WHY A PARADIGM SHIFT IN SCORING IS REQUIRED

1. With scores for testing and assessing, instructors usually mean numbers on some bounded scale between a given minimum and maximum score. We will focus here on the standard scale of percentages from 0% to 100%, with anchor points 10%, 20%, ..., 90% in between.

2. Percentages are numbers between 0 and 1, e.g. 0.70, usually multiplied by 100 and then written as 70%. A percentage of 50% represents the break-even score between passing and failing. Moreover, a score of 0% signifies the highest degree of failing and a score p = 100% signifies the highest degree of passing. So far, so good.

3. Common sense and fairness principles imply that the conventional rules of arithmetic cannot be applied to percentages. Below are some counter-examples dealing with inverses, sums and products of percentages, which clearly violate our intuition.

<table>
<thead>
<tr>
<th>− 70% &lt; 0%</th>
<th>70% + 20% = 90%</th>
<th>70% + 40% = 110%</th>
<th>2 * 70% = 140%</th>
<th>70% / 2 = 35%</th>
</tr>
</thead>
<tbody>
<tr>
<td>the inverse of 70% does not exist!</td>
<td>20% signifies failing, the sum should be &lt; 70%</td>
<td>the sum of percentages can't be &gt; 100%</td>
<td>the product of 70% with a number shall be ≤ 100%</td>
<td>two failing scores of 35% lead to a passing score!</td>
</tr>
</tbody>
</table>

5. The list of counter-intuitive results of applying ordinary arithmetic to percentages can be extended at will. Many ad hoc solutions have been proposed.

For example, the capping operations min() and max() are often applied, however, they are neither fair nor convenient to work with.

The weighted arithmetic mean of scores is often used, too, because it technically solves the problem of range violation. However, being based on ordinary arithmetic, it is just a mean trick, because it hides the other underlying anomalies.

6. Is the concept of percentage scores basically flawed and useless? Not at all. We need adequate operations!

When we adopt the rules of quasi-arithmetic, developed over almost a century and applied in many areas of measuring, we can avoid all problems mentioned above, and work out a consistent and complete scoring algebra.

Actually, we only have to replace arithmetic addition by its quasi-arithmetic counterpart:

\[
\text{the quasi-sum of } p \text{ and } q \text{ is } \frac{pq}{p+q(1-p)(1-q)}. \]

A quasi-sum lies between 0% and 100%. All other properties of addition follow easily.

Moreover, quasi-multiplication of scores with any real number r can be easily constructed in terms of this new quasi-sum.

Finally, we get a well-behaved inverse for any score p, if we multiply it by −1. Then, we will have all we need for quasi-arithmetic means and scoring rules.

7. A quasi-arithmetic scoring rule using group-peer assessment may contain an impact parameter which moderates the effect which the peer ratings will have on the group score.

Moreover, scoring rules may also include the so-called tolerance parameter which restricts the final student scores to a specified subrange of the percentage scale around the group score.

Our favorite quasi-arithmetic scoring rule, with \( \text{impact} = 1 \) and \( \text{tolerance} = 2 \), is a simple function of student rating – with the group score and a calibration factor as parameters, such that all student scores will fall within a symmetric range around the group score.

8. Conclusion: When you use the quasi-arithmetic rules of scoring, you will have no problems of working with bounded scales, e.g. the percentage scale, for group-peer assessment in SE education.

**tool:** http://www.computing.northampton.ac.uk/~gpm/#!login

**paper:** https://ieeexplore.ieee.org/document/9206197