Document generated on 04/10/2025 5:41 a.m.

The Canadian Journal of Action Research

Exploring Anonymous Marking to Mitigate Marking Bias: A Self-Study Through Mixed Methods Action Research

Len Chan

Volume 25, Number 1, 2025

URI: https://id.erudit.org/iderudit/1117180ar
DOI: https://doi.org/10.33524/cjar.v25i1.685

See table of contents

Publisher(s)

Canadian Association of Action Research in Education

ISSN

1925-7147 (digital)

Explore this journal

Cite this article

Chan, L. (2025). Exploring Anonymous Marking to Mitigate Marking Bias: A Self-Study Through Mixed Methods Action Research. *The Canadian Journal of Action Research*, *25*(1), 113–135. https://doi.org/10.33524/cjar.v25i1.685

Article abstract

Anonymous marking, as a means to mitigate bias in grading, involves evaluating student work with their identities withheld. Anonymous marking is explored in this self-study to mitigate implicit bias, which negated a practitioner-researcher's educational values. The mixed methods action research findings show withholding student identities during grading alleviates confirmation bias and the halo effect. Despite a short period of adjustment, anonymous marking promotes objectivity and fosters more consistent feedback. However, it prevents personalized feedback, jeopardizes relationship building, and undermines the detection of contract cheating. Moreover, anonymity cannot avert affectual influences and is impracticable for scaffolded formative assessments requiring follow-up feedback. Overall, anonymous marking is shown to be but one measure to counter assessment bias; strategies to mitigate bias unrelated to student identities need to be explored. This self-study also helped the author better understand her role as a practitioner and researcher, enabling her to generate her living-theory.

All Rights Reserved © The Canadian Journal of Action Research, 2025

This document is protected by copyright law. Use of the services of Érudit (including reproduction) is subject to its terms and conditions, which can be viewed online.

https://apropos.erudit.org/en/users/policy-on-use/

This article is disseminated and preserved by Érudit.

Érudit is a non-profit inter-university consortium of the Université de Montréal, Université Laval, and the Université du Québec à Montréal. Its mission is to promote and disseminate research.

https://www.erudit.org/en/

CJAR

Canadian Journal of Action Research
Volume 25, Issue 1, 2025, pages 113-135

EXPLORING ANONYMOUS MARKING TO MITIGATE MARKING BIAS: A SELF-STUDY THROUGH MIXED METHODS ACTION RESEARCH

Len Chan
Conestoga College

ABSTRACT

Anonymous marking, as a means to mitigate bias in grading, involves evaluating student work with their identities withheld. Anonymous marking is explored in this self-study to mitigate implicit bias, which negated a practitioner-researcher's educational values. The mixed methods action research findings show withholding student identities during grading alleviates confirmation bias and the halo effect. Despite a short period of adjustment, anonymous marking promotes objectivity and fosters more consistent feedback. However, it prevents personalized feedback, jeopardizes relationship building, and undermines the detection of contract cheating. Moreover, anonymity cannot avert affectual influences and is impracticable for scaffolded formative assessments requiring follow-up feedback. Overall, anonymous marking is shown to be but one measure to counter assessment bias; strategies to mitigate bias unrelated to student identities need to be explored. This self-study also helped the author better understand her role as a practitioner and researcher, enabling her to generate her living-theory.

KEY WORDS: Anonymous grading; Anonymous marking; Blind grading; Confirmation bias; Feedback bias; Halo effect; Implicit bias

INTRODUCTION

In 2008, the UK's National Union of Students (NUS, n.d.) launched the "Mark My Words, Not My Name" campaign to call for all post-secondary institutions to implement anonymous marking. Students had voiced their concerns about discrimination in marking practices and

believed their work would be graded differently if their identities were hidden. This perceived unfairness is not unfounded, as research has shown that faculty's awareness of student identities can lead to marking bias related to gender (Brennan, 2008; Kiekkas et al., 2016), ethnicity (Bygren, 2020; Lindsey & Crusan, 2011), appearance, and knowledge of previous performances (MacDougall et al., 2008; Malouff et al., 2014). Faculty can be affected by implicit bias—"the attitudes or stereotypes that affect our understanding, actions, and decisions in an unconscious manner" (Staats et al., 2017, p. 10)—which results in actions that do not align with the faculty's explicit intentions and educational values.

Anonymous marking (AM) involves evaluating student work with their identities withheld. When the AM feature is turned on in the learning management system (LMS), each submission is coded, e.g., Anonymous User 10. According to the LMS Brightspace developer, the anonymizing tool serves to "avoid unconscious bias [implicit bias] in the grading and feedback process… [so that learners] can be assured that their submissions are assessed fairly (*About Anonymous Marking*, 2022, para. 1). However, research evidence on AM's effectiveness in mitigating bias has been inconclusive (Sharp & Zhu, 2020). While some studies report AM's benefits (Kiekkas et al., 2016; Malouff et al., 2014), some reveal no significant impact on grade outcome and feedback (Batten et al., 2013; Owen et al., 2010), and some others highlight unfavourable effects on student learning and teacher-student relationships (Pitt & Winstone, 2018; Whitelegg, 2002).

Considering that implicit bias is pervasive and activated involuntarily in my teaching practice, I was prompted to explore AM to ensure objectivity and consistency in my assessments. This research aimed to examine how marking bias could be mitigated in my teaching context using AM. This paper begins with a literature review on assessment bias and the impact of AM, followed by the rationale and context for the self-study and the methodology. Next, the mixed methods action research stages are discussed: diagnosing, reconnaissance, planning, acting, and evaluation. After that, quantitative and qualitative data are analyzed and then reflected on in tandem. Finally, the implications, the research limitations, and the steps in a future research cycle are discussed.

LITERATURE REVIEW

Assessment Bias

Assessment bias in higher education leads to unequal learner outcomes. A meta-analysis of 20 experimental studies shows that having information about students while grading leads to significant grade differences between students in biased conditions (viz. ethnicity, educational deficiencies, physical appearance, and prior performance) and those who are not (Malouff & Thorsteinsson, 2016). In one experiment (Malouff et al., 2014), assessors across disciplines awarded higher grades to the written work of students who performed well in a previous, unrelated assessment and lower grades to students with prior poor performance. This finding demonstrates the halo effect; the tendency for an earlier impression formed in one area to influence the judgements of other unrelated areas. MacDougall et al. (2008) also report halo effects in project supervisors' grading of their fourth-year medical students' research reports. Comparing the grades from blind, second

marking of those reports, it was evident that knowledge of student performance influenced the supervisors' grading despite the use of detailed rubric descriptors.

Forgas (2011) shows that students' appearance can also impact assessment outcomes and a marker's mood can affect the direction of the halo effect. First, a philosophical essay with a picture of a middle-aged, bespectacled man attached was graded more positively than an essay with a photo of a young woman—impressions formed from the target photos led to markers' judgmental bias (halo effect). Second, following a mood-induction exercise, markers in a negative mood were observed to be more vigilant and critical in their marking. In contrast, markers in a positive mood were less vigilant and more lenient (mood-congruency effect). Interestingly, positive mood increased the halo effect, while negative mood eliminated it. In another study (Brackett et al., 2013), markers recalled a sad or happy life event before they assessed narrative essays. Those with happy memories gave higher grades, and the opposite occurred with those who recalled sad memories (emotion-congruent effect); notably, a greater halo effect was observed among experienced markers.

O'Hagan and Wigglesworth (2015) report that non-native English speaking (NNES) students received an average of 11.3% lower on their argumentative research essays than their English speaking counterparts. This discovery, along with a lower degree of grade variability among NNES essays and assessor feedback comments about grammar accuracy, implies that "language (where non-native grammatical errors are salient) may obfuscate assessor judgements of content" (p. 1743). Conversely, faculty in another study demonstrated sympathetic bias when they graded composition class essays with fabricated names and nationalities. Although the essays were written by native-speaking Americans, scripts with "international" names were consistently rated higher when scored holistically but lower analytically. The holistic scores and focus group comments reveal that the faculty's recognition of challenges faced by "international" writers led to their tolerance of these students' errors that they would not accept from native-speaking students (Lindsey & Crusan, 2011). In a similar experiment involving marking scripts with fictitious names, Chowdhury et al. (2020) observed that Chinese names and adopter names (a White first name and a Chinese last name), in comparison to White names, were less likely to be awarded a score just above the pass/fail grade threshold. The researchers point out that discrimination can occur "where it matters" (p. 19). Bias can also occur in feedback provision, although its ramifications are less noticeable than bias in grading. Batey (2018) found that essays in an undergraduate research methods course with non-white student names received more checkmarks than the ones with White names, which received more comments. The former also received less educative and helpful feedback compared to the latter. Moreover, essays with white female names received less positive and more negative feedback than those with white male names. However, there were fewer gender differences seen in feedback for non-white student names; Batey (2018) suggests the examiners had difficulty identifying gender in those names.

Impact of Anonymous Marking

A literature review on the impact of AM in post-secondary written assessments across disciplines shows that empirical findings of AM reducing marking bias are equivocal. Sharp

and Zhu (2020), from their review of 33 seminal works, posit the inconclusiveness and inconsistency of AM's effectiveness. Although some studies report no significant impact on grade outcomes (Owen et al., 2010) or on outcomes in relation to the knowledge of prior student performance (Batten et al., 2013) and demographic characteristics such as ethnicity, gender, and socio-environmental factors (Hinton & Higson, 2017; Pitt & Winstone, 2018; Shay & Jones, 2008), other studies attest to the value of AM.

Malouff et al. (2014) claim that AM can mitigate the halo effect. In their experiment, 159 academics across disciplines from tertiary institutions in Australia and New Zealand, were assigned to three groups to induce positive or negative impressions of student ability, effort, and appearance: (1) a group who watched a student's well-performed oral presentation, (2) a group who watched a poorly performed presentation by the same student, and (3) a group who did not watch a presentation. The participants then assessed a paragraph (on psychology) written by the student they watched (or did not watch). Results show that significantly higher scores were given to paragraphs associated with a good presentation and a lower grade to those with a poor presentation. Meanwhile, markers who did not watch an oral presentation gave intermediate scores. The results indicate that positive and negative halo effects occurred among both non-psychology and psychology markers and that having no prior knowledge of the students reduced this effect.

The introduction of AM at a UK university alleviated concerns of gender stereotype bias and led to a remarkable summative assessment outcome in the Faculty of Arts. In the data collected over five years before AM, first-class honours were awarded to 35% of female and 42% of male students. However, in the year AM was implemented, 47% of women achieved first-class honours, whereas the ratio for men remained at 42% (Belsey, 1988, as cited in Brennan, 2008). Despite the argument that students' gender could still be identified through their writing, AM prevented examiners from guessing genders before they read the examination script. In a crossover study at a Greek university, examiners anonymously re-marked their undergraduate nursing students' written examination scripts (n=400) after a washout period of two months. The results show a significant drop in the female students' grades (p> 0.001) but not in the male students' grades, implying examiners' bias in favour of female nursing students (Kiekkas et al., 2016).

AM has also been found to reduce "reverse" bias toward students with "foreign" names. A large-scale comparison of grades (n=25,077) before and after AM implementation was conducted at a Swedish university, involving undergraduate law, economics, political science, and sociology courses. A significant decrease (6%) in the average passing grade of students with non-Swedish names was observed, suggesting that examiners demonstrated weak positive bias toward those students when marking non-anonymously (Bygren, 2020). Another Swedish study (Huskanovic & Adem Nur, 2020) also reports a reduction of "reverse" bias in the anonymized assessment of foreign-named students' macroeconomics exam essays. The researchers propose that since the essays were written in Swedish, examiners may have been sympathetic toward the language ability of students with foreign names; these findings align with Lindsey & Crusan's (2011) bias study (see previous section).

Studies report that AM has no apparent or negative impact on written feedback. While Batten et al. (2013) claim no noticeable difference in feedback comments, another study (Pitt & Winstone, 2018) revealed that students perceived feedback from AM as less effective than personalized feedback in terms of clarification, improving learning, and motivation for better future performance. In anonymized feedback, students' previous performance could not be referenced, nor could emotional sensitivity to individual students be applied. Likewise, Whitelegg (2002) emphasizes that AM disrupts the feedback loop, i.e., faculty is prevented from providing feedback effectively as they cannot consider their students' progress or be able to foster students' self-esteem by way of personal remarks. Furthermore, AM places the onus on the weaker students to come forward for help. It also increases the distance between faculty and students since relationships cannot be developed with non-personal feedback.

Documentation of faculty's lived experience with AM is limited. In her blog post, Novaes (2013), a philosophy faculty member from the Netherlands, recounts an interesting observation with AM—how her "mind was always desperately trying to figure out who wrote each paper" (para. 5). She pointed out that marking identified work was comparatively less time- and energy-consuming because she could use "preexisting ideas about each student and would thus be able to skip a couple of passages here and there" (Novaes, 2013, para. 5). Nursing faculty McDaniel (1994) remarks on her surprise with AM when a seemingly poor student achieved an A, and a perceived A student got a B. She recalls an incident using AM in which a less-than-average student, who sat at the front and smiled constantly in class, purposely wrote their name in the hope that making their identity known could bring up their grade. According to McDaniel (1994), AM prevented manipulation and protected both her and the student. Scott (1995) expresses freedom from "expectations, biases, guilt, sympathy, or pique" (p. 214) when using AM on her students' literature essays and remembers her astonishment over outcomes where a student scored an A on one paper and a C on the next. This occurrence, according to her, rarely happened before AM. She commented that she felt "lighter, freer, as if a burden had been lifted" (p. 214), unencumbered by the knowledge that "this student needs to transfer, that student's mother is dying, and the other one tries so hard" (p. 215). With AM, she also experienced less tension with grade-conscious students who commented that they felt judged solely on their work, not their persons. However, Scott (1995) highlights that at the end of a semester of AM, she sensed a lack of bonding with her students, unlike the closer relationship she experienced before AM. Also, AM could not wholly avert her prejudices for or against differing ideas and writing styles in the essays she graded.

When I first discovered AM in a faculty workshop on grading practices, I attempted it in one assignment. I was surprised by how different the marking experience felt when students' identities were withheld. This highlighted the likelihood of implicit bias in my marking practice. I realized there was a 'contradiction' between my espoused value of fairness and my conveyed actions. Therefore, I was prompted to explore AM as an intervention.

This study addresses marking bias in my teaching context and contributes to the current AM research in a couple of ways. First, unlike previous studies (experiments, surveys, and focus groups on selected populations), this self-study utilizes action research to improve the

practitioner-researcher's own practice. Additionally, it explores a practitioner's lived experience of AM, which is scarcely covered in existing literature. This study also plays a significant role in my professional development. I had never conducted a systematic inquiry into my own practice. However, upon reflection during the pandemic lockdown, I realized the opportunity to do research was a door waiting to be opened.

This study addresses the following questions:

1. How will my students' grades be affected by AM?
2. How will my marking experience and feedback provision change with AM?

METHODOLOGY

Self-Study

Self-study is "the study of one's self, one's actions, one's ideas ... it draws on one's life, but it is more than that. Self-study also involves a thoughtful look at texts read, experiences had, people known and ideas considered" (Hamilton & Pinnegar, 1998, p. 236). In other words, the focus of self-study is not on the *self* in isolation but on the *self* in relation to others in the teaching practice and to existing knowledge in the teaching community. Educators engage in self-study to improve their teaching practice by exploring ways to reduce the gaps they identified between whom they desire to be and who they are in their teaching practice. Loughran (2007) points out that a tacit catalyst for self-study is the "overarching desire to better align theory and practice, to be more fully informed about the nature of a knowledge of practice, and to explore and build on these 'learnings' in public ways" (p. 14). Self-study involves bringing the practitioner-researcher's self to the forefront, using their lived experiences as a resource for research, and engaging in critical reflection about their roles as practitioner and researcher (Feldman et al., 2004). This study highlighted the gap between my educational values and my actions (due to implicit bias) and how it was alleviated using AM. Grade outcomes and my marking experiences with AM provided the research data. I also engaged in critical reflection to understand my past and current selves as a practitioner and my responsibilities as a researcher. As a novice practitioner-researcher, this self-study enabled me to "start small," focusing on my teaching experiences, and to conduct the study at my own pace during the pandemic lockdown.

According to Loughran (2007), a self-study must demonstrate scholarship with transparency in methods, rigorous data collection and analysis, and knowledge development that moves beyond the self so that the knowledge created can be used and built on by the teaching community. LaBoskey (2004) adds that a self-study requires evidence of improved thinking and practice; interaction with educational literature, students, and colleagues; use of multiple methods; and formalization of the work in the professional community for deliberation. Considering these requirements and the cyclical process of actions and reflections in teaching, this study was best implemented through mixed methods action research, with the current study representing the first cycle.

Mixed Methods Action Research

When Kurt Lewin (1948) first introduced the term *action research,* he described this research approach as a reiterative process of planning, acting, observing, and reflecting.

Ivankova (2015) specifies the methodological steps involved in this cyclical process: identifying a problem, fact-finding or reconnaissance, planning an action, implementing it, evaluating and reflecting on the outcomes, deciding on changes required, and modifying the action for another research cycle. The process is repeated until the identified problem is resolved or an improvement is observed. I chose action research for this self-study because it provided a "self-reflective, systematic and critical approach" (Burns, 2010, p. 2) for addressing an immediate practical and personal concern, i.e., grading bias. It allowed me to explore an intervention, evaluate and reflect on its impact, and make changes until a satisfactory outcome is achieved.

Mixed methods research involves collecting quantitative (close-ended, numerical) and qualitative (open-ended, reflective) data to answer research questions posed. This study uses the convergent parallel mixed methods design, where both forms of data use the same or parallel variable, are collected simultaneously, analyzed separately, and compared to see if they confirm or disconfirm each other (Creswell, 2014). According to Ivankova (2015), mixed methods action research provides a comprehensive, systematic, and more rigorous evaluation of an implemented intervention. Furthermore, integrating quantitative and qualitative data improves the credibility of findings and their transferability to other contexts. In this study, I compared grade outcomes between marking anonymously and non-anonymously to see AM's impact on my grading, i.e., quantitative data collection. Second, to capture changes in my marking experience and feedback provision, I journaled the lived AM experience (i.e., qualitative data) and compared it with experiences prior to AM. I identified intersections between the two data forms to understand how implicit bias was mitigated. Mixed methods added rigour to this self-study, for quantitative data enabled me to see beyond the qualitative observations and helped increase the legitimacy of my claims.

The research framework for this study comprised six phases: diagnosing, reconnaissance, planning, acting, evaluating, and monitoring. Figure 1 captures the flow of research activities. Solid arrows indicate the actual sequence of the activities in this study, whereas dotted arrows represent potential next steps or iterations in the cycle.

Diagnosing

Having established the need to mitigate implicit bias in my marking practice and explore AM as the intervention, I developed the research questions and conceptualized a mixed-methods action research self-study guided by the Living Educational Theory (Whitehead, 1989). The Living Educational Theory emphasizes aligning research with one's values and improving oneself professionally and personally through systematic inquiry and reflection. The process begins with acknowledging oneself as a living contradiction, "holding educational values whilst at the same time negating them" (Whitehead, 1989, p. 45) and asking the question; "How do I improve my practice?" Through inquiry and reflection, a practitioner-researcher generates knowledge of their practice (their living-educational-theory). In life and in my teaching practice, I value fairness, empathy, and conscientiousness. When implicit bias negated my value of fairness, I took action by exploring AM to improve my practice. This led to two research questions: (1) How will my students' grades be affected by AM? (2) How will my marking experience and feedback provision change with AM?

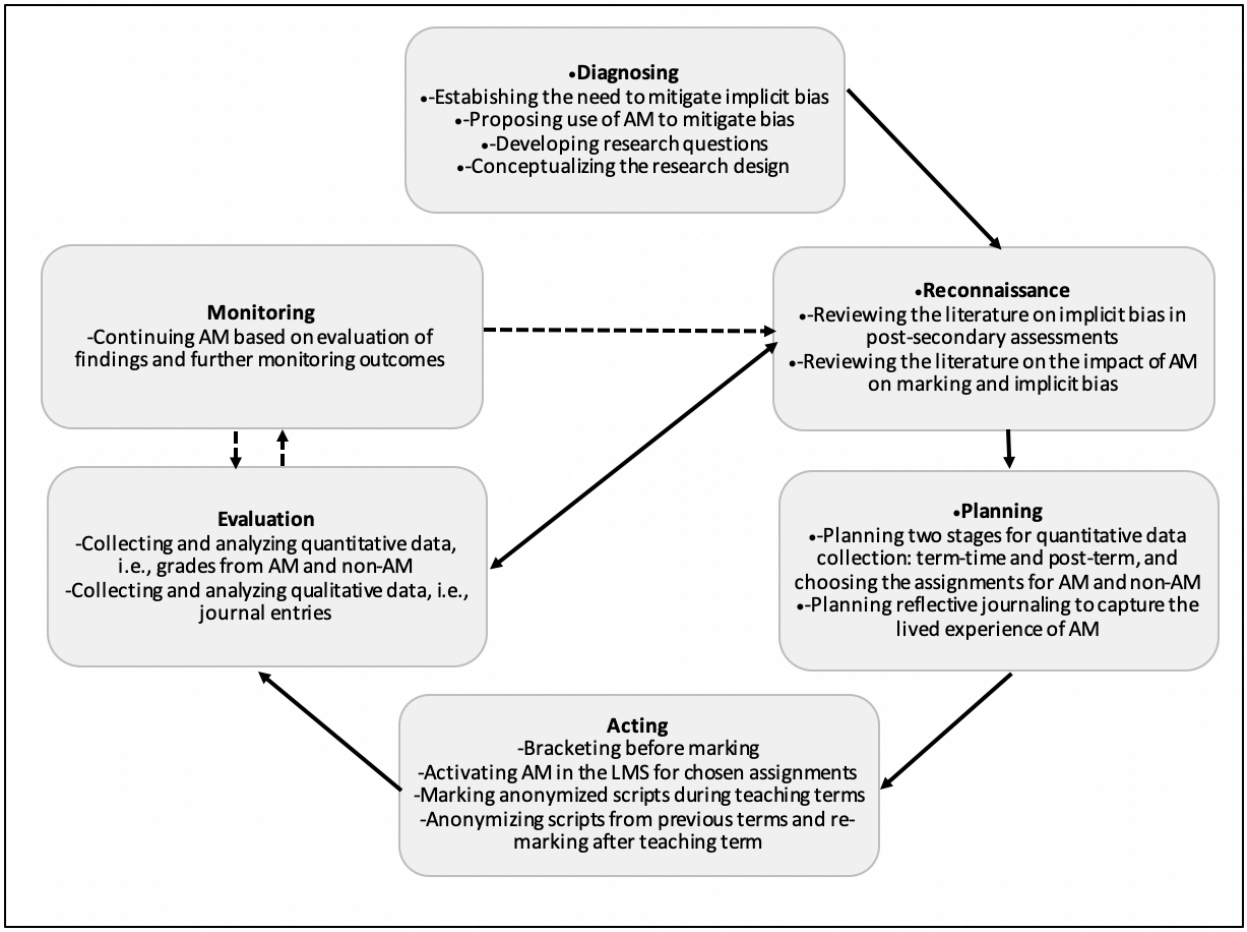

Figure 1. Mixed methods Action Research Framework

Reconnaissance

This fact-finding phase clarified the issue in my practice and established the need for an intervention. Reflecting on assessment bias, although I endeavoured to be fair in my assessments, I realized subjectivity was inevitable in writing assessments. Furthermore, I could not deny the influence of students' attributes or previous performance on my grading. *Perhaps I had been less vigilant with better performers when fatigued. Perhaps I provided more constructive comments to students who read their feedback. Maybe my mood had affected my grading.* The more I reflected, the more my fairness came into question, motivating me to conduct this study.

Elliott (1998) explains that reconnaissance occurs not just at the beginning as a fact-finding phase but also in different activities in the research cycle, e.g., the evaluation and monitoring phases. Revisiting previous research during the evaluation phase helped me understand and place my findings in context, confirming them and providing insights.

Planning

In this phase, I developed a plan to collect quantitative and qualitative data, as shown in Figure 2. Data were collected in two stages.

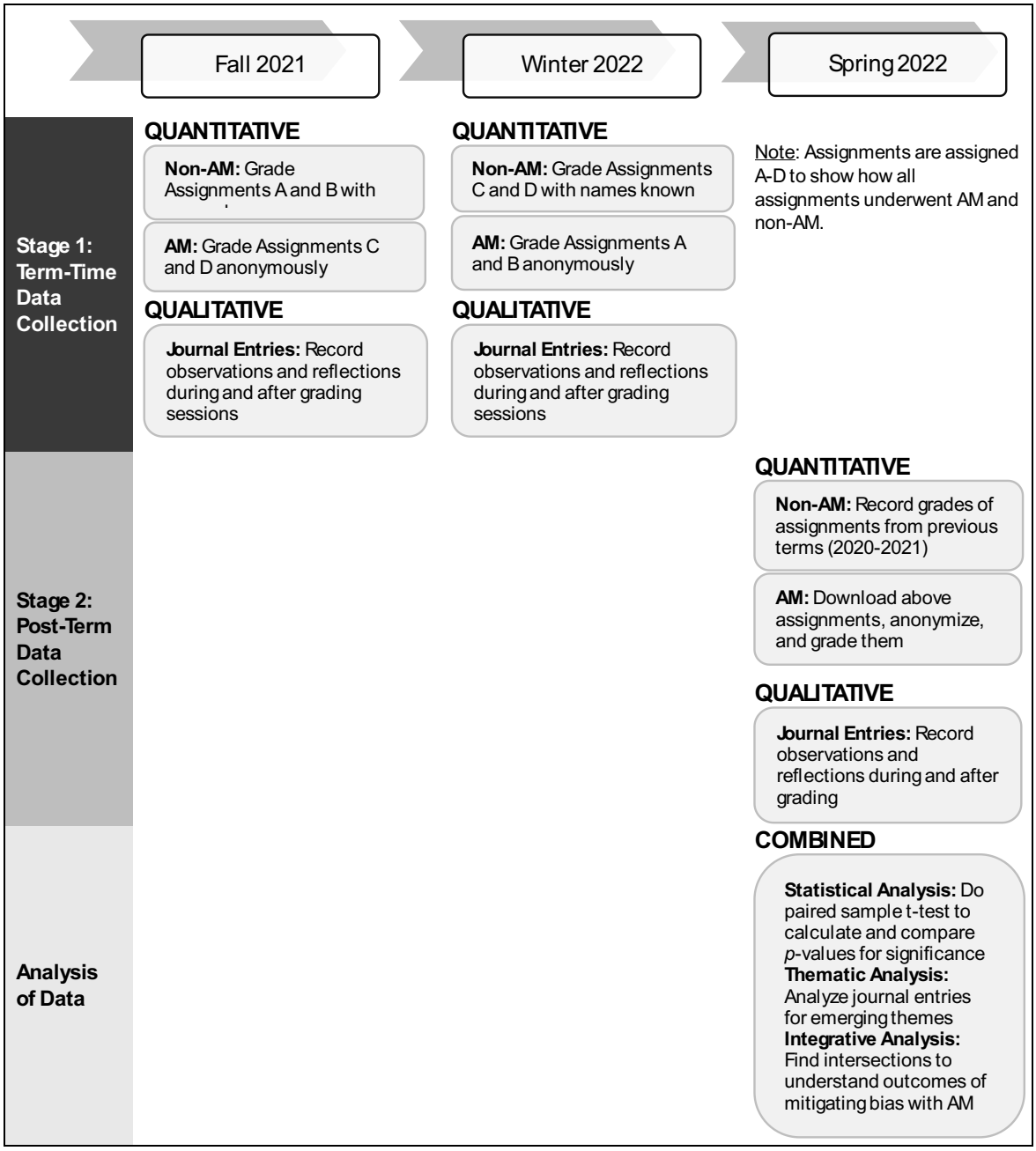

Figure 2. Action Plan Timeline

I used a one-group pretest-post-test design for quantitative data collection without a control group, as all students completed the AM and non-AM assignments. In Stage 1, during two teaching terms (2021-2022), two pairs of similar assignments (involving paragraph writing and library research skills) were selected for the study, one marked anonymously and the

other with identities known. These assignments were reversed the following term. A total of 200 scripts were marked.

Journaling was used for qualitative data collection. According to Ovens and Garbett (2020), it is a tool that enables the practitioner-researcher to "embrace the uncertainty, non-linearity, and inevitable 'messiness'" (p. 325) inherent in their teaching practice to make sense of it. Journaling involves "a process of intermingling description, commentary, introspection, and analysis in ways that enable deeper reflection and transformation" (Ovens & Garbett, 2020, p. 326). Journal entries were made after (and sometimes during) every marking session. They included descriptive notes of my thoughts, feelings, and (re)actions during the grading and feedback provision. Along with the observations were critical reflections—the questioning and reasoning of what was observed using AM.

For the post-term collection (Stage 2), I downloaded 192 paragraph-writing and research essay scripts from three previous terms (2020-2021), anonymized and re-marked them in the spring of 2022. This allowed a one-term washout period to minimize carryover effects, in which students might still be identifiable. No research ethics board review was required as the study fit with quality improvement and posed a minimal risk of student identification. I conducted a paired sample *t*-test, using the Statistical Package for Social Sciences (SPSS v.27) for statistical analysis, with the significance *p*-value set at $p<0.05$ for the comparison. The significance testing determined if any grade differences from AM were unlikely by chance (Creswell, 2014). Observations and reflections were journaled similarly to Stage 1.

Acting

At the beginning of each term (Stage 1), I informed students about the study and instructed them not to include their names in submissions. I turned on the AM feature on the LMS; once a submission was received, the feature was irreversible. All submissions had to be marked before feedback could be published. Before the AM sessions, I bracketed my preconceptions about marking anonymously: I presumed the marking would be more objective and efficient or taxing and time-consuming. I did not expect changes in the feedback provision (except for the absence of names), and I wondered if avoiding implicit bias would lead to noticeable outcomes. Tufford and Newman (2012) stress that bracketing preconceptions and monitoring them throughout the study increases research rigour. Making these notions explicit enabled me to approach the marking sessions with an open, neutral mind.

The first assignment I marked anonymously showed a Turnitin similarity index of 80%. When I opened the similarity report, the student's name appeared at the top of the screen! I realized the AM feature in the LMS did not sync with Turnitin, nor was there an AM setting on Turnitin. However, this was not problematic as I covered the name when I needed to view a report. This "loophole" allowed me to identify and handle academic integrity cases immediately.

I documented and reflected on my thoughts and behaviour after (and sometimes during) each AM session. The process was more involved than expected. Reflecting on the experience raised questions, which, upon deliberation, raised more. I was aware of the range of

emotions expressed—some uncomfortable—but I remained open to what unfolded. According to Heen (2015), "feelings as a sense, which, like other senses, conveys information to the self" (p. 266); hence, one should not suppress or withdraw from them. As I documented these emotions, I waited for personal insights to emerge.

Evaluation of Findings

In this phase, I analyzed the data to address the research questions, looking for intersections between the two data forms, relating to research literature to contextualize my findings, and reflecting on my practitioner and researcher roles to generate my own living-educational-theory.

How were my students' grades affected by AM?

Withholding student identities affected the grades I gave. Overall, using AM led to lower mean grades in both data collection stages. The mean differences in the term-time and the post-term grades were 0.55% and 5.25%, respectively. A paired sample *t*-test (Table 1) showed an insignificant term-time mean difference (p=0.775) but a significant post-term mean difference (p<0.001). Nevertheless, the effect of the post-term difference was small based on Cohen's *d* value (d=0.1).

Table 1
Students' Mean Scores and Statistical Values from the Paired Sample t-Test.

Methods of collection	Non-AM mean % (SD)	AM mean % (SD)	Mean difference (SD)	*t*	*p*	Cohen's *d*
Term-time scripts (n=100)	71.00 (17.91)	70.45 (17.58)	0.55 (19.20)	0.287	0.775	--
Post-term scripts (n=192)	68.82 (17.13)	63.56 (18.57)	5.25 (10.58)	6.881	<0.001	.1058

AM's impact on the grade (albeit small) corroborated with the impact on my initial AM experience. Without student identities, i.e., no knowledge of their prior performances to help me quickly judge their work, grading became uncomfortable and slow going. This could be due to increased cognitive load because AM prevented confirmation bias and the halo effect, which likely led to the grade difference. The lowered AM grades indicated that I tended to award more marks when students' names were available. It was possible that seeing their names might have subconsciously reinforced my desire to help them succeed.

Unsurprisingly, post-term AM led to a more significant grade decrease (5.25%) than AM during the teaching terms (0.55%). Marking anonymized scripts outside a teaching term involved no students under my care, no tight timelines, and no required feedback. Without these demands, I was more focused and marked more objectively and consistently. In contrast, I was invested in my students' success and emotionally connected with them during

teaching terms, so anonymity made little difference to my grading. However, I could not discount the fact that increased awareness of bias due to this study might also have resulted in this insignificant impact.

I was disheartened by the discrepancies between the original and post-term AM grades and questioned the consistency of my marking; I wondered if the AM grades were more reflective of students' achievement. Curious, I further analyzed post-term grade differences by exploring the relationship between the original grades and the extent of the grade reduction with AM. I grouped the original grades in 10% ranges (except for the lowest category, 0-30%, which had only four scripts), calculated the mean difference post-AM for each range, and then graphed the two variables (Figure 3). As shown by the descending dotted line in the graph, the highest-performing scripts showed the most significant reduction (-8%) in mean grade post-AM, while the lowest-performing scripts showed an increase (+6%). In other words, I tended to rate not only my stronger students' work more favourably but also my weaker students' work less favourably. This discovery demonstrated the positive and negative halo effect in my marking practice—my expectancy of students' achievement led me to be less or more vigilant during grading. This aligns with Malouff et al.'s (2014) study, in which faculty assigned higher grades to students' written work following a better prior presentation than a poorer one.

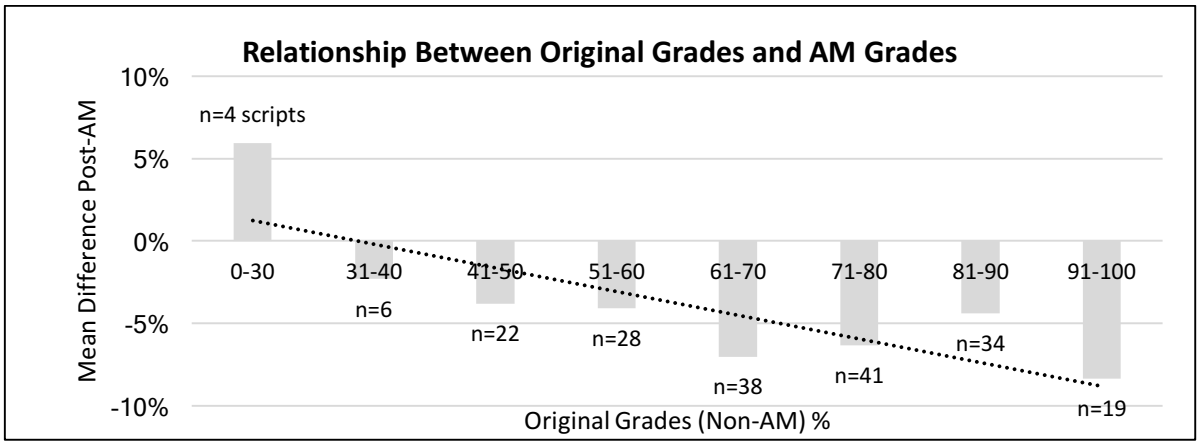

Figure 3. Mean Differences between Original Scores and AM Scores in Post-Term Marking

How did my marking experience and feedback provision change with AM?
Three themes emerged from critical observations of my marking experience and feedback provision with AM: (1) I needed to adjust to using AM, (2) I provided more consistent overall feedback, and (3) emotions played a significant role in the grading process. These themes are expressed by the italicized excerpts from my journal entries.

Adjusting to Anonymous Marking. I observed a notable difference between my first and last AM sessions during the term-time data collection. Initially, I needed to adjust to marking anonymously. The first session felt slow going, strange, and uncomfortable. I frequently guessed whose work I was assessing and felt uncertain about my grading. This distraction pointed to my tacit reliance on prior knowledge of students in my grading decisions.

> *There's more to the absence of names... I feel restless… Could it be that, in my case, the student identities provided so much cognitive offload that I'm now noticeably affected without them? … could I have paid more or less attention to students' work because of my expectations?... I keep debating over half points … why [do] I seem to be overthinking?*

Since the first anonymously marked assignment was the fourth in the course, I already had a fair idea of students' levels from their previous work. Pressured to complete the marking, I relied on this knowledge to make quick judgments of their work (halo effect). Additionally, I became more likely to notice features that confirmed my beliefs about their abilities (confirmation bias). However, when the names were withheld, the increased cognitive load led me to guess the names. This resulted in distracted and prolonged marking, which aligned with my bracketed perception that avoiding implicit bias could lead to tangible experiences and that marking could be taxing and more time-consuming. Wyatt–Smith and Castleton's (2005) study shows that teachers actively sought students' gender in the assignments, citing that being unable to draw on their knowledge of the students led to uncertainty and difficulty in reaching a judgement.

By the second AM session, I adjusted to marking anonymously, and the subsequent sessions went smoothly. Marking became more objective and streamlined (as was bracketed) once I overcame the need for student identities. My focus on how well an assignment met the rubric criteria was not influenced by perceptions of students' effort, prior performance, likability, or previous academic integrity incidents.

> *… I feel confident, being assured that I'm assessing solely the student's work and not being influenced by my perceptions of the student. I thought about students with previous academic violations and am glad the submissions are anonymized. In this way, I won't be unduly wary of previous offenders ...*

Scott (1995) reports a similar experience after a semester of marking anonymously:

> I have freed myself from the kind of bias that makes evaluation so burdensome for the conscientious and sensitive instructor… Because I am not thinking about the individual student as I grade, grading has become cleaner and simpler, and students get the message that they are being treated equitably. (p. 215)

However, I was concerned that contact cheating would be more challenging to detect with anonymity. A contracted submission might be mistaken for a strong student's work. Moreover, checking the Turnitin report for the name whenever I was suspicious defeated the purpose of AM. The only solution was to review the grades for discrepancies after publishing the feedback. Whitelegg (2002) proposes marking assignments anonymously and giving feedback *non*-anonymously. In other words, after releasing the feedback, I return to the submissions to provide personalized feedback. Although the turnaround time would increase, detecting cheating would not be undermined.

Providing More Consistent Overall Feedback. Contrary to what I bracketed, the feedback provision was affected by AM. While my inline feedback and criterion feedback on the LMS were unaffected, I noticed my behaviour when providing the overall feedback. To ensure consistency in the overall feedback, I typically followed a three-step frame: a thank you, positive comments, and points for improvement. When marking anonymously, I found myself revisiting submissions to be sure I had included positive comments. I realized I was thinking about the students who regularly read their feedback and wanted to be sure they received the comments. This pricked my conscience. Curious about the consistency in my previous overall feedback, I reviewed past term submissions for the same assignment. I was alarmed by what I found! Some feedback lacked positive comments and only listed to-dos and must-dos. This discovery somehow unnerved me, realizing I had negated my educational values (Whitehead, 1989). As a researcher, I felt anxious, knowing the expectations of trustworthy reporting.

> *I have no explanation … I am alone in my room, yet I feel exposed. ... All the while, I perceived myself as a fair and nurturing educator… now I feel like an impostor… Worse still, research demands transparency… does that mean I have to share this discovery?*

Seeing evidence of implicit bias in my practice could not be compared to simply recognizing the pervasiveness of the bias. Recognizing it was intellectual, but seeing the evidence hit me emotionally as the research findings were personal. I felt remorse for not upholding fairness and for shortchanging my students despite their not knowing what happened. I felt like an impostor because I was not who I thought I was. McNiff (2013) describes the crossing of the gap between normative assumptions and their dismantling as being so uncomfortable that one's mind is changed forever, and this change leads to changing one's practice. This incident led to an internal change in me. I finally understood that while endeavouring to uphold fairness in work and life, I could not entirely avoid implicit bias. Therefore, making changes to my practice before I went further in teaching was imperative.

Batey (2018) suggests that because feedback is a social practice, the feedback provided is influenced by whom it is intended. She contends that students' identities can activate markers' expectancies of the students based on prior beliefs and affect the feedback given. In my case, my double-checking the feedback to ensure I had positive comments revealed I had unfairly intended those comments for students I expected would read them. Since I did not know which submissions belonged to those students, I had to ensure that all feedback had positive comments. As a result, my feedback provision became consistent. Upon reflection, my failure to provide positive comments in previous feedback provisions revealed the influence of ingrained values in my childhood and my educational background on my marking practice. In Asia, my teachers were more inclined to focus on improvement than to highlight achievements in their feedback. Consequently, as an international student during my university days, I valued prescriptive feedback more than positive remarks because the former was crucial for improving my grades. As an educator, I understood the importance of balancing supportive and critical feedback and giving constructive comments to build students' confidence and motivation (Ferguson, 2011). Reflecting on why I had failed to use

the feedback frame, I could have unintentionally reverted to embedded values from childhood due to time and marking load pressures. Prioritizing the need for improvement, I listed steps for improvement and overlooked the necessity of providing positive reinforcement and building my students' confidence.

While marking anonymously reinforced consistency in feedback provision, it prevented personalized feedback. Non-personalized feedback did not seem problematic since I included positive comments; however, it undermined my efforts to build rapport with my students. Although student input was not part of this study, a few students commented that they preferred that I was aware I was marking their work. Nevertheless, they accepted my implementation of AM to safeguard fairness in assessments. However, I did not know if they viewed my implementation as a preventive measure against or an intervention for existing marking bias. Scott (1995) reports a lack of bonding with her students at the end of a term of AM. I also felt a lack of connection but could not be sure if it was due to AM or online teaching, considering this study was conducted during the pandemic lockdown. Additionally, with AM, I had a poorer sense of how my average students were progressing. After publishing the feedback, I typically scrolled through the LMS gradebook to identify students with high or low grades in case of contract cheaters and at-risk students. As a result, I was more aware of my strong or weak students and less so of the ones in between. Several studies report how AM depersonalizes feedback and undermines the potential learning feedback offers, reduces the chance of students seeking help following feedback, prevents recognition of student progress across assignments, and strains the teacher-student relationship (Batten et al., 2013; Pitt & Winstone, 2018; Whitelegg, 2002). Still, I would argue that measures can be taken in and beyond feedback to promote student engagement despite anonymity. Some best practices include pointing out successes to motivate students (i.e., positive feedback), using supportive and understandable language, being specific and selective in feedback comments (i.e., giving detailed inline feedback and highlighting a few items that students can work on to improve), providing timely written feedback, promptly responding to students' emails for help, and offering meetings for verbal feedback and clarification (Ferguson, 2011; Lowe & Shaw, 2019).

Significance of Emotions in the Grading Process. I experienced a range of emotions in both the AM and non-AM sessions. These emotions were affectual reactions to students' work. I felt satisfaction when students did well, disappointment or anxiety when they did poorly, delight or relief when they performed better than expected, frustration or annoyance when mistakes were repeated, and anger or betrayal when students committed academic violations. However, marking anonymously added an element of surprise when I saw unexpected grade outcomes after the feedback was published. The pleasure or disappointment I felt seemed intensified, for I had not reacted as deeply before using AM. This could be that I had downplayed my disappointment with students who did poorer than expected by giving them the benefit of the doubt and awarding them half points.

With AM, the first piece of sloppy work (paragraph writing assignment) I encountered displeased me. However, remembering I was in this study to mitigate bias prevented me from succumbing to my emotions, and I stayed intentionally objective throughout the

assessment. As I reflected, I wondered if my annoyance originated from my assumption that the student could not be bothered with the assignment instructions or from a worrisome thought that I failed to explain the instructions clearly. My mind went to guessing who the student was, but I chided myself immediately.

> *Why? How will the student's identity help in my marking except to influence it? Do I want to know so I can place the student in a category? Or is it to affirm the category they are already in?*

Regardless of anonymity, emotional reactions to student work can trigger bias. Interestingly, research shows that even emotions elicited from positive or negative memories unrelated to students can significantly affect experienced faculty's grading decisions in emotion-congruent ways without their awareness (Brackett et al., 2013). In other words, the faculty's emotional state approaching a student's work can affect their judgment even before reading it.

In the post-term stage, anonymously marking the previous term assignments was emotionally uneventful. I felt detached from the scripts I was grading. Nevertheless, to my surprise, with a few scripts, I recalled the pride or disappointment I had felt when grading them, although I could not remember the students' names. These scripts belonged to struggling students I had worked with or those who started strong but lost momentum. The fact that I could still recall these emotions suggested that marking was not only cognitive but also emotion-laden work. According to enabling educator Henderson-Brooks (2021), marking is considered emotional labour because navigating between the academic criteria and the desire to be supportive of students is "'deeply personal' and emotionally tiring" (p. 115). Regardless of whether I marked anonymously during the teaching terms, every assignment represented a student, and the emotions that came with marking seemed to be the cost of desiring to see students succeed..

WHAT I LEARNED

I value fairness, empathy, and conscientiousness in my teaching practice. However, my values of fairness were negated by the implicit bias in my marking practice, so I engaged with AM to mitigate the bias, improve my practice, and generate my knowledge of practice (my living-educational-theory). Through the action research process, I discovered the benefits and drawbacks of using AM to address bias in my teaching context and gained insights into myself as a practitioner and researcher.

About Marking Anonymously

Marking anonymously enabled me to uphold fairness by preventing confirmation bias and the halo effect, allowing me to mark objectively and provide more consistent feedback. Even though AM's impact on grade outcomes was insignificant, my students and I were assured that I was judging their work and not them as persons. Additionally, students were discouraged from "getting into my good books" to earn more marks, and I was in a better position to use the rubric to justify the grades I gave. Although I was hindered in personalizing feedback and found contract cheating harder to detect, I could provide

feedback *non*-anonymously to circumvent these issues. Unfortunately, I could only implement AM in "stand-alone" assignments. AM was impracticable for scaffolded assignments requiring students to follow up on previous feedback or where students were identifiable by their research topics, such as in a research project. Although using AM prevented bias related to student identities, it did not address bias from my reaction to student writing or my emotional state when I graded assignments. On reflection, the influence of implicit bias could be more extensive than what I have observed in my marking practice. This implies that AM alone was insufficient to tackle assessment bias.

About Me, the Practitioner

Seeing myself in my "contradiction" helped me better understand how negating my educational values impacted me. Having experienced discrimination, I value fairness and have always endeavoured to be a fair and inclusive educator. For that reason, seeing evidence of my selective feedback provision impacted me deeply because it was unacceptable that I discriminated in my marking. The revelation that I was not who I thought I was disoriented me, and I was compelled to take action. McNiff (2013) explains that the dismantling of one's initial self-perception drives change in one's practice. Consequently, I became more reflective in my teaching and more vigilant of my emotions during marking, finally understanding the subtle influence of implicit bias. I paused more often to consider how my feedback would affect students. When providing feedback, I imagined a student sitting across from me "listening" as I typed the comments, and I intuitively adopted an encouraging tone. When marking became mundane, and I found myself repeating the same feedback phrases, I reminded myself that these comments were personal to the student receiving them. According to Henderson-Brooks (2021), marking is an interpersonal activity in which the marker engages with the student through their work. Although the work is judged according to academic criteria, the feedback is received through an interpersonal filter and may create strong feelings for the student. Aside from changing my attitude and behaviour, I attended workshops on combatting discrimination in the post-secondary classroom, finding and forgiving our teaching selves, and becoming a critically reflective professor. According to Gordon (2021), educators must deeply explore their biases through bias education and training to reap long-term effects and adopt a growth mindset to counter stereotype threats. Faculty mindset beliefs have been shown to affect their feedback provision and student achievement and motivation.

I also learned in this self-study that my educational background and values ingrained in my childhood can influence my teaching practice. My failure to provide positive feedback when pressured to complete marking was attributed to my preference for prescriptive feedback over positive remarks in my student days. Under strain, I inadvertently reverted to providing the kind of feedback that mattered to me as a student.

Coming into this study, I had not understood the subtle influence of implicit bias on my behaviour despite acknowledging its general pervasiveness. I had the perception that I did not typically seek out students' names when I assessed their work because I was focused on reading their work. Hence, I did not think the absence of names would make any difference. This study proved otherwise—my reliance on the names to guide my grading led to an

uncomfortable experience of marking anonymously. When I shared my research findings with a few critical friends, none of whom used AM, they shared my previous sentiment and did not see the need for AM. However, marking anonymously may lead to surprising outcomes. It could be worth a try, even as a diagnostic tool, if not an intervention, for implicit bias.

About Me, the Researcher

As a self-study researcher, I learned that maintaining trustworthiness in my reporting was crucial because I had a personal stake when representing my *self* and all the "messiness" of my practice. Loughran (2007) stresses that trustworthiness is "central to creating a platform from which data sets, learnings, and conclusions might be critiqued and questioned to establish the significance and legitimacy of the outcomes being claimed" (p. 15). Therefore, "developing accurate, coherent trustworthy accounts has value… [and] ethical and moral obligations have public as well as personal strength" (Pinnegar & Hamilton, 2009, p. 13). Initially, I feared that reporting my vulnerability would invite public criticism. However, I realized that formalizing the self-study work was taking responsibility for my biased actions and upholding the values that I espoused. Moreover, making the study available to the professional community was the final requirement of the self-study methodology (LaBoskey, 2004) and must be done for the completion of my study. According to Feldman (2003), self-study is not only scholarly work but also moral work—practitioner-researchers study their practice in order to improve it in a specific direction that will affect the teaching community. I hope my work will pique readers' curiosity and prompt them to consider using AM; it could reveal hidden aspects of their teaching selves and the covert operation of implicit bias in their practice.

Overall, this study implies that marking anonymously is just one strategy for combatting assessment bias, which has widespread consequences. More attention is needed to research mitigation strategies for grading bias unrelated to student identities. This study also suggests that implicit bias could be more prevalent in negating educational values we espouse than we might realize.

A few limitations were noted in this study. AM was implemented in only four assignments of one course. It would be interesting to see AM's impacts on other courses. Second, this study was conducted during the pandemic when classes were online. The dynamics of online classes are different from in-person delivery. The perceptions I formed about students when teaching black squares (because students tend not to turn on their videos) would differ from those formed in in-person instruction. Likewise, the extent of relationship-building would vary. In addition, how I perceived myself could have distorted my observations of my lived experience. Other faculty may react to the AM differently or not at all. Considering these limitations, plans for the following action cycle include implementing AM across all possible assignments in the course (and other courses) and inviting collaboration with colleagues. Extending the use of AM and collaborating with other practitioners to understand AM's impact would impart a more accurate picture of AM's benefits, challenges, and consistency.

Conclusion

This action research self-study highlighted the influence of confirmation bias and the halo effect in my practice and the benefits and drawbacks of using AM to mitigate them. It helped me better understand myself as a practitioner and researcher and improved my practice. My awareness of the subtlety of implicit bias has motivated me to stay reflective in my teaching and vigilant in my assessments.

Doing this research has been an eye-opening experience. I realized not only the level of rigour it required but also the commitment and integrity needed to accept and report the findings about me as revealed to me. This self-study has changed how I view myself and my bias, revitalized my teaching practice, and created a deep desire for professional growth. Finally, this fulfilling experience of doing action research has spurred me to look for more "research doors" to open. I hope educators considering a systematic inquiry into their practice will be inspired to "open their first door." 

References

About anonymous marking. (2022). D2L Brightspace. https://documentation.brightspace.com/EN/le/assignments/instructor/about_anonymous_marking.htm

Batey, J. (2018). *Expectancies, marking and feedback in undergraduate student assessment* [Dissertation]. University of Winchester.

Batten, J., Batey, J., Shafe, L., Gubby, L., & Birch, P. (2013). The influence of reputation information on the assessment of undergraduate student work. *Assessment & Evaluation in Higher Education*, *38*(4), 417–435. https://doi.org/10.1080/02602938.2011.640928

Brackett, M. A., Floman, J. L., Ashton-James, C., Cherkasskiy, L., & Salovey, P. (2013). The influence of teacher emotion on grading practices: A preliminary look at the evaluation of student writing. *Teachers and Teaching*, *19*(6), 634–646. https://doi.org/10.1080/13540602.2013.827453

Brennan, D. J. (2008). University student anonymity in the summative assessment of written work. *Higher Education Research & Development*, *27*(1), 43–54. https://doi.org/10.1080/07294360701658724

Burns, A. (2010). *Doing action research in English language teaching: A guide for practitioners*. Routledge.

Bygren, M. (2020). Biased grades? Changes in grading after a blinding of examinations reform. *Assessment & Evaluation in Higher Education*, *45*(2), 292–303. https://doi.org/10.1080/02602938.2019.1638885

Chowdhury, S., Klauzner, I., & Slonim, R. (2020). *What's in a name? Does racial or gender discrimination in marking exist?* (13890; IZA Discussion Paper). Institute of Labor Economics (IZA).

Creswell, J. W. (2014). *Research design: Qualitative, quantitative, and mixed methods approaches* (4th ed.). Sage Publications, Inc.

Elliott, J. (1998). *Action research for educational change* (repr). Oxford University Press.

Feldman, A. (2003). Validity and quality in self-study. *Educational Researcher*, *32*(3), 26–28. https://doi.org/10.3102/0013189X032003026

Feldman, A., Paugh, P., & Mills, G. (2004). Self-study through action research. In J. J. Loughran, M. L. Hamilton, V. K. LaBoskey, & T. Russell (Eds.), *International Handbook of Self-Study of Teaching and Teacher Education Practices* (pp. 943–977). Springer Netherlands. https://doi.org/10.1007/978-1-4020-6545-3_24

Ferguson, P. (2011). Student perceptions of quality feedback in teacher education. *Assessment & Evaluation in Higher Education*, *36*(1), 51–62. https://doi.org/10.1080/02602930903197883

Forgas, J. P. (2011). She just doesn't look like a philosopher…? Affective influences on the halo effect in impression formation: Mood and halo effects. *European Journal of Social Psychology*, *41*(7), 812–817. https://doi.org/10.1002/ejsp.842

Gordon, A. (2021). Better than our biases: Using psychological research to inform our approach to inclusive, effective feedback. *Clinical Law Review*, *27*(2), 195–252.

Hamilton, M. L., & Pinnegar, S. (1998). Conclusion: The value and the promise of self-study. In *Reconceptualizing teaching practice: Developing competence through self-study* (pp. 235–246). Routledge.

Henderson-Brooks, C. (2021). Marking as emotional labour: A discussion of the affective impact of assessment feedback on enabling educators. *Henderson-Brooks*, *8*(1), 110–121.

Hinton, D. P., & Higson, H. (2017). A large-scale examination of the effectiveness of anonymous marking in reducing group performance differences in higher education assessment. *PLOS ONE*, *12*(8), e0182711. https://doi.org/10.1371/journal.pone.0182711

Huskanovic, J., & Adem Nur, B. (2020). *Discrimination against foreign named students: A study based on a Swedish university* [Thesis, Linnaeus University]. https://www.diva-portal.org/smash/get/diva2:1441293/FULLTEXT01.pdf

Ivankova, N. V. (2015). *Mixed methods applications in action research: From methods to community action*. SAGE Publications.

Kiekkas, P., Igoumenidis, M., Stefanopoulos, N., Bakalis, N., Kefaliakos, A., & Aretha, D. (2016). Gender bias favors female nursing students in the written examination evaluation: Crossover study. *Nurse Education Today*, *45*, 57–62. https://doi.org/10.1016/j.nedt.2016.06.010

LaBoskey, V. K. (2004). The methodology of self-study and Its theoretical underpinnings. In J. J. Loughran, M. L. Hamilton, V. K. LaBoskey, & T. Russell (Eds.), *International Handbook of Self-Study of Teaching and Teacher Education Practices* (pp. 817–869). Springer Netherlands. https://doi.org/10.1007/978-1-4020-6545-3_21

Lewin, K. (1948). Resolving social conflicts: Selected papers on group dynamics. *American Sociological Review*, *13*(6), 778. https://doi.org/10.2307/2086832

Lindsey, P., & Crusan, D. (2011). How faculty attitudes and expectations toward student nationality affect writing assessment. *Across the Disciplines*, *8*(4), 1–19. https://doi.org/10.37514/ATD-J.2011.8.4.23

Loughran, J. (2007). Researching teacher education practices: Responding to the challenges, demands, and expectations of self-study. *Journal of Teacher Education*, *58*(1), 12–20. https://doi.org/10.1177/0022487106296217

Lowe, T., & Shaw, C. (2019). Student perceptions of the 'best' feedback practices: An evaluation of student-led teaching award nominations at a higher education institution. *Teaching & Learning Inquiry*, *7*(2), 121–135. https://doi.org/10.20343/teachlearninqu.7.2.8

MacDougall, M., Riley, S. C., Cameron, H. S., & McKinstry, R. (2008). Halos and horns in the assessment of undergraduate medical students: A consistency-based approach. *Journal of Applied Quantitative Methods*, *3*(2), 116–128.

Malouff, J. M., Stein, S. J., Bothma, L. N., Coulter, K., & Emmerton, A. J. (2014). Preventing halo bias in grading the work of university students. *Cogent Psychology*, *1*(1), 988937. https://doi.org/10.1080/23311908.2014.988937

Malouff, J. M., & Thorsteinsson, E. B. (2016). Bias in grading: A meta-analysis of experimental research findings. *Australian Journal of Education*, *60*(3), 245–256. https://doi.org/10.1177/0004944116664618

McDaniel, C. (1994). Anonymous grading: An ethical teaching strategy. *Nurse Educator*, *19*(5), 11–12.

McNiff, J. (2013). *Action research: Principles and practice* (Third edition). Routledge.

Novaes, C. D. (2013, May 2). On anonymous grading. *New APPS: Art, Politics, Philosophy, Science*. https://www.newappsblog.com/2013/05/on-anonymous-grading.html

O'Hagan, S. R., & Wigglesworth, G. (2015). Who's marking my essay? The assessment of non-native-speaker and native-speaker undergraduate essays in an Australian higher education context. *Studies in Higher Education*, *40*(9), 1729–1747. https://doi.org/10.1080/03075079.2014.896890

Ovens, A., & Garbett, D. (2020). Weaving self-studies through journaling: A systematic review. In *Textiles and tapestries: Self-study for envisioning new ways of knowing* (pp. 325–332). EdTech Books.

Owen, C., Stefaniak, J., & Corrigan, G. (2010). Marking identifiable scripts: Following up student concerns. *Assessment & Evaluation in Higher Education*, *35*(1), 33–40. https://doi.org/10.1080/02602930802474177

Pinnegar, S., & Hamilton, M. L. (2009). *Self-study of Practice as a Genre of Qualitative Research: Theory, Methodology, and Practice*. Springer Netherlands. https://doi.org/10.1007/978-1-4020-9512-2

Pitt, E., & Winstone, N. (2018). The impact of anonymous marking on students' perceptions of fairness, feedback and relationships with lecturers. *Assessment & Evaluation in Higher Education*, *43*(7), 1183–1193. https://doi.org/10.1080/02602938.2018.1437594

Scott, E. L. (1995). "Mokita," the truth that everybody knows, but nobody talks about: Bias in grading. *Teaching English in the Two-Year College*, *22*(3), 211–216.

Sharp, J. G., & Zhu, X. (2020). Anonymous marking in Higher Education: A review of literature for practitioners. *Journal of Higher Education Research*, *3*(1). https://cpb-eu-w2.wpmucdn.com/blogs.lincoln.ac.uk/dist/c/8316/files/2020/04/Anonymous-Marking-Sharp-2020.pdf

Shay, S., & Jones, B. (2008). Anonymous examination marking at University of Cape Town: The quest for an `agonising-free zone'. *South African Journal of Higher Education*, *20*(4), 528–546. https://doi.org/10.4314/sajhe.v20i4.25682

Staats, C., Capatosto, K., Tenney, L., & Mamo, S. (2017). *State of the science: Implicit bias review 2017*. Kirwan Institute For The Study Of Race & Ethnicity. http://kirwaninstitute.osu.edu/implicit-bias-training/resources/2017-implicit-bias-review.pdf

Tufford, L., & Newman, P. (2012). Bracketing in qualitative research. *Qualitative Social Work*, *11*(1), 80–96. https://doi.org/10.1177/1473325010368316

Whitehead, J. (1989). Creating a living educational theory from questions of the kind, 'How do I Improve my practice?' *Cambridge Journal of Education*, *19*(1), 41–52. https://doi.org/10.1080/0305764890190106

Whitelegg, D. (2002). Breaking the feedback loop: Problems with anonymous assessment. *Planet*, *5*(1), 7–8. https://doi.org/10.11120/plan.2002.00050007

Wyatt–Smith, C., & Castleton, G. (2005). Examining how teachers judge student writing: An Australian case study. *Journal of Curriculum Studies*, *37*(2), 131–154. https://doi.org/10.1080/0022027032000242887

BIOGRAPHICAL NOTE:

Len Chan teaches communications courses at Conestoga College, Ontario. She holds a Master of Education in Language and Literacy and has over 17 years of experience in adult ESL and EAP instruction. A novice practitioner-researcher, her areas of research interest include second language writing, written corrective feedback, pedagogical grammar, evaluation of learning, teacher development, and teacher identity.