; the marking load negatively affects their marking. Most would prefer complete anonymity of marking, including blind negotiation. A third of assessors have had issues marking dissertations outside their area of expertise. Thematic analysis of the qualitative data corroborated those findings and revealed three overarching themes: 1) I felt that I betrayed my ethical principles, 2) I tried to acknowledge my bias/familiarity with the topic, and 3) Improving the marking experience, second marking and reconciliation process. The paper highlights four key points to consider to improve the master's dissertation marking process.**

Keywords—Master's education, dissertation marking, marking schemes, master's theses, higher education, master's dissertation

I. INTRODUCTION

The individual project or dissertation is the most significant element in a master's program. In the UK, this assessment is worth a third or more credits out of 180 and contributes to student degrees, which is seen as a benchmark for employability [1]. Marking or grading of dissertations is part of assessment and feedback, key elements in the student learning development within and beyond formal education settings [2]. However, despite the increased regulation and accountability regarding academic standards, there is very little debate on grading practices [3] and dissertation marking in particular [4]. While the literature on marking undergraduate and PhD dissertations is well established, there is limited empirical research on marking master's dissertation projects. Reference [5] notes there are not many studies on master's dissertation assessment despite master's students being the larger group when compared to PhD students.

Determining and ensuring that dissertation marking meets all stakeholders' requirements is one of the quality assurance processes. One of the processes is ensuring that marking is reliable and the student gets a fair grade and avoids potential bias through the use of marking rubrics, also known as marking schemes or marking criteria. Marking rubrics can improve reliability in marking essays style or dissertation work and are often hailed as a method for a fairer marking process [6] and enable a relatively common interpretation of student performance [7]. However, some still argue that assessment is not "universally fair" (p. 66) [8] because assessors "are bounded by their social and cultural environment and expectations" (p. 647) [9]. Summative marking practices, in particular, can be unfair and 'intellectually and morally indefensible" (p. 233) [10]. Quality assurance procedures such as second marking (double-marking) have not always guaranteed reliability and fairness because they are based on 'idiosyncratic institution rules' (p. 233–234).

Despite the greater potential for reliable marking in local settings because the assessor may have developed a shared understanding as they work closely together (p. 218), more research on the subjectivity and differences in marking within and across universities is needed [11]. However, the current study focuses on the differences within one School (School of Computing Science) at one University.

As identified more than ten years ago by [4], there is still little literature related to marking dissertations compared to the literature on assessments. The limited research on master's dissertation marking indicates the need for more research in that area. As such, this paper's focus is on investigating master's dissertation marking practices in Computing Science.

II. BACKGROUND

A. Marking practices

Reference [3] investigated academics' perceptions of standards through marking practices. Participants, 12 lecturers, were asked to think aloud as they marked written assignments and then interviewed. In asking the question "What influences staff marking practices and judgements?" the authors investigated the source of tutors' standards, their sense of accountability for their grading judgements, and their use of artefacts. Findings showed no evidence of significant pressure or practice related to lowering of standards. However, bias can emerge because of the differences in tutors' tacit 'standards' frameworks. The authors argued that the "debate about academic standards and marking is very difficult to have with tutors facing a much more intangible picture of standards in use" due to the opacity of concepts and their slippery nature in assessment criteria when interpreted in relation to complex and diverse student work and reduced to a single grade or mark (p. 621).

Reference [12] investigated the reliability of gradings master's theses. They concluded that while a more objective

rubric might raise the marking reliability, this could also affect the creativity and unique nature of theses and lower individual student contribution to a particular discipline. This implies that rigorous and highly structured marking rubrics can be unfair. Moreover, work completed by students with complex disabilities such as "physical impairments on their practical work, and the relatively slower speed at which they were able to assimilate new learning" could contribute to the sense of unfairness in marking judgements, thus leading tutors to reduce their expectations of what they consider fair by classifying student achievement as different rather than inferior and adjusting their interpretation of assessment rubrics in order to widen participation on grading decisions [13].

In their study, [14] found that some early career supervisors inflated their marks, fearing that awarding low marks could affect their careers as it will be seen as a failure to educate and inspire students. Citing literature, [14] list the factors that influence supervisors' marks as the length of student contact, the supervisor wanting to protect the student-teacher relationship and a reluctance to award a failing grade. However, their study also showed there exists an 'internal' marker bias even though the internal markers do not know the students. However, that bias is eliminated following the moderation process due to the key role that subject expertise plays during moderation discussions. This is the case for postgraduate report marking, where both the internal and external assessors did not know the students.

B. Moderation

Moderation, traditionally seen as an agreement on marks between multiple markers, does not assure the accuracy or the best marks and has its limitation in terms of ensuring fair and consistent marking [11]. But moderation is crucial for new staff to learn more about standards in the private, autonomous act of marking as it gives them the opportunity to discuss standards with experienced markers, although these discussions could be affected by the power relationship [3]. Moderation is not without issues. The issues include workload, limited assessment choices, slowed feedback time and limited increase in reliability [11].

In their paper that reviewed common approaches to moderation, [9] found that there are four purposes for moderation directly related to assuring standards: a) to ensure consistency and fairness in standards (equity), b) to create confidence for academics in their grading decisions (justification),c) to align with systemic requirements (accountability), and d) to calibrate judgements and build shared interpretations of criteria and standards (community building). Their investigation revealed that the reasons for moderation that were used the most were justification and accountability.

III. Context of the study

The context of this study was the School of Computing Science at one institution. Due to an increase in the number of students completing their master's degree in computing science (nearly 600 students), academics mark dissertations on topics that are not in their area of research or teaching. Moreover, students can choose to come up with their own projects for which the supervisor might not be an expert. Assessors used two generic marking schemes, one for projects in the Computing Science strand (CS+ – mainly research projects or advanced software development), and one for

TABLE I. MARKING SCHEMES EVALUATIVE CRITERIA OR ASSESSMENT COMPONENTS

IT+	IT+ and CS +	CS+
Development projects	Research-style projects	Any type of CS projects
Analysis 15%	Analysis 15%	Problem Analysis (15%)
Software Product 40%	Research Product 40%	Outcome (Software Product Research Insight) (40%)
Evaluation 15%	Evaluation 15%	Reflection (10%)
Dissertation 20%	Dissertation 20%	Dissertation Quality (25%)
Professional Conduct 10%	Professional Conduct 10%	Professional Conduct (10%)

projects in the Information Technology strand (IT+ - mainly development projects for conversion students) with a double-marking process for fair marking. The marking schemes consist of 5 evaluative criteria or assessment components, as seen in Table 1. However, students' professional conduct is marked only by supervisors. Each criterion consists of sub-criteria for quality descriptors divided into eight quality levels. These levels are A1-A5 Excellent, B1-B3 Very Good, C1-C3 Good, D1-D3 Adequate, E1-E3 Weak, F1-F3 Poor, G1-G2 Very Poor, and H. The main difference between the marking schemes is the percentage of weight for the dissertation. Also, the CS+ students are expected to provide a reflection. The marking schemes also provide instructions for reconciliation and arbitration.

The marking process is not anonymous as both the supervisors and readers can see their names on the project system, and students also include their supervisors' names in the acknowledgement section of the dissertation. Both the students and assessors use the same marking schemes. Students are expected to use it as a formative tool to plan and self-assess their work. Further support is provided through exemplars, another helpful method to support the knowing of standards beyond telling [15] and indicate to students, supervisors, and assessors the quality of expected work.

IV. Methodology

A. Research Questions

This study aimed to investigate master's project dissertations marking practices in Computing Science at one UK university. The following questions frame the study:

- What are academics' master's dissertations marking practices in Computing Science?

- Is there any difference in the perceptions of CS+ and IT+ marking schemes and the marking practices based on dissertations marking load?

- Is there any difference in the perceptions of CS+ and IT+ marking schemes and the marking practices based on dissertation marking experience?

B. Data collection and analysis

A survey was used to get participants' perceptions of master's dissertation marking in Computing Science. The survey consisted of close and open-ended questions. The self-reported instrument also included three scales measuring assessors' perceptions of the CS+ and IT+ marking schemes and their master's dissertation marking practices. The rating questions used Likert scales of 1 to 7 (1 means "strongly disagree" and 7 means "strongly agree"). Convenience sampling was used for the survey. It provides a representative sample (n=31) of staff in the School of Computing Science

(n~80). The survey was sent to all those who marked master's project dissertations in 2020-2021. Quantitative data were analysed in Statistical Package for Social Science (SPSS) version 27 using descriptive statistics and Kruskal-Wallis tests, non-parametric tests to compare three or more groups. The total score of participants for the scale items was used.

The open-ended questions explored participants' feelings about project dissertation marking and questions on marking and moderation. The questions included, "Think of a recent dissertation you have marked. In your own words, please tell us how you went about marking it (Describe your dissertation marking process). Did you follow the same procedure to mark all your dissertations? What were your emotions or your thinking or your immediate reactions to what you were reading?", "If the dissertations you have marked were not in your area of expertise or research field, how did you feel about it?", "What do you think are good MSc dissertation marking practices?", "How do or would you ensure fair marking and equity of MSc dissertations?", "If the number of project dissertations to mark has negatively affected your marking, tell us how marking was affected", "How can your marking experience be improved?", "How can the second marking and reconciliation process be improved?", "How can the current marking schemes (CS+ / IT+) be improved?". Qualitative data from open-ended questions were analysed using Thematic Analysis [16] because of its flexibility in the data collection method, sample size, and analysis. The analysis followed Braun and Clarke's six-phase analytic process of familiarisation with data, code generation, combining related codes into themes, reviewing themes (and recoding, if necessary), naming the themes and subthemes and writing up. As in [17] guidelines, no codebooks and coding frames were used as they do not "cohere with the qualitative sensibility that underpins and shapes our approach" (p. 108). Ethical approval was obtained.

V. FINDINGS

A. Quantitative results

Out of the 31 assessors who self-reported their views, the majority (68%) had between 1 and more than 5 years of dissertation marking experience, while the remaining (32%) did not have any marking MSc dissertation marking experience prior to marking the 2020-2021 cohort (Fig. 1).

Many (40%) marked more than 10 MSc project dissertations. Regardless of their experience level, 83.87% stated that some of the dissertations they marked required negotiation. The number of dissertations requiring negotiation per person varied between 1 and 7 dissertations. While 54.8%, mainly participants who have been in the University longer (5+ years), teach a course that corresponds to their research field, 22.6%, primarily new assessors (up to 3 years), teach a course unrelated to their interest, and 22.6% of assessors do not teach any course. Meanwhile, 34.4% have marked research-based only dissertations, 35.9% have marked development-based only dissertations, 18.8% have marked dissertations that are both research and development-based, and 9.4% of the respondents were not sure of the type of dissertations they have marked. Most assessors used the CS+ and IT+ marking schemes; however, 20% did not read the marking schemes in full, and more than 10% did not even refer to them. Meanwhile, 20% did not read the short form of the marking schemes, nor did they refer to the extra support material provided on Moodle. Some of the reasons include not

Fig. 1. Master's dissertation marking experience

knowing what a short form marking scheme is; the marking scheme is not needed because of years of experience referring to the marking schemes, and because the marking schemes were too long and irrelevant.

Descriptive statistics in Table 2 show that most of the 51.6% who tend to assign grades based on the main dissertation document only have more than five years of experience and with a marking load of 10+ dissertations. However, they only read supplement documents when they struggle to find evidence in the main dissertation document. On the other hand, most of those who read everything before assigning grades (51.6%) have a lower marking load of $1 - 4$ dissertations and no marking experience. Regardless of their marking load and marking experience, many (41.4%) do a search on Google when marking a dissertation outside their expertise, including experienced assessors. More than half of the respondents (58%) would submit each individual dissertation grade after marking it. This practice is mainly done by experienced assessors (5+) regardless of their load. In contrast, 35.5% of participants, primarily those with no experience, would wait until they have completed all markings before uploading all the results as they tend to compare them, leading to modifying some of their markings before submission. However, a few experienced (5+ years) assessors also compare them. Regardless of their marking experience, 28.6% believe that how they grade a project dissertation is affected when the supervisor is the expert. In comparison, 20.7%, even those with 5+ years of marking experience, feel reserved and would award lower grades because they are not experts in the project dissertation topic they are marking; many more in that group, 31%, would award higher grades for the same reason. A few participants (19.5%), mainly those with 5+ years of experience, do not use the marking schemes when grading the projects that they have supervised because they already know them.

A Kruskal-Wallis tests revealed that participants' perceptions of the CS+ ($n = 27$, $p = .617$) and IT+ ($n = 20$, $p = .253$) marking schemes and marking practices ($n = 29$, $p = .790$) were not statistically different across the three MSc dissertation marking load groups (1- 4; $5 - 9$; 10+ dissertations) (Fig. 2). Furthermore, their perceptions on the three scales did not statistically differ (CS+ marking scheme: $n=29$, $p = .514$; IT+ marking scheme: $n= 22$, $p = .205$; Marking practices: $n = 31$, $p = .751$) between those who had no marking experience, 1 to 3 years and more than 5 years of MSc dissertations marking experience (Fig. 3). None of the participants had 4 years of marking experience.
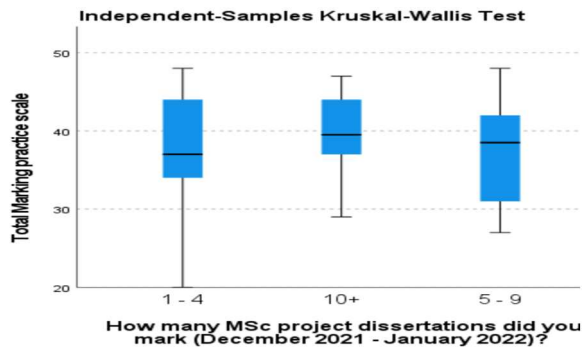
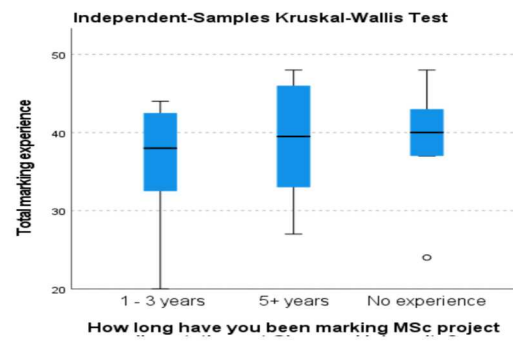Fig. 2. Participants' marking practices based on dissertations marking load

Fig. 3. Participants' marking practices based on marking experience

TABLE II.   MARKING PRACTICES SCALE RATINGS

MSc dissertation marking practice attitude	N	Md	% rating >= 5	Marking load	Year of experience
14.1. I tend to assign grades based on the main dissertation document only.	31	5	51.6%	10+	5+
14.2. Before assigning the grades, I read the main dissertation and supplementary documents (including videos demonstration and URLs).	31	5	51.6%	1-4	No experience
14.3. I only read any supplementary documents in the compressed folder when I struggle to find evidence in the main dissertation document.	29	5	51.7%	10+	5+
14.4. I have had to search on Google when marking a dissertation outside my expertise.	29	3	41.4%	10+	5+
14.5. As soon as I complete marking a project dissertation, I submit its result on LTC.	31	5	58.1%	10+	5+
14.6. When I have more than one project dissertation to mark, I wait until I have completed all the markings before uploading all the results.	31	3	35.5%	5-9, 10+	No experience
14.7. When I mark more than one dissertation, I can't help but compare them, and this has led me to modify some of my markings.	31	3	35.5%	10+	No experience, 5+
14.8. I often feel that how I grade a project dissertation is affected when the project supervisor is the expert.	28	3.5	28.6%	10+	No experience, 5+
14.9. I feel more reserved in marking when I am not an expert in the project dissertation topic. So, I award lower grades.	29	2	20.7%	10+	5+
14.10. I feel I award higher grades because I am not an expert in the project dissertation topic.	29	3	31.0%	10+	5+
14.11. I do not refer to the marking schemes when grading the projects, I have supervised because I already know them.	31	2	19.4%	10+	5+

Table 3 and 4 present participants' perceptions of the CS+ and IT+ marking schemes, respectively. Overall, while 51.7% of participants thought that the CS+ marking scheme was very useful, only 34.5% believe it ensures fairness in marking and makes it easy. The percentage of those who struggle with marking is higher when marking IT+ dissertation. Indeed, only 40.9% thought the IT+ marking scheme was useful. Only 30.8% (CS+) and 40.9% (IT) had relied on the students' context when markings project dissertations they did not supervise.

As shown in Table 5, many participants (40%) agreed that the number of dissertations negatively affected their marking practices. While only a few participants thought that knowing who the student project supervisor (19.4%) is or the supervisor's seniority (12.9%) would not affect their marking practices, more than half (53.3%) would rather not know who supervised the projects they are marking. A third of the participants had an issue with marking dissertations outside their area of expertise. Results also show support for double marking, with only 16.1% finding it tolerable to use one marker per dissertation. Two-thirds of participants (64%) had to third mark a project, and their comments highlighted that most agreed more with the readers.

B. Qualitative results

The survey included many open-ended questions which revealed three overarching themes: 1) I felt that I betrayed my ethical principles, 2) I tried to acknowledge my bias/familiarity with the topic, and 3) Improving marking experience, second marking and reconciliation process.

1)   I felt that I betrayed my ethical principles
I have been allocated too many dissertations. The marking load is the main issue that affects marking practices. Participants indicated that marking becomes very tiring, with less time spent on marking and reading in detail. This is exacerbated when the dissertation topic is not in the marker's teaching or research area, which, in this case, also affects the marker's understanding of the dissertation, leading to the extra material not being scrutinised to the same extent. However, the lack of time due to having a big marking load also prevents the supervisors to scrutinise extra material from their own students' projects.

Human nature is to play marking like this as a two-player game. Marking practices can be affected when the project supervisor is a senior staff member. This is especially the case for new assessors, who would then be stricter (Participant 687) in some cases and could feel unable to mark (Participant 144) in other cases. Knowing who the supervisor is can also impact

the marking process regardless of the assessors' marking experience. Most assessors would rather not know who supervised the dissertations they are requested to second or third mark. Indeed, knowing the supervisor lets assessors identify the marking patterns of the other assessors (Participant 030) and the "difficult characters" and adapt marking to the supervisor's character, thus fostering the reader's bias.

"... I began to expect when I saw their name that there would be a big difference in the marks. If I knew while marking that they were the supervisor I'd be worried that I'd be extra generous to make up for my perception that they are extra harsh." (Participant 030)

This leads to marking being done more defensively with more justification, as explained by Participant 291, "In practice, I know almost all of them, so I know who the difficult characters are. So, I will mark more defensively for these, with more justification." This seems to resonate with Participant 081, who mentions the "psychological biases" that would affect marking in that case. At the same time, Participant 687 notes the unfairness that results as sometimes they try to imagine how the supervisor would mark. Assessors' changes in marking practices are, therefore, a result of assessor's subjectivity which invites them to adapt marking based on the supervisor's marking habits:

"As marking projects is SO subjective if I knew the person was a harsh marker, I would probably increase my score, the reverse also holds. I think human nature is to play marking like this as a two-player game where the overall score is the one you want, probably that is bad though!" (Participant 861)

Many markers would prefer project supervisor anonymity so they can concentrate only on the produced work (Participants 209) and be objective as they believe objectivity would exclude considerations for negotiation discussions (Participants 291, 861, 144). Indeed, marking is also affected during negotiations when the other assessors have "strong motivation for certain scores" (Participant 013) and "famously defend the grades they award" (Participant 206), implying some power struggle between the reader and the supervisor, where the supervisor always wins. This is even more important when the assessor is new to the University and does not want to take "risks" and therefore conforms to "existing practices" (Participant 171).

I don't want to look like I don't know how to mark properly. The lack of subject expertise also influences assessors' preference for anonymity when marking. This is to avoid being, as Participant 465 wrote, "tempted" to be over-generous with marks. Indeed, that unfamiliarity with the topic and supervisors' advantage of knowing their students' context make non-supervisor assessors worried about their peers' perceptions of them. The argument is that since the reader does not have a clear picture of the student's context, they may not provide an accurate grade. This may make them look like they do not understand the contribution, or they do not know how to mark properly. Participant 687 explains,

"I think the supervisor has a better picture of the performance of the student, hence awards a rather accurate grade; in that case, especially if we know each other, I don't want to look like I didn't understand properly the contributions of the dissertation or don't know how to mark properly." (Participant 687)

As a result, non-supervisors mark more "carefully" because of who they know and how they may appear to those supervisors as they do not want to "feel observed," as mentioned in this comment,

"I want to not feel observed by that person when I grade. It would feel more of a fair process for the student." (Participant 081).

Marking disagreement can also have a deeper effect on project supervisors. For example, Participant 291, supervisor and experienced marker, portrayed the negotiation with other assessors as a betrayal of ethical principles, as both the reader and third marker awarded higher undeserved grades to a student even though they were not familiar with the topic. The experienced supervisor had explained,

"... They were motivated by the process aspect, and both were *highly* unfamiliar with the topic ... I felt that I betrayed my ethical principles..." (Participant 291).

I found marking really stressful. Assessors' well-being is affected by dissertation marking. Many have mentioned being stressed, tired, worried, disappointed and even resentful of the time that must be committed to marking. For example, stress also emerges from the subjectivity of marking, because similar projects could get different marks causing further distress to the assessors. Moreover, stress and anxiety can be caused by the intercultural difference when marking MSc dissertations for the first time at a new institution. New assessors come with a background of quality and marking standards they gained through their previous institutions' marking practices. These differences are more pronounced when coming from different countries where the academic culture differs from the current country.

"Marking itself is a bit more difficult, perhaps because being new at [removed] and in the UK in general, I'm not too familiar with the students' level and with the expectations on dissertations... Reconciling these different expectations and intercultural differences takes a bit of time." (Participant 171)

2) I tried to acknowledge my bias/familiarity with the topic

Many participants recognised that they are biased toward their own student's dissertation because they are familiar with the topic as they are "'committed' to the work and/or to the student" (Participant 206). Supervisor bias, which is not always in favour of the student, is also a source of concern during moderation. However, their deepest interest in the topic influences their marking practices the most; otherwise, it would be "only a chore". They are also motivated by the quality of the project dissertation. The marker's own background is also a strong motivation. For example, Participant 171 wrote, "My non-UK background influences my dissertation marking practice; I may not have the same expectations as others." (Participant 171). For some participants, their marking process is dependent on their role. As supervisors, they would skim the documents because they have already reviewed them (Participant 206) and would mark "more with the knowledge of the actual product in mind" (Participant 291). Indeed, it can be very difficult for a supervisor to "forget" about what happened during supervision and student work and focus exclusively on what their student wrote, as requested by the marking guidelines. For example, Participant 030 thinks it would be "naïve" to imagine they can make sure they are always equally fair. This

TABLE III. CS MARKING SCHEME SCALE RATINGS

Perceptions of CS marking scheme	N	Md	% rating >= 5	Marking load	Year of experience
11.1. The current marking scheme is very useful.	29	5	51.7%	10+	5+
11.2. Without the marking scheme, I would not manage to mark accurately.	29	4	44.8%	10+	5+
11.3. I understand the difference between development and research-based project dissertation.	29	5	72.4%	10+	5+
11.4. The marking scheme document ensures fairness in marking.	29	4	34.5%	10+	5+
11.5. The marking scheme document makes my marking easy.	29	4	37.9%	1 - 4	5+
11.6. The current marking scheme document leaves too much to interpretation.	29	4	44.8%	10+	5+
11.7. The current marking scheme document helps me mark MSc project dissertation in different fields effectively.	29	4	37.9%	10+	5+
11.8. The current marking scheme helps me mark MSc project dissertation in my teaching/research field effectively.	29	4	44.8%	10+	5+
11.9. The marking scheme document is not flexible enough.	29	3	27.6%	10+	1-3
11.10. I rely on my experience in marking rather than the marking scheme.	29	4	37.9%	10+	5+
11.11. I rely on the context when marking dissertations I did not supervise.	26	4	30.8%	10+	5+
11.12. I have struggled to mark some dissertations using the marking scheme.	29	3	27.6%	10+	5+
11.13. I struggle to make use of the marking scheme when marking a research-based MSc project dissertation.	28	3.5	28.6%	10+	5+
11.14. I struggle to make use of the marking scheme when marking a development-based MSc project dissertation.	28	3	25.0%	10+	5+, 1-3
11.15. I prefer bidding to be a reader of a dissertation in my field to ensure I do not mark dissertations outside my expertise.	28	5	60.7%	10+	5+ no experience
11.16. I believe research-based project dissertations and development-based project dissertations should not have the same marking scheme.	29	4	44.8%	1 - 4	5+

TABLE IV. IT MARKING SCHEME SCALE RATINGS

Perceptions of IT marking scheme	N	Md	% rating >= 5	Marking load	Year of experience
13.1. The current marking scheme is very useful.	22	4	40.9%	10+	5+
13.2. Without the marking scheme, I would not manage to mark accurately.	22	3	27.3%	10+	5+
13.3. I understand the difference between development and research-based project dissertation.	22	5.5	77.3%	1-4, 5-9	5+
13.4. The marking scheme document ensures fairness in marking.	22	4	31.8%	10+	5+
13.5. The marking scheme document makes my marking easy.	22	4	40.9%	10+	5+
13.6. The current marking scheme document leaves too much to interpretation.	22	4.5	50.0%	5-9	No experience, 5+
13.7. The current marking scheme document helps me mark MSc project dissertation in different fields effectively.	22	4	27.3%	10+	5+
13.8. The current marking scheme helps me mark MSc project dissertation in my teaching/research field effectively.	22	4	36.4%	10+	5+
13.9. The marking scheme document is not flexible enough.	22	3	31.8%	10+	1-3
13.10. I rely on my experience in marking rather than the marking scheme.	22	4.5	50.0%	5-9,10+	5+
13.11. I rely on the context when marking dissertations I did not supervise.	22	4	40.9%	10+	5+
13.12. I have struggled to mark some dissertations using the marking scheme.	22	4	45.5%	5-9,10+	5+
13.13. I struggle to make use of the marking scheme when marking a research-based MSc project dissertation.	22	3	31.8%	10+	1-3, 5+
13.14. I struggle to make use of the marking scheme when marking a development-based MSc project dissertation.	22	3	22.7%	10+	1-3
13.15. I prefer bidding to be a reader of a dissertation in my field to ensure I do not mark dissertations outside my expertise.	22	4.5	50.0%	10+	5+
13.16. I believe research-based project dissertations and development-based project dissertations should not have the same marking scheme.	22	5	54.5%	10+	5+

reinforces the issue of the supervisor's unconscious or conscious bias when marking student projects, which is one of the causes of unfairness.

> *"It is a bit hard to only take the dissertation as basis when marking your own students, as you know how the project went. I tried very hard to forget that and concentrate on the dissertation, being aware of the bias I have by knowing the student." (Participant 405)*

Many respondents, experienced and inexperienced assessors, think that the marking schemes are vague or lack clarity as they typically contain variants of the same sentences with some synonym for "good" and "bad" changed. They should be detailed enough to clarify what the product is; where the user evaluation goes, especially if it is a research-only project, explicitly referring to the design and implementation aspects as it is unclear which marking category they belong to. Unfairness in dissertation marking is also caused by the disparity in complexity and requirements among dissertations, which are not addressed in the marking schemes. Other aspects that are not captured in the marking schemes are the level of project difficulty and student contribution. For

instance, as commented by Participant 956, an IT$^+$ student could get an A in an IT project that should only be five credits worth of work, which would not be the case had the project been appropriately scoped.

3) Improving marking experience, second marking and reconciliation process

In Computing Science, there are far too many fields even within the same research group, which means that many assessors end up second or third-marking project dissertations that are entirely outside their research area. It explains why, when asked what could improve their marking experiences, many participants suggested marking allocation be done within the assessors' areas of expertise, supported with exemplars and case studies in the marking schemes, and especially, the provision of mentoring for inexperienced assessors. There should be consistency in documenting the negotiation process, although second marking and reconciliation can seem less worthwhile for some assessors because it is time-consuming. For example, Participant 461 think that solving the issues of mismatch in topic expertise instead, which leads to good projects being awarded lower grades, would be more beneficial:

"There is no need for second marking. It is extra work for everyone and does not yield much benefit. If it makes any difference, it'll probably be that good projects are given lower grades because of the mismatch in topic expertise." (Participant 461).

But overwhelmingly, the marking experience will improve with a reduction of marking load or not being done at the same time as exam marking and assigning more marking time. Other good practices include asking students to provide interactive information, such as a video demonstration of their work as "It is sometimes hard to know from looking into the report alone" (Participant 081), spending a long time reading and trying out the project. Blind double and blind third markings and blind negotiation are some of the good practices that would help ensure "excellent" arbitration practices. One participant also suggested marking master's dissertations using the same rigour as top conference and journal reviews. Finally, looking after assessors' well-being is important. Some mentioned taking regular breaks and not marking when tired.

VI. DISCUSSION AND CONCLUSIONS

This study investigated master's project dissertations marking practices in Computing Science in a higher education institution. It is part of a bigger project that aims to investigate fairness in MSc project dissertation marking in Computing Science and, through reflection, to develop a conceptual framework aiding the provision of fair marking and grading of MSc projects dissertations. The focus is on postgraduate master's education because it is an under-researched area.

Findings showed that participants' marking practices were not statistically different despite having different marking loads and marking experiences. Furthermore, despite marking schemes being critical to ensuring fairness in marking and grading [5, 18], they are not being used by some. One of the main reasons for that is their vagueness. That sense of unhelpfulness of the marking schemes is increased when marking project dissertations outside one's research and teaching area, making markers feel "uncomfortable" and lack confidence in their judgement. The vagueness of marking schemes can affect reliability because the design of the marking rubric is one of the factors that affect marking reliability [19]. Therefore, there is a need to design detailed, clear and calibrated marking schemes and supporting material while considering a) the disparity in the complexity of the various types of project dissertations, b) the student contribution, c) intercultural differences of assessors, and d) the level of the marker's subject knowledge, which are all often sources of disagreement between assessors. These marking schemes should be used in conjunction with exemplars. Studies have found that rubrics or marking criteria, or marking schemes scaffold or as pedagogical tools, support the development of evaluative judgement, while exemplars give students (and assessors) a notion of quality and an opportunity to exercise their evaluative judgement [20]. Exemplars also 'anchor' the criteria to a community of practice [21].

Many participants also highlighted their desire for complete anonymity during the marking process. They would prefer double and third marking, including at the negotiation stage, to be blind to avoid feeling "observed" by the supervisors and other assessors, appearing to others as if they lack marking skills, to avoid the power struggle and maintain objectivity in marking. This would also help second and third assessors not succumb to the pressure upon them that is caused by the supervisor's bias. Indeed, supervisor bias is one of the factors affecting fairness and equity in marking practices [14, 22]. During marking, the supervisors' knowledge of their students' background put the second assessors in an unfair position as these would mark based on what was presented to them. This is even more noticeable when they assess dissertations on unfamiliar topics. However, the second assessors' most obvious unfair practice is their strategic approach to marking when they know who the supervisors are.

TABLE V. MARKING AND MODERATION ISSUES

Marking and moderation issues	Yes/True	No/False
The number of project dissertations to mark has negatively affected my marking.	12 (40%)	18 (40%)
I feel that how I mark a project dissertation is affected when the student project supervisor is my senior.	4 (12.9%)	27 (87.1%)
I feel that how I grade a project dissertation is affected when I know who the student project supervisor is.	6 (19.4%)	25 (80.6%)
As a reader or third marker, I would rather NOT know who supervised it.	16 (53.3%)	14 (46.7%)
Have you ever had any issues with marking a dissertation because it is not in your area of expertise?	10 (33.3%)	19 (63.3%)
Have you ever had any issues with project dissertation moderation where you are the primary marker (e.g: supervisor of the project)?	6 (20%)	24 (80%)
Have you ever had any issues with project dissertation moderation where you are the secondary marker (e.g: reader)?	6 (20%)	24 (80%)
Have you ever been asked to be the third marker of an MSc project dissertation?	20 (64%)	11 (35.5%)
Would it be tolerable to use only one marker per dissertation?	5 (16.1%)	26 (83.9%)

Indeed, the second assessors would adapt their marking, which would no longer reflect the objectivity required to ensure fairness in marking. This could be why many participants would prefer complete anonymity in the marking process to ensure they are not 'observed'. However, anonymity does not always ensure fairness [23]. Findings also show that their current institutions' standards can inform assessors' marking standards. But they can also be influenced by the intercultural academic difference, thus, agreeing with the literature that assessors "are bounded by their social and cultural environment and expectations" (p. 647) [9].

In summary, there are four key points to consider in order to improve the marking experience and reconciliation process:

- Reduction of marking load or/and assigning more marking time. However, being given more time for project markings means exam board deadlines may not be met or that time to complete the project is reduced.

- Assigning marking within assessors' knowledge or research area. This can be challenging when there are not enough assessors.

- Provide detailed, clear and calibrated marking schemes. Improvement of the context of CS$^+$ could come from deeper consideration of the marking scheme and what markers felt was 'unclear' or how it could be clearer.

- Complete anonymity of the marking process (supported by technology). While the supervisor may know their own student projects, they should not know who is second or third marking them. And those other assessors should not see the students' and supervisors' names. However, this is difficult to achieve in practice, particularly in smaller institutions, as most markers would either directly know who had supervised a project or who is currently working in that area.

The main limitation is that the study was conducted at one institution. Further research should investigate master's dissertation marking practices in other computing science departments at other universities in the UK and internationally. It should also consider including non-CS schools in the University to get an overall view of their marking and moderation practices. Further, data were collected using a self-reported questionnaire which provided quantitative and qualitative data from open-ended questions. Future work should consider using interviews as the qualitative data collection method to understand better assessors' dissertation marking practices and experiences. The author will follow up this study with interviews of willing participants identified in this survey for further probing, including understanding the impact of having less time on the marking, whether participants' practices change, and whether they spend less time looking at the code as a result.

REFERENCES

[1] Quality Assurance Agency. Characteristics Statement - Master's Degree. 2020.

[2] M. Bikanga Ada. "Using MyFeedBack, a mobile web 2.0 system, to help students engage with their feedback - a case study at a Scottish university", *EDULEARN14 Proceedings*, pp. 4910-4919.

[3] S. Bloxham and P. Boyd. "Accountability in grading student work: securing academic standards in a twenty-first century quality assurance context," *British Educational Research Journal*, vol. 38, no. 4, pp. 615-634, 2012.

[4] J. Bettany - Saltikov, S. Kilinc and K. Stow. "Bones, boys, bombs and booze: an exploratory study of the reliability of marking dissertations across disciplines," *Assessment & Evaluation in Higher Education*, vol. 34, no. 6, pp. 621-639, 2009.

[5] D. Man, Y. Xu, M. H. Chau, J. M. O'Toole, and K. Shunmugam. "Assessment feedback in examiner reports on master's dissertations in translation studies," *Studies in Educational Evaluation*, vol. 64, 2020, https://doi.org/10.1016/j.stueduc.2019.100823.

[6] S. M. Brookhart and F. Chen. "The quality and effectiveness of descriptive rubrics," *Educational Review*, vol. 67, no. 3, pp. 343-368, 2015.

[7] Y. M. Reddy and H. Andrade. "A review of rubric use in higher education," *Assessment & Evaluation in Higher Education*, vol. 35, no. 4, pp. 435-448, 2010.

[8] R. D. Tierney. "Fairness as a multifaceted quality in classroom assessment," *Studies in Educational Evaluation*, vol. 43, pp. 55-69, 2014, https://doi.org/10.1016/j.stueduc.2013.12.003.

[9] S. Bloxham, C. Hughes, and L. Adie. "What's the point of moderation? A discussion of the purposes achieved through contemporary moderation practices," *Assessment & Evaluation in Higher Education*, vol. 41, no. 4, pp. 638-653, 2015.

[10] C. Rust. "Towards a scholarship of assessment," *Assessment & Evaluation in Higher Education*, vol. 32, no. 2, pp. 229-237, 2007.

[11] S. Bloxham. "Marking and moderation in the UK: false assumptions and wasted resources," *Assessment & Evaluation in Higher Education*, vol. 34, no. 2, pp. 209-220, 2009.

[12] L. Williams and S. Kemp. "Independent assessors of master's theses show low levels of agreement," *Assessment & Evaluation in Higher Education*, vol. 44, no. 5, pp. 764-771, 2019.

[13] M. Ashworth, S. Bloxham and L. Pearce "Examining the tension between academic standards and inclusion for disabled students: the impact on marking of individual academics' frameworks for assessment," *Studies in Higher Education*, vol. 35, no. 2, pp. 209-223, 2010.

[14] R. McQuade, S. Kometa, J. Brown, D. Bevitt and J. Hall. "Research project assessments and supervisor marking: maintaining academic rigour through robust reconciliation processes," *Assessment & Evaluation in Higher Education*, vol. 45, no. 8, pp. 1181-1191, 2020.

[15] D. C. Carless, K. K. H. Chan, J. To, M. Lo, and E. Barrett. "Developing students' capacities for evaluative judgement through analysing exemplars," In *Developing Evaluative Judgement in Higher Education: Assessment for knowing and producing quality work*, D. Boud, R. Ajjawi, P. Dawson & J. Tai, Eds, London: Routledge, 2018.

[16] V. Braun and V. Clarke. "Reflecting on reflexive thematic analysis. Qualitative Research in Sport," *Exercise and Health*, vol. 11, no. 4, pp. 589-597, 2019.

[17] V. Clarke and V. Braun. "Using thematic analysis in counselling and psychotherapy research: A critical reflection," *Counselling Psychotherapy Research*, vol. 18, no. 2, pp. 107 – 110, 2018.

[18] M. Bearman and R. Ajjawi. 'From "Seeing Through" to "Seeing With": Assessment Criteria and the Myths of Transparency,' *Frontiers in Education,* vol. 3, 2018, doi:10.3389/feduc.2018.00096.

[19] J. Tisi, G. Whitehouse, S. Maughan and N. Burdett. "*A Review of Literature on Marking Reliability Research (Report for Ofqual)",* Slough: NFER, 2013.

[20] J. Tai, R. Ajjawi, D. Boud, P. Dawson and E. Panadero. "Developing evaluative judgement: enabling students to make decisions about the quality of work," *Higher Education*, vol. 76, no. 3, 467-481, 2017.

[21] A. Jönsson and F. Prins. "Editorial: Transparency in Assessment— Exploring the Influence of Explicit Assessment Criteria," *Frontiers in Education,* vol. 3, 2019, doi: 10.3389/feduc.2018.00119.

[22] M. K. Russell and P. W. Airasian. "*Classroom Assessment: Concepts and Applications,*" McGraw-Hill Education, 2012.

[23] E. Pitt and N. Winstone. (2018) "The impact of anonymous marking on students' perceptions of fairness, feedback and relationships with lecturers," *Assessment & Evaluation in Higher Education*, vol. 43, no. 7, pp. 1183-1193, 2018.