The Mathematical Structure of a Bias-Minimized Assessment Algorithm

ZHUHAN JIANG and JIANSHENG HUANG
School of Computing and Mathematics
University of Western Sydney
New South Wales 1797
AUSTRALIA
z.jiang {j.huang}@scm.uws.edu.au

Abstract: - We propose an assessor reallocation algorithm that aims to objectively reduce the marking biases of the individual markers based on their earlier marking statistics. The underlying mathematical structure along with a number of pertinent statistical properties and relationships has been analyzed for the model to assure the validity of the proposed methodology. Experiments on simulated data and on the real cases have also been conducted to illustrate the effectiveness on the reduction of the accumulated marking biases over multiple assessment items involving multiple assessors.

Key-Words: - algorithm, modeling marking biases, mathematical structure, multiple assessments and assessors.

1 Introduction

Even though computer technologies are now widely employed in education [1,2] to provide additional support and to reduce the necessary repetitions, a teaching team is still almost always required to co-teach a large university subject of several hundred students or more, with the lectures, tutorials or practicals being repeated for different timeslots or campuses. Quite a few casual tutors are often recruited within a very short space of time, right at the beginning of a semester, each of whom is typically assigned to marking all the assessment items other than the final exam for their own group of students. When an assessment item is marked by different tutors, the marks inevitably vary, and can vary to the extent of about 20% when casuals are heavily involved, which was exactly what happened some 2 years back that prompted this research work. Most of the studies [3] on the assessment have been somewhat subjective and are also closely linked with the individual subjects. One exception is the so-called grading on the curve [4,5] which determines the student grades according to the normal distribution of the marks. The main aim of this work is to devise a marking allocation scheme that is universally applicable to all subjects of large cohorts, and can be utilized to improve the overall marking fairness and consistency without having to formally evaluate the performance of the individual assessors. This scheme will thus reduce effectively the assessment inconsistencies due to the involvement of a large number of disparate or inhomogeneous teaching team members.

The studies on the assessment consistency have been perhaps as old as the education itself. However we will here only mention some of the work that are most pertinent to our current methodology. To start with, it was found [6], quite consistent to our intuitive understanding, that the marking of a given assessment item tends to receive a better mark if the marking is immediately preceded by marking a poorer work of the same item. If the marking of the same work is preceded immediately by a better work then the marker tends to give a poorer mark. This is a typical example of *subjective* biases one perhaps can't always avoid completely, just like people can have illusions of the same shape when it is placed at different backgrounds. However, it is also found [7] that individual assessors vary in their level of leniency or "biases", and the leniency of most assessors remains internally consistent throughout the local marking batches. While a generic marking reliability [8] is perhaps not yet conclusive, the prevailing marking consistency over a longer period of time [9] shows that there is a minor decrease in leniency and that most assessors' marking clusters remain nonetheless consistent within. These observations are in fact very much consistent with our currently work, or are somewhat implicitly utilized. In a different perspective, the acute issues of marking consistency across multiple assessors for the large student cohorts have also been analyzed recently in [10,11], illustrating that the students' perception of inconsistent difference in grade was not unfounded, and the inconsistency problem can be exacerbated by other factors such as the inconsistencies in the language the markers use when providing the feedback. However there doesn't appear to exist so far any quantitatively-based algorithms with which the marking inconsistency can be objectively and systematically reduced, and this is why it has been main driving force of this work. It is worth noting that while we hereby concentrate on fairer schemes to assess mostly written student work, multiple choice questions are also known to be capable of assessing knowledge [12], and it would be even more interesting to incorporate the assessment schemes more closely to the cooperative learning strategies [13].

The traditional approach of entailing the marking consistency involves training the new staff as well as supplying very detailed marking guides. However these are not always sufficient, and the additional staff training is typically not even feasible due to the time and funding constraints. Moreover, subjective marking biases are often unavoidable for the design based or opinion based assessment items. Unless there are obvious and substantial marking errors, it is simply not reasonable to ask a marker to go back to assessment item and re-mark it because it was somehow felt that the marking might have been subconsciously prejudicial. Marks rescaling for certain groups of students may also seem a good alternative, but it can be difficult to formally justify which student marks are to be rescaled and which are not. Our purpose here is thus to reallocate the assessment tasks to different markers for the different assessment items, according to the marks statistics of all the markers for the previous assessment items. This way, the potential and often unconscious marking biases generic to the individuality of the assessors is well spread out and compensated over several items as much as possible. The obvious advantage of this over the marks rescaling is that all marks from all the markers are still formally deemed “accurate” to all the students.

In order that each marker be profiled for his potential “biases” or leniency before any marker reallocation can be sensibly applied, some marks must already exist for all the participating markers. This should not be a problem as even an assessor joining the marking team in the *middle* of a semester can simply just not participate in the reallocation scheme in the 1st assessment instance, or more precisely, they will join the marker reallocation at a priority lower than those who already marked some assessment items. Moreover, the more items an assessor has already marked, the more accurately he or she will be profiled for the “bias” with which the proposed reallocation scheme will reassign the markers to the different students. The marks coming off the assessment items via the reallocation scheme will again be added back to further profile the individual markers. Hence this scheme can also be applied repeatedly for all of the later assessment items, or simply applied selectively to just some major assignments. We note that for a large student cohort, some students will transfer from one activity group to another from time to time due to their changed commitment elsewhere. Hence it makes sense to expect that a given student may be marked by different assessors for his or her different assessment items, and will be even more so if the reallocation scheme is applied repeatedly over several assessment items.

This paper is organized as follows. In section II, we first set up a simple formalism to rescale marks, and then illustrate the general procedure of our scheme to reassign the new marking duties according to the marks distribution of the previous assessment items. Section III then derives the individual markers’ biases and shows how to measure the accumulated marking biases. Section IV proposes the actual marker reallocation algorithm to minimize the accumulated marking inconsistencies or biases while maintaining the homogeneity of the statistical characteristics as much as possible. We then further justify in Section V our proposed methodology via both the experimental simulations and an actual subject delivery. Finally Section VI gives a conclusion.

2 The Main Marks Statistics

In this section, we will first set up a marks rescaling formalism that can be applied directly to realign the available marks, typically coming from a number of different markers. We will then explore how to reassign markers to the students for their later assessment items according to the existing marks statistics, based on the previous assessment items, for both the individual makers and for the individual students, for the purpose of minimizing the accumulation of the marking discrepancies or biases.

2.1 Marks Rescaling and the Broad Marking Reallocation Scheme

Let $J=\{1, 2, \ldots, n\}$ denote the set of all the markers, and $\mu^{(j)}$ and $\sigma^{(j)}$ denote the mean and the standard deviation of the marks for the group of students marked by the j-th marker. We also assume that student marks will observe a normal distribution [14,4]. Then one way of consistently rescaling the student marks is to transform them via

$$Y_i^{(j)} = \mu' + (X_i^{(j)} - \mu^{(j)})\sigma'/\sigma^{(j)},\ i=1, \ldots, N^{(j)}, \qquad (1)$$

where all the marks $\{ X_i^{(j)} \}$, i=1, …, $N^{(j)}$, in the j-th group of students, are transformed into $Y_i^{(j)}$ respectively, and μ (μ') and σ (σ') are the mean and standard deviation for the original (respectively the transformed) marks. Since $\Sigma_i (X_i^{(j)} - \mu^{(j)})^2 = N^{(j)} \times [\sigma^{(j)2} + (\mu - \mu^{(j)})^2]$ holds for all j's, we note that the μ and σ can also be derived from those $\mu^{(j)}$ and $\sigma^{(j)}$ directly via

$$N = (\Sigma_{j\in J} N^{(j)}),\quad \mu = (\Sigma_{j\in J} N^{(j)} \times \mu^{(j)})/ N,$$

$$\sigma = [(\Sigma_{j\in J} (N^{(j)} \times (\sigma^{(j)2} + (\mu - \mu^{(j)})^2))/\mathrm{N}]^{1/2}. \qquad (2)$$

The target μ' and σ' thus become the mean and the standard deviation of the transformed new marks $\{ Y_i^{(j)} \}$, i=1, …, $N^{(j)}$. Even though rescaling via (1) seems to offer a quick remedy to transform all marks towards a more suitable μ' and σ' anticipated by the unit (subject) instructor, the hidden biases due to the different markers have not been addressed in any way. While such subjective biases may be reduced by staff training, they are often inherent of each individuals, and a timely training is not always possible when new casual staff is recruited at the last minute. Hence this calls for an alternative approach that will maximally cancel out the potential marking biases.

Suppose K assignments for each student have already been marked by several different markers. Then for the next assessment item, the (K+1)-th assignment, for each given student which assessor should be assigned to mark the student's assessment item so as to minimise the accumulation of the bias that may be intrinsic to each individual assessor? In order to answer this question systematically, we need to first set up properly the relevant mathematical structure and notations.

Let $I=\{ i \}$ be the set of all students, $J=\{ j \}$ with $j\neq 0$ be the set of all the markers, and $J^*=J\cup \{ 0 \}$ for each $i\in I$. For each integer k with $1\leq k\leq K$, we also assume the existence of ρ_k: $I\rightarrow J^*$ such that $j=\rho_k(i)$ indicates that the k-th assignment for student i had been assessed by marker j if $j\in J$, or the assignment was not submitted at all if j=0. Our task is to determine a ρ_{K+1}: $I\rightarrow J^*$ such that the newer marking allocation ρ_{K+1} best compensates the potential marking biases already experienced in marking the first K assessment items. Let $I_k=\{i\in I:\ \rho_k(i)\in J\}$ for $1\leq k\leq K$, and $I_k^{(j)}=\{i\in I_k:\ j=\rho_k(i)\}$, then $I_k = \cup I_k^{(j)}$, and $I_k^{(j)} \cap I_k^{(j')} = \emptyset$, the empty set, $\forall j\neq j'$. Hence for the k-th assessment item, its marking details are completely determined by the tuple (I_k, ρ_k, w_k, ν_k) where w_k represents the positive weight of the assessment item and ν_k: $I\rightarrow\mathbb{R}^+$, ($\mathbb{R}^+$ being the set of non-negative real numbers), gives the grading percentage mark $x_{k,i}$ via $\nu_k(i)$ for each student $i\in I$, i.e. $x_{k,i}=\nu_k(i)$ and $0\leq x_{k,i}\leq 1$. The actual mark student i received for this k-th assignment is thus $w_k \cdot x_{k,i}$. We note however that in practice one may choose $x_{k,i}$ to be within the range of 0 to 10 because this would essentially have no impact on our proposed algorithm but the marks out of 10 would make a better intuitive sense.

If $\forall i\in I$ marker $j=\rho_k(i)$ has marked the k-th assignment for i, then the percentage mark $x_{k,i}$ will be denoted by $x_{k,i}^{(j)}$. Hence $\{ x_{k,i}^{(j)} \}$ for $i\in I_k^{(j)}$ has $N_k^{(j)} \equiv |I_k^{(j)}|$ elements, $N_k = \Sigma_{j\in J} N_k^{(j)}$ is the total number of students who submitted the k-th assignment, and $N=\Sigma^K_{k=1} N_k$ is the total number of student submissions. $\forall j\in J$, we denote by $\mu_k^{(j)}$ and $\sigma_k^{(j)}$ respectively the mean and the standard deviation for the (percentage) marks $\{ x_{k,i}^{(j)} \}$ for $i\in I_k^{(j)}$ given by j for the k-th assignment. Likewise we denote by $\mu^{(j)}$ and $\sigma^{(j)}$ respectively the mean and the standard deviation for the $\{ x_{k,i}^{(j)} \}$ over all the i and k with $i\in I_k^{(j)}$ and $1\leq k\leq K$. We also denote by μ_k and σ_k respectively the overall mean and standard deviation of the marks of the k-th assignment for the submitted students. It also makes sense to set the "statistics" for marks of those who didn't submit to those for the corresponding overall, hence we set $\mu_k^{(0)} = \mu_k$ and $\sigma_k^{(0)} = \sigma_k$. As for all the marks for the submitted K assignments, we naturally denote by μ and σ the corresponding mean and standard deviation.

Since each marker has his own marking tendency in leniency or harshness, we will first establish for each individual marker his "bias" profile by analysing the past marking statistics, and then allocate suitable markers to the next assessment item so that the total accumulated marking biases are shared evenly or pro-rata among all the students. The main procedure of this strategy is summarised in Figure 1, of which the details will be expounded in the subsequent sections.

2.2 Derivation of the Underlying Statistical Properties

The purpose of this subsection is to establish the relevant underlying statistical properties and formulas that will be helpful or necessary for the later analysis and development of our methodology.

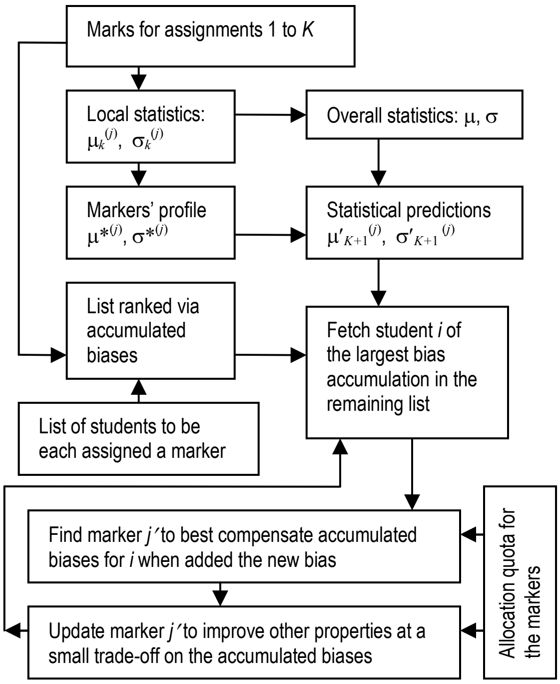

Figure 1: Procedure to reallocate markers

Since the transformation (1) is a linear filter and linear filters are known to exhibit rich properties, see e.g. [15,16], we expect to see a similar set of nice properties as shown below.

First we see that if the data $\{X_i\}$ have the mean μ and the standard deviation σ, then $X'_i=\mu+(X_i-\mu)\sigma'/\sigma$ (and $X''_i=\mu'+(X'_i-\mu')\sigma''/\sigma'$, respectively) will have their mean μ' and standard deviation σ' (and μ'' and σ'' respectively). The same X''_i can be derived *directly* from $X''_i=\mu+(X_i-\mu)\sigma''/\sigma$. In other words, the transformed results X''_i are independent of the transformation paths. Secondly, it is easy to verify that the transform $X_i \to X'_i$ is invertible as long as both σ and σ' are greater than 0.

Next we need to decide how to meaningfully synthesise the statistics for the existing marks of the past assessment items. In terms of profiling an assessor's marking behaviour, if w_k and $w_{k'}$ are the marks weight for the k-th and k'-th assignment respectively and $w_k \neq w_{k'}$, then the corresponding percentage marks should contribute to the profiling in proportion to the marks weights. We observe that if one modifies all marks weights uniformly by a fixed factor δ>0, the resulting μ and σ will remain the same. But if δ is large enough, then the effect of rounding the weights to integers will become less significant. Therefore we can assume without loss of generality that the marks weights w_k are integers. Hence we have for the overall mean

$$\begin{aligned}\mu &= \sum_{k=1}^{K}\sum_{i\in I_k} w_k x_{k,i} \Big/ \sum_{k=1}^{K} w_k N_k \\ &= \sum_{k=1}^{K} w_k N_k \mu_k \Big/ \sum_{k=1}^{K} w_k N_k\end{aligned} \tag{3}$$

and for the overall standard deviation

$$\begin{aligned}\sigma^2 &= \sum_{k=1}^{K}\sum_{i\in I_k} w_k (x_{k,i}-\mu)^2 \Big/ \sum_{k=1}^{K} w_k N_k \\ &= \sum_k w_k \sum_{i\in I_k}\left[(x_{k,i}-\mu_k)^2 + N_k(\mu-\mu_k)^2\right] \Big/ \sum_k w_k N_k \\ &= \sum_k w_k N_k\left[\sigma_k^{\,2} + (\mu-\mu_k)^2\right] \Big/ \sum_k w_k N_k\end{aligned} \tag{4}$$

in which we have also used the identity

$$\begin{aligned}&\sum_{i\in I_k}(x_{k,i}-\mu)^2 \\ &= \sum_i\left[(x_{k,i}-\mu_k)^2 + (\mu-\mu_k)^2\right] + 2\left[\sum_i (x_{k,i}-\mu_k)\right](\mu_k-\mu) \\ &= \left[\sum_i (x_{k,i}-\mu_k)^2\right] + N_k(\mu-\mu_k)^2 + 2(\mu_k-\mu)\left[\left(\sum_{i\in I_k} x_{k,i}\right) - N_k\mu_k\right] \\ &= \left[\sum_i (x_{k,i}-\mu_k)^2\right] + N_k(\mu-\mu_k)^2 .\end{aligned}$$

We note that (3) and (4) show how to derive the overall μ and σ from those μ_k and σ_k for the k-th assignment, and they also imply $\min_k(\mu_k) \le \mu \le \max_k(\mu_k)$, and $\min_k(\sigma_k^{\,2}) + \min_k((\mu-\mu_k)^2) \le \sigma^2 \le \max_k(\sigma_k^{\,2}) + \max_k((\mu-\mu_k)^2)$, which are consistent with our intuitive expectations.

For the student marks of the assignments, the smallest granularity for the statistics in this work is on the j-th group for the k-th assignment. More precisely, we will have to always calculate

$$\mu_k^{(j)} = \sum_{i\in I_k^{(j)}} x_{k,i}^{(j)} / |I_k^{(j)}|, \quad \left[\sigma_k^{(j)}\right]^2 = \sum_{i\in I_k^{(j)}}\left(x_{k,i}^{(j)} - \mu_k^{(j)}\right)^2 / N_k^{(j)} \tag{5}$$

for all the j's and k's. Hence it makes sense to also find other statistics in terms of these $\mu_k^{(j)}$ and $\sigma_k^{(j)}$ whenever possible. Since μ_k and σ_k can be calculated from $\mu_k^{(j)}$ and $\sigma_k^{(j)}$ via (2), the overall μ and σ can also be derived from them via (3) and (4). If we apply the derivation in (3) and (4) to only the marks provided by marker j, then we obtain

$$\begin{aligned}&\mu^{(j)} = \sum_{k=1}^{K} w_k N_k^{(j)}\mu_k^{(j)} \Big/ \left(\sum_{k=1}^{K} w_k N_k^{(j)}\right), \quad N_k^{(j)} = \left|I_k^{(j)}\right|, \\ &\left[\sigma^{(j)}\right]^2 = \sum_{k=1}^{K} w_k N_k^{(j)}\left[\left(\sigma_k^{(j)}\right)^2 + (\mu^{(j)} - \mu_k^{(j)})^2\right] \Big/ \left(\sum_{k=1}^{K} w_k N_k^{(j)}\right)\end{aligned} \tag{6}$$

Similar to those assignment-wise statistics μ_k and σ_k, these marker-wise statistics $\mu^{(j)}$ and $\sigma^{(j)}$ are also sufficient to determine the overall mean and standard deviation via

$$\begin{aligned}&\mu = \sum_{j\in J} N^{(j)}\mu^{(j)} \Big/ \left(\sum_{j\in J} N^{(j)}\right), \quad N^{(j)} = \sum_{k=1}^{K} w_k N_k^{(j)}, \\ &\sigma^2 = \sum_{j\in J} N^{(j)}\left[\left(\sigma^{(j)}\right)^2 + (\mu-\mu^{(j)})^2\right] \Big/ \left(\sum_{j\in J} N^{(j)}\right).\end{aligned} \tag{7}$$

We note that the denominators in (3) and (4) are the same as those in (6) and (7) because $N = \sum_k w_k N_k = \sum_j N^{(j)}$, and (6) and (7) illustrate the direct connection of μ and σ to the marker-wise statistics which are also needed to determine the bias profiles for the markers via (11) below.

Suppose for a given number K of marked assignments, one has already calculated the statistics $\mu^{(j)}$, $\sigma^{(j)}$, μ and σ. When the marks of an additional assessment item are added to the consideration, by definition these statistics need to be recalculated from scratches with K replaced by K+1. We will however show here how to update these statistics for K+1 assignments through those for the first K assignments and those for the last (K+1)-th assignment. Let these new statistics be denoted by the primed counterpart such as $\mu'^{(j)}$, $\sigma'^{(j)}$, μ′ and σ′ for the case of K+1 marked assignments. Then we have

$$\begin{aligned}
&\mu'^{(j)} = \left[\mu^{(j)} + \gamma_{K+1}^{(j)}\mu_{K+1}^{(j)}\right] / \left[1+\gamma_{K+1}^{(j)}\right],\\
&\left[\sigma'^{(j)}\right]^2 = \left[(\sigma^{(j)})^2 + (\mu'^{(j)} - \mu^{(j)})^2\right.\\
&\qquad \left. + \gamma_{K+1}^{(j)}\left[(\sigma_{K+1}^{(j)})^2 + (\mu'^{(j)} - \mu_{K+1}^{(j)})^2\right]\right] / \left[1+\gamma_{K+1}^{(j)}\right],\\
&\gamma_{K+1}^{(j)} = w_{K+1}N_{K+1}^{(j)} / \sum_{k=1}^{K} w_k N_k^{(j)} .
\end{aligned} \tag{8}$$

To prove this, we first observe from (6)

$$\begin{aligned}
\mu'^{(j)} &= \left(\sum_{k=1}^{K} w_k N_k^{(j)}\mu_k^{(j)} + w_{K+1}N_{K+1}^{(j)}\mu_{K+1}^{(j)}\right) / \left(\sum_{k=1}^{K} w_k N_k^{(j)} + w_{K+1}N_{K+1}^{(j)}\right)\\
&= \left(\mu^{(j)} + \mu_{K+1}^{(j)} w_{K+1}N_{K+1}^{(j)} / \sum_{k=1}^{K} w_k N_k^{(j)}\right) / \left[1 + w_{K+1}N_{K+1}^{(j)} / \sum_{k=1}^{K} w_k N_k^{(j)}\right]\\
&= \left[\mu^{(j)} + \gamma_{K+1}^{(j)}\mu_{K+1}^{(j)}\right] / \left[1+\gamma_{K+1}^{(j)}\right],
\end{aligned}$$

which is the first part of (8), and then from (6) again

$$\left[\sigma'^{(j)}\right]^2 = \sum_{k=1}^{K+1} w_k N_k^{(j)}\left(\left[\sigma_k^{(j)}\right]^2 + (\mu'^{(j)} - \mu_k^{(j)})^2\right) / \left(\sum_{k=1}^{K+1} w_k N_k^{(j)}\right).$$

Since

$$\begin{aligned}
&\sum_{k=1}^{K} w_k N_k^{(j)}\left(\left[\sigma_k^{(j)}\right]^2 + (\mu'^{(j)} - \mu_k^{(j)})^2\right) = \sum_{k=1}^{K} w_k N_k^{(j)}\left((\sigma_k^{(j)})^2\right.\\
&\quad \left. + (\mu^{(j)} - \mu_k^{(j)})^2 + (\mu'^{(j)} - \mu^{(j)})^2 + 2(\mu'^{(j)} - \mu^{(j)})(\mu^{(j)} - \mu_k^{(j)})\right)\\
&= \left[(\sigma^{(j)})^2 + (\mu'^{(j)} - \mu^{(j)})^2\right]\cdot\left[\sum_{k\le K} w_k N_k^{(j)}\right] + 2(\mu'^{(j)} - \mu^{(j)})\sum_{k\le K} w_k N_k^{(j)}(\mu^{(j)} - \mu_k^{(j)})\\
&= \left[(\sigma^{(j)})^2 + (\mu'^{(j)} - \mu^{(j)})^2\right]\cdot\left[\sum_{k\le K} w_k N_k^{(j)}\right],
\end{aligned}$$

we have

$$\begin{aligned}
&\left[\sigma'^{(j)}\right]^2\cdot\left(\sum_{k=1}^{K+1} w_k N_k^{(j)} + w_{K+1}N_{K+1}^{(j)}\right) = \left[(\sigma^{(j)})^2 + (\mu'^{(j)} - \mu^{(j)})^2\right]\\
&\quad \times\left[\sum_{k\le K} w_k N_k^{(j)}\right] + w_{K+1}N_{K+1}^{(j)}\left[(\sigma_{K+1}^{(j)})^2 + (\mu'^{(j)} - \mu_{K+1}^{(j)})^2\right]
\end{aligned}$$

which leads to the second equation of (8). For the update of μ and σ to μ′ and σ′, we can make use of (3) and (4), rather than (7), to derive similar to (8) the following formulas

$$\begin{aligned}
&\mu' = \left[\mu + \gamma_{K+1}\mu_{K+1}\right] / \left[1+\gamma_{K+1}\right], \gamma_{K+1} = w_{K+1}N_{K+1} / \sum_{k=1}^{K} w_k N_k ,\\
&\sigma'^2 = \left[\sigma^2 + (\mu' - \mu)^2 + \gamma_{K+1}\left[(\sigma_{K+1})^2 + (\mu' - \mu_{K+1})^2\right]\right] / \left[1+\gamma_{K+1}\right] .
\end{aligned} \tag{9}$$

It is perhaps worth noting here that iterative formulas such as those in (8) and (9) may be represented in terms of weighted finite automata [17] as well.

3 Measuring Markers' Biases

We now examine how to effectively profile the markers' biases exhibited in the past marked K assignments. First if a single assessor marked both the k-th assignment and the k'-th assignment for the whole cohort, we have to expect that the statistics for these two different assignments will in general differ due to the different nature or complexity of the assignments. In order to homogenize the statistics across the marks for the different assignments, we choose a base pair $\bar{\mu}$ and $\bar{\sigma}$, and normalize the marks for each assignment via

$$x^*{}_{k,i}^{(j)} = \bar{\mu} + (x_{k,i}^{(j)} - \mu_k)\,\bar{\sigma}/\sigma_k,$$
$$\forall j\in J,\ i=1,..,N_k,\ k=1,..,K \tag{10}$$

which implies also $x^*_{k,i} = \bar{\mu} + (x_{k,i} - \mu_k)\,\bar{\sigma}/\sigma_k$. For the rescaled marks $\{x^*{}_{k,i}^{(j)}\}$ we will calculate the statistics $\mu^*{}_k^{(j)}$, $\sigma^*{}_k^{(j)}$, $\mu^{*(j)}$, $\sigma^{*(j)}$, μ^*_k, σ^*_k, μ*, and σ* respectively in parallel to those without the *'s. These *-ed statistics constitute the bias profiling for all the markers, and $\mu^* = \bar{\mu}$ and $\sigma^* = \bar{\sigma}$. Technically, our proposed marker reassignment procedure uses $\{x_{k,i}^{(j)}\}$ to measure the biases accumulated over the marked assignments, uses $\{x^*{}_{k,i}^{(j)}\}$ to predict the biases in the next assignment to be marked, and then reallocates the markers so that the combined biases are as evenly spread out as possible across all the students. In fact, $\mu^{*(j)}$ and $\sigma^{*(j)}$ can be explicitly represented by those non-starred statistics via

$$\mu^{*(j)} = \bar{\mu} + \bar{\sigma}\cdot\sum_{k=1}^{K}\alpha_k^{(j)}\cdot(\mu_k^{(j)} - \mu_k)/\sigma_k ,$$

$$\alpha_k^{(j)} = w_k N_k^{(j)} / \sum_{k'=1}^{K} w_{k'} N_{k'}^{(j)} ,$$

$$\begin{aligned}
[\sigma^{*(j)}]^2 = \bar{\sigma}^2\cdot\Big\{&\sum_{k=1}^{K}\alpha_k^{(j)}[\sigma_k^{(j)}/\sigma_k]^2\\
&+ \sum_{k=1}^{K}\alpha_k^{(j)}[(\mu_k^{(j)} - \mu_k)/\sigma_k]^2\\
&- [\sum_{k=1}^{K}\alpha_k^{(j)}(\mu_k^{(j)} - \mu_k)/\sigma_k]^2\Big\} .
\end{aligned} \tag{11}$$

The 1st half of (11) follows immediately from (6) and

$$\mu^{*(j)} = \sum_{k=1}^{K} \sum_{i\in I_k^{(j)}} w_k x^{*(j)}_{k,i} \Big/ \left(\sum_{k=1}^{K} w_k \,|\, I_k^{(j)} \,|\right)$$
$$= \bar{\mu} + \bar{\sigma} \cdot \left(\sum_{k=1}^{K} \sum_{i\in I_k^{(j)}} w_k (x_{k,i}^{(j)} - \mu_k)\bar{\sigma} / \sigma_k\right) \Big/ \sum_{k=1}^{K} w_k N_k^{(j)} ,$$

which made use of (10). For the 2nd half of (11), we first observe from (10) and the 1st half of (11)

$$[\sigma^{*(j)}]^2 / \bar{\sigma}^2$$
$$= \sum_{k=1}^{K} \sum_{i\in I_k^{(j)}} w_k \cdot (x^{*(j)}_{k,i} - \mu^{*(j)})^2 \Big/ \left(\sum_{k=1}^{K} w_k N_k^{(j)}\right)$$
$$= \sum_{k=1}^{K} \sum_{i\in I_k^{(j)}} w_k [(x_{k,i}^{(j)} - \mu_k)/\sigma_k - \sum_{m=1}^{K} \alpha_m^{(j)} (\mu_m^{(j)} - \mu_m)/\sigma_m]^2 / N^{(j)}$$
$$= (\mathrm{I} + \mathrm{II} + \mathrm{III}) / N^{(j)} \qquad (12)$$

where terms I, II and III are

$$\mathrm{I} = \sum_{k=1}^{K} \sum_i w_k [(x_{k,i}^{(j)} - \mu_k)/\sigma_k]^2 ,$$
$$\mathrm{II} = \sum_{k=1}^{K} \sum_i w_k [\sum_{m=1}^{K} \alpha_m^{(j)} (\mu_m^{(j)} - \mu_m)/\sigma_m]^2 ,$$
$$\mathrm{III} = -2 [\sum_{k=1}^{K} \sum_i w_k (x_{k,i}^{(j)} - \mu_k)/\sigma_k] \cdot [\sum_{m=1}^{K} \alpha_m^{(j)} (\mu_m^{(j)} - \mu_m)/\sigma_m]$$

Expanding $(x_{k,i}^{(j)} - \mu_k)^2 = [(x_{k,i}^{(j)} - \mu_k^{(j)}) + (\mu_k^{(j)} - \mu_k)]^2$ one has

$$\mathrm{I} = \sum_{k=1}^{K} \sum_i w_k (x_{k,i}^{(j)} - \mu_k^{(j)})^2 / \sigma_k^2 + \sum_{k=1}^{K} w_k N_k^{(j)} (\mu_k^{(j)} - \mu_k)^2 / \sigma_k^2$$
$$+ \sum_{k=1}^{K} w_k \left((\mu_k^{(j)} - \mu_k)/\sigma_k^2\right) \cdot \left(\sum_{i\in I_k^{(j)}} (x_{k,i}^{(j)} - \mu_k^{(j)})\right)$$
$$= \sum_{k=1}^{K} w_k N_k^{(j)} \sigma_k^{(j)2} / \sigma_k^2 + \sum_{k=1}^{K} w_k N_k^{(j)} (\mu_k^{(j)} - \mu_k)^2 / \sigma_k^2 .$$

Summing up over i one has also

$$\mathrm{II} = \left(\sum_{k=1}^{K} w_k N_k^{(j)}\right) \left(\sum_{m=1}^{K} \alpha_m^{(j)} (\mu_m^{(j)} - \mu_m)/\sigma_m\right)^2$$
$$\mathrm{III} = -2 \left(\sum_{k=1}^{K} w_k N_k^{(j)} (\mu_k^{(j)} - \mu_k)/\sigma_k\right) \cdot \left(\sum_{m=1}^{K} \alpha_m^{(j)} (\mu_m^{(j)} - \mu_m)/\sigma_m\right)$$
$$= -2 \times (\mathrm{II}).$$

Hence (12) is simplified to the 2nd half of (11). We note that the important formulae or relationships such as (11), (8), (7), (6) and (2) are also verified numerically.

Let the tilded statistics $\tilde{\mu}^{(j)}$, $\tilde{\sigma}^{(j)}$ denote those for the marks of the (K+1)-th assessment item. Then they can be derived from

$$\tilde{\mu}^{(j)} = \tilde{\mu} + (\mu^{*(j)} - \bar{\mu})\tilde{\sigma}/\bar{\sigma} , \quad \tilde{\sigma}^{(j)} = \sigma^{*(j)} \cdot \tilde{\sigma}/\bar{\sigma} , \qquad (13)$$

and may also make use of $\mu^* = \bar{\mu}$, $\sigma^* = \bar{\sigma}$, and $\tilde{\mu} = \mu$, $\tilde{\sigma} = \sigma$ due to respectively (10) and the marks prediction. To show this, we let $z_i^{(j)}$ denote the marks for the (K+1)-th assignment by the j-th marker, and $z^{*(j)}_i$, the normalised marks as in (10). Then

$$z_i^{(j)} = \tilde{\mu} + (z^{*(j)}_i - \bar{\mu})\, \tilde{\sigma}/\bar{\sigma} ,$$
$$\tilde{\mu}^{(j)} = \sum_{i\in I_{K+1}^{(j)}} z_i^{(j)} / |I_{K+1}^{(j)}|$$
$$= \tilde{\mu} + (\sum_{i\in I_{K+1}^{(j)}} z^{*(j)}_i / |I_{K+1}^{(j)}| - \bar{\mu})\tilde{\sigma}/\bar{\sigma} , \qquad (14)$$

which proves the first half of (13). Since (13) and (14) imply

$$z_i^{(j)} - \tilde{\mu}^{(j)} = (z^{*(j)}_i - \mu^{*(j)})\tilde{\sigma}/\bar{\sigma} ,$$
$$[\tilde{\sigma}^{(j)}]^2 = \sum_{i\in I_{K+1}^{(j)}} (z_i^{(j)} - \tilde{\mu}^{(j)})^2 / |I_{K+1}^{(j)}|$$
$$= [\tilde{\sigma}/\bar{\sigma}]^2 \cdot \sum_{i\in I_{K+1}^{(j)}} (z^{*(j)}_i - \mu^{*(j)})^2 / |I_{K+1}^{(j)}|$$
$$= [\tilde{\sigma}/\bar{\sigma}]^2 \cdot [\sigma^{*(j)}]^2,$$

the second half of (13) also holds. We note that the homogenization via (10) is essentially redundant if K=1, and this is because having just 1 past assessment item offers no room for the marking inconsistency to be ironed out.

Now that we know how to predict the statistical behaviour or bias of each marker for the (K+1)-th assignment, we also need to measure the accumulation of such biases for each marker. For a given student i, if marker j' is assigned to mark the student's (K+1)-th assignment, then the total amount of the accumulated mean biases can be estimated as $\Delta_\mu(i, j', w_{K+1})$ where

$$\Delta_\mu(i, j', w') = w'(\tilde{\mu}^{(j')} - \tilde{\mu}) + \Delta_i \mu,$$
$$\Delta_i \mu = \sum_{k=1}^{K} w_k (\mu_k^{\rho_k(i)} - \mu_k) , \qquad (15)$$

and $\mu_k^{\rho_k(i)} - \mu_k$ becomes 0 if student i didn't submit the k-th assignment. As for the standard deviation, the meaning of $\sigma' \equiv \sigma \pm \delta$ is not the same for the different signs even though $|\sigma' - \sigma|$ remains the same in both cases. To better gauge the closeness of σ and σ', we measure both the difference $|\sigma' - \sigma|$ and the overlap, denoted by $\wp(\sigma,\sigma') \equiv \min(\sigma,\sigma')$. Hence the total amount of the accumulated bias widths can be modelled by $\Delta_\sigma(i, j', w_{K+1})$ defined for $0 \le \lambda \le 2$ by

$$\Delta_\sigma(i, j', w') = |\Delta^+_\sigma - \Delta^-_\sigma| + (2-\lambda) \cdot \wp(\Delta^+_\sigma, \Delta^-_\sigma), \qquad (16)$$

where $\Delta^\pm_\sigma(i,j',w')$ are respectively the sum of the positive and negative terms in

$$\Delta^\oplus_\sigma(i, j', w') \equiv w'(\tilde{\sigma}^{(j')} - \tilde{\sigma})/\tilde{\sigma} + \Delta^+_i \sigma - \Delta^-_i \sigma \qquad (17)$$

for $w' = w_{K+1}$, and $\Delta^\pm_i \sigma \ge 0$ are respectively the sum of the positive and negative terms in

$$\sum_{k=1}^{K} w_k (\sigma_k^{\rho_k(i)} - \sigma_k)/\sigma_k , \qquad (18)$$

in which $\sigma_k^{\rho_k(i)}$ is just σ_k if student i didn't submit the k-th assignment. We note that λ=0 corresponds to the obvious case of $\Delta_\sigma = \Delta^+_\sigma + \Delta^-_\sigma$, λ=1 to $\Delta_\sigma=\max(\Delta^\pm_\sigma)$, while λ=2 corresponds to the other extreme case of $\Delta_\sigma = |\Delta^+_\sigma - \Delta^-_\sigma|$. The role of the denominators in (17) and (18) is to make the calculation percentage-wise. Also, (15) and (18) have made the implicit use of

$$\mu_k^{(0)} \equiv \mu_k,\ \sigma_k^{(0)} \equiv \sigma_k,\ \tilde{\mu}^{(0)} \equiv \tilde{\mu},\ \tilde{\sigma}^{(0)} \equiv \tilde{\sigma}, \qquad (19)$$

which implies $\Delta_i\mu = \Delta_i\sigma = 0$ holds for any student i if that student has not submitted any of first K assignments. We finally note that Δ_μ and Δ_σ are independent of $\bar{\mu}$ and $\bar{\sigma}$ according to (11) and (13). However, using fixed $\bar{\mu}$ and $\bar{\sigma}$ will make the values $x^*_{k,i}{}^{(j)}$ unchanged as K changes. In the case of choosing simply $\bar{\mu}=\mu$ and $\bar{\sigma}=\sigma$, we can rewrite (15), (17) and $\Delta_\sigma^\pm$ in (16) more explicitly as

$$\Delta_\mu(i,j,w) = \Delta_\mu^{(j)} + \Delta_i\mu, \quad \Delta_\mu^{(j)} = w(\mu^{*(j)} - \mu),$$
$$\Delta_\sigma^\oplus(i,j,w) = \Delta_\sigma^{(j)+} - \Delta_\sigma^{(j)-} + \Delta_i^+\sigma - \Delta_i^-\sigma,$$
$$\Delta_\sigma^{(j)+} - \Delta_\sigma^{(j)-} = w(\sigma^{*(j)} - \sigma)/\sigma, \ \Delta_\sigma^\pm = \Delta_\sigma^{(j)\pm} + \Delta_i^\pm\sigma, \quad (20)$$

where $\Delta_\sigma^{(j)-}=0$ and $\Delta_\sigma^{(j)+}=w(\sigma^{*(j)}-\sigma)/\sigma$ if $\sigma^{*(j)}\geq\sigma$, and $\Delta_\sigma^{(j)+}=0$ and $\Delta_\sigma^{(j)-}=w(\sigma^{*(j)}-\sigma)/\sigma$ if $\sigma^{*(j)}\leq\sigma$. Moreover, we can set in (20) $\mu^{*(j)}=\mu$ and $\sigma^{*(j)}=\sigma$, similar to (19), for every tutor j who hasn't marked any previous assignments.

4 Markers' Reallocation

By following the main procedure outline in Figure 1, we can now proceed to design the marker reallocation algorithm. For the clarity of the algorithm, we first specify formally the input, which can be symbolically represented by

(I.1) $w_k>0$, $\rho_k: I\to J^*$, $\nu_k: I\to\mathbb{R}^+$, k=1, .., K; $w_{K+1}>0$

(I.2) $I'\subset I$, $N'^{(j)}\geq 0$, $\sum_{j\in J} N'^{(j)} \geq |I'|$, (21)

where I' denotes the set of students whose new, i.e. (K+1)-th, assignment need to be allocated to a suitable marker, and the allocation quota $N'^{(j)}$ denotes the maximum number of students whose new assignment could be contracted for being marked by the j-th marker. We note that providing the mapping ν_k is essentially the same as providing all the marks $\{ x_{k,i}^{(j)} \}$. The expected output of the allocation algorithms is then a marker-assigning mapping $\rho: I\to J^*$ such that

(O.1) $\rho(I')\subset J$, $\rho(I-I')\subset\{0\}$;

(O.2) $|\rho^{-1}(j)|\leq N'^{(j)}$, $\forall j\in J$;

(O.3) $\exists\ 1\geq\kappa>0$ such that $|\rho^{-1}(j)|\approx\kappa\cdot N'^{(j)}$, $\forall j\in J$, (22)

where (O.1) implies the new assignment by the students in I' will be assigned to the markers in J, and the rest of the students didn't submit the new assignment. (O.2) simply says the total number of students assigned to the j-th marker should not exceed the quota $N'^{(j)}$. And (O.3) is actually optional, and implies that the total number of students assigned to each marker should be proportional to their quota, if not all their quotas can be exactly met.

When students are being one-by-one assigned a suitable marker for the new assignment, students who suffered the most "biases" accumulated over the past K assignments should be given higher priority in finding the most suitable marker so as to best compensate the past marking biases. For this purpose, we let Ψ be an ordered list of the set I', sorted in the decreasing order of

$$(\ |\Delta_i\mu| + \tau\cdot|\Delta_i\sigma|, |\Delta_i\sigma|\ ) \qquad (23)$$

for a given user-selected coupling factor $\tau\geq0$, where $|\Delta_i\mu|$ is shown in (15) and

$$\Delta_i\sigma = |\Delta^+_i\sigma - \Delta^-_i\sigma| + (2-\lambda)\,\wp(\Delta^+_i\sigma, \Delta^-_i\sigma) \qquad (24)$$

is defined along the same line as (16) and (17). In the case of two students having exactly the same ranking, the student with a higher total weight of the previously submitted assignments will be ahead of the other student. Hence the students who didn't submit any previous assignments will sit at the bottom of the list.

For a fairer scheme, we also randomise the ordering of the elements of Ψ within the same band, i.e. the elements of the same $|\Delta_i\mu|$ and $|\Delta_i\sigma|$. For each marker $j\in J$, we use the set $\Psi^{(j)}$ to collect the students assigned to the j-th marker, and use $M^{(j)}$ to denote the total number of elements in $\Psi^{(j)}$, i.e. $M^{(j)}=|\Psi^{(j)}|$. We thus initialise the $\Psi^{(j)}$ by setting $M^{(j)}$ =0 and $\Psi^{(j)}=\emptyset$. The marker reallocation procedure is to continuously assign the top element on the student list Ψ to a marker, then removing the student fromΨ, until the list Ψ becomes empty.

Let $i\in I$ be the top element on the list Ψ, and let $||i||\equiv|\Delta_i\mu|+|\Delta_i\sigma|$. If $||i||$=0, i.e. $\Delta_i\mu=\Delta_i\sigma=0$, we can just randomly associate the remaining students on the list Ψ with the markers so that each marker j marks the new assessment item for the allocated $N'^{(j)}$ number of students or less. More precisely, For all $j\in J$ such that $M^{(j)} < N'^{(j)}$, we can randomly pick $N'^{(j)}-M^{(j)}$ students Φ from Ψ, assign their new assignment to be marked by the j-th marker, and then remove them from Ψ.

If $||i||\neq 0$, we find a $j'\in J$ from all those $j\in J$ such that $M^{(j)} < N'^{(j)}$ and $|\Delta_\mu(i,j',w')|$ is the smallest, and then assign student i's new assignment to the j'-th marker. If we allow a certain degree of error tolerance $\varepsilon\geq 0$ for the $\Delta_\mu(i,j',w')$, i.e. trading ε amount of $\Delta_\mu(i,j',w')$ for a better $\Delta_\sigma(i,j',w')$, then we find a $j''\in J$ from all those $j\in J$ such that

$$M^{(j)} < N'^{(j)},\ \ |(\ |\Delta_\mu(i,j',w')| - |\Delta_\mu(i,j'',w')|\ )| \leq \varepsilon \qquad (25)$$

and $|\Delta_\sigma(i,j'',w')|$ is the smallest. If $j'\neq j''$, we set $j'\neq j''$. Hence the new j' best compensates the mean as well as the standard deviation.

Another strategy to even out the potential biases is to avoid as much as possible having different assignments for the same student to be assessed by the same marker. To achieve this goal, we introduce error tolerance $\varepsilon_\mu\geq 0$ and $\varepsilon_\sigma\geq 0$ so that the allocated new marker will be as much different as those who marked the student's previous assignments. For any student i and any candidate marker j, we denote by $\chi(i,j)$ the number of student i's previous assignments marked by j. Hence we find a $j''\in J$ from all those $j\in J$ such that

$$M^{(j)}<N'^{(j)},\ \ |(\ |\Delta_\mu(i,j',w')| - |\Delta_\mu(i,j'',w')|\ )| \leq \varepsilon_\mu,$$
$$|(\ |\Delta_\sigma(i,j',w')| - |\Delta_\sigma(i,j'',w')|\ )| \leq \varepsilon_\sigma, \qquad (26)$$

and $\chi(i,j)$ is the smallest. If $j'\neq j''$, we set $j'\neq j''$.

Now that we have shown assigning marker j' to student i is the best, we can carry on with the same procedure for the next unassigned student. However, if there are students of a similar status, i.e. their k-th assignment is marked by the same marker as student i for each k with $1\leq k\leq K$, then these students can also be optimally assigned to the j-th marker for the same reason, within the marker's allocation quota. This completes the design of our main marker reallocation algorithm although (O.3) in (22) has not been addressed yet. Property (O.3) comes into consideration when markers have already been contracted to mark more than the actual number of students due to such as the student attrition. Hence (O.3) is to ensure the marking duties for the markers are reduced, if necessary, proportional to their contracted workload.

Once one selects the coupling constant $\tau\geq 0$, the error tolerance $\varepsilon\geq 0$ to compensate σ, the trade-off error tolerance $\varepsilon_\mu\geq 0$ and $\varepsilon_\sigma\geq 0$ for Δ_μ and Δ_σ, and the coupling λ for (24), all of them defaulting to 0, one can also set the pivot mean $\bar{\mu}$ (default μ) and the pivot standard deviation $\bar{\sigma}$ (default σ). If one further chooses Ψ for the list of students to be each allocated to a marker, Ψ_j for the set of students assigned to marker j, L_j for the adjusted marking quota for marker j, M_j for the number of elements currently in Ψ_j, and sets $w'=w_{K+1}$, then the algorithm is essentially composed of the following major steps.

i) Pre-processing: $L_j \leftarrow N'^{(j)}\ \forall j\in J$, and then reduce L_j proportionally so that $\sum_{j\in J} L_j = |I'|$. Round L_j up or down to the nearby integer while maintaining $\sum_{j\in J} L_j = |I'|$.
ii) Initialisation: $\Psi \leftarrow I'$; $\Psi_j \leftarrow\emptyset$, $M_j \leftarrow 0$, $\forall\, j\in J$.
iii) Calculate $\mu^{(j)}$, $\sigma^{(j)}$, μ, σ, $\Delta_i\mu$, $\Delta_i\sigma$, $\forall\, j\in J$, via (6) etc. Set $\bar{\mu}=\mu$, $\bar{\sigma}=\sigma$, and calculate $\mu^{*(j)}$, $\sigma^{*(j)}$ via for all the j's.
iv) Sort Ψ in the decreasing order of $(|\Delta_i\mu|+\tau\cdot|\Delta_i\sigma|, |\Delta_i\sigma|)$ via (15) and (20).
v) If $\Psi=\emptyset$, go to step xii).
vi) Let i be the 1st element of Ψ, set $||i||=|\Delta_i\mu|+|\Delta_i\sigma|$.
vii) If $||i||=0$, randomly allocate all the students in Ψ:
    a) $\forall j\in J$ such that $M_j < L_j$, randomly pick L_j-M_j students Φ from Ψ, set $\Psi_j=\Psi_j\cup\Phi$, $\Psi=\Psi-\Phi$, and then set $M_j=L_j$
    b) Go back to step v).
viii) Otherwise (i.e. $||i||\neq 0$), find a $j'\in J$ such that $|\Delta_\mu(i,j',w')|$ is the smallest among all those $j\in J$ with $M_j<L_j$.
ix) Find a $j''\in J$ such that $|\Delta_\sigma(i,j'',w')|$ is the smallest among all those $j\in J$ with $M_j<L_j$ and

$$|(\ |\Delta_\mu(i,j',w')| - |\Delta_\mu(i,j'',w')|\ )| \leq \varepsilon.$$

If $j'\neq j''$, set $j'=j''$.
x) Find a $j''\in J$ such that $\chi(i,j)$, defined just above (26), is the smallest among all those $j\in J$ with $M_j<L_j$ and

$$|(\ |\Delta_\mu(i,j',w')| - |\Delta_\mu(i,j'',w')|\ )| \leq \varepsilon_\mu\ ,$$
$$|(\ |\Delta_\sigma(i,j',w')| - |\Delta_\sigma(i,j'',w')|\ )| \leq \varepsilon_\sigma\,.$$

If $j'\neq j''$, set $j'=j''$.
xi) Go back to step v).
xii) Each marker $j\in J$ is assigned to mark the new assignment for those students precisely contained in Ψ_j.

5 Experiments via Simulations

We will now evaluate our assessor reallocation method in two main ways. One is to apply it to an actual unit delivery and then evaluate its impact. The other is to experiment the algorithm on the simulation data. The complete reallocation algorithm and the simulation experiments are written in the form of a single program in PERL and the Box-Muller transform [18] is used to generate random normal distributions.

We first conduct our experiments on the simulated data. Let assessor $j\in J$ be assigned to marking T_j students, and let $I=\{i: 1\le i\le T\}$ with $T=\Sigma_{j\in J} T_j$ denote all the student IDs. For simplicity, we assume these assessors have marked all the previous K different assessment items for the same number of students although who those students are may vary, and that they are now to be allocated to mark the (K+1)-th assessment item.

Assume all marks are between 0 and 10, and that the students' true marks are of the normal distributions with μ'_k and σ'_k being the mean and the standard deviation for the k-th assignment respectively. The "true" marks refer to the marks the students would receive if the most accurate marking has been conducted. Likewise, $x'_{k,i}$ and $x_{k,i}^{(j)}$ will respectively denote student i's the true and actual mark of the k-th assignment by marker j. Hence the actual marks can be generated from the true marks via $x_{k,i}^{(j)} = \mu_k^{(j)} + (x'_{k,i} - \mu'_k)\ \sigma_k^{(j)} / \sigma'_k$. If we calculate the total marks for each student, $\sum_{1\le k\le K+1} w_k{\cdot}x_{k,i}$, and then the average error against the true marks $\sum_{1\le k\le K+1} w_k{\cdot}x'_{k,i}$, we will then find that the marking reallocation does lead to less such errors.

In the marks simulations or predictions, we need to be aware that the marks for the different assignments of the same given student are most likely correlated. For any given k with $1\le k\le m$, let ξ_k be the random variable that generates the true marks for the k-th assignment for all the students. Then the random variable ξ_{m+1} that generates the true marks for the (m+1)-th assignment is likely to observe the distribution

$$\xi_{m+1} = (\alpha + \beta\Theta)\xi + \gamma\theta + \delta,$$
$$\xi = (\textstyle\sum_{1\le k\le m} w_k{\cdot}\xi_k) / (\sum_{1\le k\le m} w_k), \tag{27}$$

where the marks ξ_{m+1} is to be truncated to the marks range if necessary, Θ and θ are two random numbers generated by the unit Gaussian, parameter α is a user-selected value close to 1, and parameters β, γ, and δ are generally selected to be close to 0. The formulation of ξ_{m+1} in (27) is in fact based on the understanding that each student typically performs consistently across all the assignments. We note that if ξ has the mean μ_ξ and the standard deviation σ_ξ, then ξ_{m+1} has the mean $\alpha\mu_\xi+\delta$ and the variance $(\sigma_\xi)^2(\alpha^2+\beta^2) + (\mu_\xi)^2\beta^2 + \gamma^2$. The new marks ξ_{m+1} can be regarded as the performance perturbation of the combined marks ξ. Given the true marks for the previous assignments, it is hence possible to predict the true marks via (27) for the next assignment.

Suppose $\Phi'=(\Phi'_i)$ is the vector of the true total marks and Φ is the vector of the approximate or actual marks, we may measure the goodness of the approximation by comparing $\|\Phi-\Phi'\|$ for a vector norm $\|.\|$. However, since the relative values of the marks are often more important than the absolute values, we may make use of measurement $<\Phi,\Phi'>_m \equiv \min_{\alpha,\beta} \|(\alpha\Phi+\beta)-\Phi'\|$. We will in particular use the Euclidean distance $\|.\|_2$ as the norm and utilise thus the Lease Squares Approximation to calculate the measurement.

When dealing with a real subject delivery, one has for all assignments only the actual marks, and the "true" marks are never known. To overcome this difficulty, we will first compensate the assessors' biases on the actual marks and use these marks to substitute for the "true" marks. Hence for the actual marks $\{x_{k,i}^{(j)}\}_i$ for a given k, the rescaling towards the overall mean $x''_{k,i} = \mu_k + (x_{k,i}^{(j)} - \mu_k^{(j)})\ \sigma_k/\sigma_k^{(j)}$ will be treated as the approximate true marks. We note that this strategy has borrowed from the error estimation via the difference of consecutive iterates in a general numerical method. With this preparation, we can then evaluate the difference between the sum of the actual marks with the sum of the estimated true marks, which will constitute another error indicator termed the *predicted error*. We expect that the predicted errors will decrease under our assessor reallocation scheme.

Another way of error estimation is to make use of the marks of an assessment item that is known to be reasonably accurate across all the students in terms of the fairness, such as the marks of the final exam of which at least each question is marked by the same person for all. If such fair marks are also available for other items at different times, then the marks for an item closest to a given assignment may be used for estimate the true marks for that assignment. In this regard we will typically utilise the least squares approximation when comparing with such substitute "true" marks. More precisely, suppose the final exam marks $\{\varphi_i: i\in I'\}$ are available for all students in $I'\subseteq I$, thus $\Phi=(\varphi_i)$ being the marks vector, and $\mathbf{x}_k=(x_{k,i})_{i\in I'}$ is the marks vector for the k-th assignment for the students in I'. Since marks for the different assignments will compensate each other's marking biases under the marker reallocation scheme, we expect that $<\mathbf{x}_k+\mathbf{x}_{k'},\Phi>_m \le <\mathbf{x}_k,\Phi>_m$, and that the value $<\sum_k \mathbf{x}_k, \Phi>_m$ should be smaller compared with the case when the reallocation is not administered.

Tables I simulates the marking process and illustrates the reduction in the overall marking errors. Row M denotes the markers 1-9, row N denotes the number of students each marker will mark, $\Delta\mu$ denotes the difference of the individual mean with the given μ=7, and likewise for $\Delta\sigma$ with σ=2. It is shown in the table that the errors for the use of "new" markers are consistently smaller than those for the use of "same" markers. We note that if N marks $\{x_i\}$ are to approximate marks $\{\varphi_i\}$, then $(\sum_i |x_i-\varphi_i|^n/N)^{1/n}$ is the linear error for n=1 and is the squared error for n=2. If the approximation is in least squares, then the corresponding squared error is referred to as the least squares error. We will also make use of the predicted errors defined earlier, and illustrate their effectiveness in comparison with the other error indicators. It is perhaps worth noting here that the data in Table I for such as M, N, μ,σ, $\Delta\mu$ and $\Delta\sigma$ are made to resemble an actual unit delivery to which Table III is also related.

TABLE I. TWO ASSESSMENTS OF EQUAL WEIGHT

M	1	2	3	4	5	6	7	8	9
N	41	53	93	87	23	16	41	39	23
$\Delta\mu$	-.19	.63	.40	-.27	-.37	-.77	-.41	.05	-.20
$\Delta\sigma$	.24	-.56	-.57	.03	.52	.83	-.04	.23	-.05

Error Type	Linear Error		Squared Error		Least Squares	
Methods	same	new	same	new	same	new
Direct errors	.7920	.4397	1.090	.5087	1.082	.4880
Predicted errors	.9936	.4019	1.310	.4828	1.292	.4587

To experiment with the case of 3 assignments, we set the weight ratio to 1:1:2 and let the same group of markers to each mark the same number of students as in Table I. We conducted 1000 simulations in which all true marks are randomly generated by making use of (27). In Table II, Case A denotes that no marker reallocation is ever done, Case B denotes that the 1st and 3rd assessment items are marked by the same tutors for the same students while the 2nd item was reallocated to new markers based on the marks statistics on the 1st item, and Case C denotes that the same tutors will mark the same set of students for the 1st and 2nd items and these tutors are reallocated to different students via our algorithm to mark the 3rd assessment item. Table II shows that the marks errors (in L^2 norm) will in general be cut into about half of what they would be when no reallocation is applied. In particular, the average and minimum errors without reallocation are 0.56 and 0.50 respectively, while after the reallocation of the 3rd assignment, they reduce to 0.30 and 0.23 respectively.

TABLE II. SIMULATION ON THREE ITEMS

Items	Case A	Case B	Case C
Average	.5591	.3874	.3041
Minimum	.5028	.3325	.2343
Maximum	.6253	.5054	.4203

As the last of our experiments, we now apply our reallocation algorithm to an actual unit delivered on 3 campuses simultaneously. The assessment of the unit is composed of 2 major assignments and a final exam, along with a few less significant items. In order to reduce the noises in the error-prone data, we first remove those incomplete samples in which one or more assessment items are missing. We then remove the extremely poor–performing students too because their marks are more likely to be "irregular" and tend to not reflect the actual marks truthfully. In fact we evaluate our algorithm only on the students of complete records and in the groups of the exam marks exceeding 20, 25, and 35 respectively, out of the full mark 50. Tables III lists the average linear errors, and shows that the errors are being reduced across the board when assessment items 1 and 2 are added together, thus compensating each other's biases.

TABLE III. EXAM MARKS ADOPTED AS THE TRUE MARKS

Exam Marks	20 - 24.9	25 - 34.9	35 - 50
Item 1	5.4516	4.5484	2.8278
Item 2	4.3467	3.5108	2.7951
Items 1+2	4.0604	3.3016	2.4051

We note here that other experiments have also been conducted in support of our proposed methodology, when a number of additional controlling features and parameters are further introduced. However this is beyond the scope of our current work, and we will present them in a different work for which the critical mathematical consideration will be mostly left out for a different audience.

4 Conclusion

We have proposed an objective methodology which ensures that none of the students of multiple assessment items will suffer any irregular loss of the overall marks due to unconscious marking biases generic to individual markers. This is mainly done through reallocating suitable markers to the students for a new or next assessment item, on the basis of all assessors' marking patterns from the previously marked multiple assessment items. We have in particular developed the mathematical structure for the modeling of this problem, along with the derivation of some important statistical properties. We also devised a systematic simulation scheme to evaluate our proposed algorithm, and have demonstrated the convincing improvement on the resulting fairness on the overall assessment marks.

References:

[1] F. Greyling, M. Kara, A. Makka, and S. Van Niekert, "IT Worked for US: Online Strategies to Faciliate Learning in Large (Undergraduate) Classes", Electronic Journal of e-Learning, vol. 6, 2008, pp. 179-188.

[2] Z. Jiang, X. Guo, N. Gangavarapu, and K. Khan, "Knowledge-based Algorithms to Optimise e-Learning Outcome", The 2009 International Conference on Frontiers in Education: Computer Science and Computer Engineering, Las Vegas, 13-16 July, 2009, pp.247-253.

[3] J. Fermelis, R. Tucker, and S. Palmer, "Online Self and Peer Assessment in Large, Multi-campus, Multi-cohort Contexts", in Atkinson, Roger (eds), Ascilite 2007 : ICT: Providing Choices for Learners and Learning, pp. 271-281.

[4] C. Wall, "Grading on the Curve", InCider, vol. 5, 1987, pp. 83-85.

[5] G. Kulick, and R. Wright, "The Impact of Grading on the Curve: A Simulation Analysis", International Journal for the Scholarship of Teaching and Learning", vol. 2, no.2, 2008.

[6] M. Spear, "The Influence of Contrast Effects upon Teachers' Marks", Educational Research vol. 39, no. 2, (1997), pp. 229-33.

[7] M. E. Lunz, and T. R. O'Neill, "A Longitudinal Study of Judge Leniency and Consistency", paper presented at the Annual Meeting of the American Educational Research Association, Chicago, 24-28 March, 1997.

[8] S. Chamberlain, "Do Marking Reliability Studies have Validity?", International Association for Education Assessment (IAEA) Conference, 7-12 Sept 2008, Cambridge.

[9] A. Moira, C. Massey, J. Baird, and M. Morrissy, "Marking Consistency over Time", Research in Education, no. 67, 2002, pp. 79-87.

[10] K. Willey, and A. Gardner, "Difference in Tutor Grading in Large Classes: Fact or Fiction?", 40th ASEE/IEEE Frontiers in Education Conference, 2010, Washington, DC.

[11] K. Willey, and A. Gardner, "Improving the Standard and Consistency of Multi-tutor Grading in Large Classes", ATN Assessment Conference, 18-19 Nov 2010, Sydney.

[12] C. Torres, A. P. Lopes, L. Babo, and J. Azevedo, "Improving Multiple-Choice Questions", US-China Education Review B, vol. 8, no. 6b, 2011, pp1-11.

[13] F. Neri, "Cooperative Evolutive Concept Learning: An Empirical Study", WSEAS Transaction on Information Science and Applications, vol. 2, no. 5, 2005, pp. 559-563.

[14] W. Hines, D. Montgomery, D. Goldsman, and C. Borror, "Probability and Statistics in Engineering", ed 4, Wiley, 2003.

[15] Z. Jiang and X. Guo, "A Note on the Extension of a Family of Biothorgonal Coifman Wavelet Systems", The Australian and New Zealand Industrial and Applied Mathematics Journal, vol. 46, 2004, pp. 111-120.

[16] Z. Q. Huang, and Z. Jiang, "Image Watermarking via Sequencing Wavelet Filters", Chinese Journal of Electronics, vol. 17, no. 4, 2008, pp.649-654.

[17] Z. Jiang, O. De Vel, and B. Litow, "Unification and Extension of Weighted Finite Automata Applicable to Image Compression", Theoretical Computer Science A, vol. 302, no. 1-3, 2003, pp. 275-294.

[18] G. Box, and M. Muller, "A Note on the Generation of Random Normal Deviates", Annals of Mathematical Statistics, vol. 29, 1958, pp. 610-611.