Analysis of the Students' Representations and Solution Methods for the PF Normalisation unit
(SAMPLE)

The following pages contain the analysis and facilitation guide of the three solutions students have produced when attempting the complex problem "***Which Subject Did Lisa Outperform Her Classmates the Most?***" targeting the concept of Normalisation. The following notations are used to describe solutions:

X_i	Lisa's score
n	The number of students of a subject group
$\overline{X}_n$ and SD_n	Mean and Standard Deviation (SD) of all students in a subject group respectively
$\overline{X}_{n-1}$	The Mean of the students in a subject group, excluding Lisa's score
X_{max} and X_{min}	The maximum and minimum scores of a subject group respectively

No	RSMs	Relevant Conceptual Points	Constraints of RSM/ Misconceptions	Possible Facilitation Approaches/Questions	Possible Counter Examples
1	The group made sense of the relative standing of Lisa's scores in the subjects by examining the <u>changes in the means with and without her score,</u> i.e., $\overline{X}_n - \overline{X}_{n-1}$ The larger the mean changes in that subject, the more outstanding is Lisa's score is in that group. Using this method, students should find that the changes in mean scores for Mathematics, Science, and English are 0.64, 0.64, and 0.69 respectively. Given that, her performance is the most outstanding in English.	Students examined the *influence of the data points in their respective groups* by looking at the changes in central tendencies that are used to describe data.	M1.1 This solution *fails to consider the spread or variability* of the data. M1.2 The changes exhibited in this method are *not scaled*. The method will see glaring issues when we compare data sets of different units or metric units. M1.3 There are possible situations when there might be two data sets with equal shifts in the central tendencies after the removal of the points of interest, but have very different consistencies. The counter example E1.1 exemplifies this.	A question teachers can consider asking for this class of solutions is if two data sets have the same change of means after excluding the data of interest (e.g., the change in mean scores without Lisa's scores), does it mean that both data sets are the same? Students can be asked to think of a situation when there might be equal shifts in means, but the data sets look very different.	E1.1 Consider the following data sets: D1: **19** 14 14 14 14 9 D2: **19** 18 17 11 10 9 Both the data sets above have the same means of 14 and when both 19 are removed from the data sets, both means will drop to 13. If we want to examine the influence of the score of 19 on each of the above data sets, we will find that percentage changes of the means (with and without 19) are the same for both data sets. However, it is obvious that D2 is less consistent compared to D1, so this method does not take into account the variability in the data sets.

No	RSMs	Relevant Conceptual Points	Constraints of RSM/ Misconceptions	Possible Facilitation Approaches/Questions	Possible Counter Examples
2	The group examined the percentage change of Lisa's score from a reference point (other than the mean) in each data set. The reasoning here is that the higher the percentage changes in a data set, the more outstanding the score. Examples include (1) the percentage changes of Lisa's scores from the minimum point in the data set, i.e., $\frac{X_i\text{-}X_{min}}{X_{min}} \times 100\%$ Regarding this, the percentage changes for all three subjects are the same – 25%. (2) the percentage change of Lisa's scores from the next highest data point in the data set. The percentage changes of Lisa's scores relative to the next highest scorer for Mathematics, Science, and English are 1.12%, 1.12%, and 3.45% respectively.	Students choose some reference points to examine the positions of the data of interest. In addition, students consider relative or scaled changes. The reference points and scaling are two important aspects of normalisation.	M2.1 The scaling factors (e.g., $\overline{X}_{min}$) are not independent of the origin. Data sets with smaller scaling factor will have higher percentage changes.	Teachers may consider asking students to think of a situation when there might be an equal percentage change of a point of interest from a reference point, but the data sets look very different. The following questions may help students think about why this method does not work. Q2.1 If we translate a data set by adding a number to each of the data points, will it change the relative position of any of the data points in the data? Q2.2 If two data points have the same percentage change from the chosen reference point, does it mean the two data points are at the same position in their respective data set?	E2.1 Consider the following data sets: D1: **19** 14 14 14 14 9 D2: **33** 28 28 28 28 23 The percentages changes of the two points, 19 and 33, from the minimum points are 111.11% and 43.48%, although the ranges are the same. E2.2 (same data sets as E1.1) Consider the following data sets: D1: **19** 14 14 14 14 9 D2: **19** 18 17 11 10 9 Both D1 and D2 have the same ranges, but it is obvious that D2 is less consistent than D1.

No	RSMs	Relevant Conceptual Points	Constraints of RSM/ Misconceptions	Possible Facilitation Approaches/Questions	Possible Counter Examples
3	After finding that the mean scores of the three subjects ($\overline{X}_n$) are the same and that the distances between Lisa's scores to the respective $\overline{X}_n$ s are the same, <u>students then went on to find the SDs for each data set</u>. The SDs for the three data sets are 6.35, 4.43, and 6.00 respectively. Given that, Lisa outperformed the most for Science, since the SD for this group is the smallest, indicating a very consistent performance.	This method contains *two critical aspects of normalisation* - the distance from the mean and the spread of the data via SD.	This method *does not combine the two aspects of normalisation together*. It works only for special cases when either the distances from the mean or SDs are the same, just like the three data sets shown in the complex problem. If both the distances and SD are different, this method will not work.	Teachers may consider asking students if two data sets with different SDs can have different distances from the data point of interest to the means. Students could be asked to think of ways of combining both the distances of the data points of interest from the means of each data sets and the SDs of the data set, to determine the relative standing of the point of interest. Counter Example E3.1 may be considered.	E3.1 Consider the following data sets D1: **19** 14 14 14 14 (mean =15; SD = 2) D2: **19** 15 13 11 7 (mean = 13; SD = 4) The two data sets above have different means and SDs. The method of comparing SD only cannot work for this data sets.