University of Northampton

Authentic Assessment of Group Work

Dr Suraj Ajit* (Speaker)

Project Members: Paul Vossen** & Andrew Dean*

*University of Northampton, UK

**SQUIRE Research Institute, Germany

Outline

- Problem
- Shortcomings/Flaws of existing models/tools
- Our Tool (Models) – Group Project Marker (GPM)
- Demo

Main Problem

- How do we combine a tutor’s group (product) score with the peer assessment (process) scores to derive a fair (authentic) individual student score in a group project?

WebPA*

- Overshooting is possible and capping is done

```
Alice's Grade = 1.14 x 80 = 91.2%
Bob's Grade = 1.47 x 80 = 117.6% = 100% (we don't give grades above 100%)
Claire's Grade = 1.11 x 80 = 88.8%
David's Grade = 0.85 x 80 = 68.8%
Elaine's Grade = 0.41 x 80 = 32.8%
```

*https://webpaproject.lboro.ac.uk/

SparkPlus*

SPA (Performance) factor

The SPA (Performance) factor is a weighting factor that can be used to change a team mark for a project (stage) into an individual mark.

$$\text{SPA Factor} = \sqrt{\frac{\text{Total ratings for individual team member}}{\text{Average of total ratings for all team members}}}$$

Individual mark = team mark * Individual's SPA

For example, if a team's project mark was 80 out of 100 and a team member receives a SPA factor of 0.9 , they would receive an individual mark of 72 to reflect a lower than average team contribution as perceived by a combination of themselves and their peers. Alternatively, if not used to moderate summative assessment the SPA factor can be used formatively to assist student development.

In applying the SPA (Performance) factor we recommend that the maximum mark be capped at 100% to reflecting the maximum available mark for demonstrating the particular learning outcome or outcomes achievement. For example, if the project mark for a high-performing team was 95% and the highest contributor to this team received an SPA factor of 1.1, then without capping this student would receive a mark greater than 100% of the marks allocated for demonstrated achievement of the associated learning outcomes.

95% * 1.1 = ~~104.5%~~ 100%

*http://sparkplus.com.au/factors/

CATME*

Yet these three methods are not without fault. One can easily see that the scale of the derived individual scores can be quite different from that of the original group grade. In the extreme case where only one student in an *n*-member group actually made any contribution, the individualized score for that student would be *n* times larger than the group score. To alleviate this scale inflation, a constraint can be put on the maximum value of the within-group weighting factor (e.g., Kaufman, Felder & Fuller, 2000) or the derived score itself. However, what the maximum value should be is arbitrary.

*Bo Zhang & Matthew W. Ohland (2009) How to Assign Individualized Scores on a Group Project: An Empirical Evaluation, Applied Measurement in Education, 22:3, 290-308

The failure of additive scoring rules

- Let $group\ score$, $student\ rating$ and $student\ score$ be achievement measures on the percentage scale 0 … 1 for group assessment, peer assessment and individual student assessment, respectively, and related by the following scoring rule:

$$student\ score = group\ score + student\ rating$$

- Because all three measures are supposed to lie on the percentage scale, we may get inconsistent results for $student\ score$, e.g.:
  - if $group\ score = .75$ and $student\ rating = .60$ then $student\ score = 1.35$, so that we will get an invalid score $student\ score > 1.00$
- Because $student\ rating \geq 0$, $group\ score + student\ rating$ will always be equal to or larger then $group\ score$, and so will be $student\ score$. Thus there are no genuine negative scores: all scores are positive (or zero).

The failure of additive scoring rules

- We might extend the measurement scale for peer and student ratings to a scale between -1 and +1, but obviously, then we aren't talking anymore about standard percentages. Moreover, we may again get inconsistent results for $student\ scores$, e.g.:
  - if $group\ score = .60$ and $student\ rating = -.70$ then $student\ score = -.10$, which is an illegal score because $student\ score < 0.00$
- The use of a cutting operation to force student scores to always lie between 0 and 1 doesn't seem to be fair and plausible because many students would get a score of 0 or 1 although their student ratings may be quite different.

The failure of additive scoring rules

- Alternatively, we might introduce weights (weightings) w and $1 - w$ for $group\ score$ and $student\ rating$, respectively, and change the scoring rule in the following way:

$$student\ score = w \times group\ score + (1 - w) \times student\ rating$$

- All student scores will come to lie between $\min(group\ score, student\ rating)$ and $\max(group\ score, student\ rating)$, and thus be valid scores. Again, this "solution" of the underlying problem looks highly artificial. Moreover, the tutor has the unthankful job of specifying an arbitrary parameter $weight$.

Our Tool – Group Project Marking (GPM)

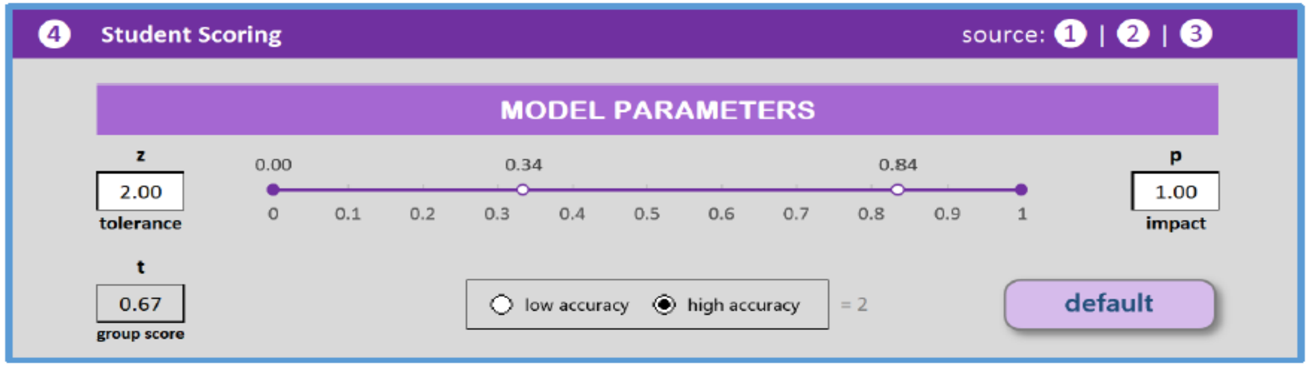
- Two models:
  - Boosted Arithmetic Student Scoring (BASS)
  - Quasi Arithmetic Student Scoring (QASS)

BASS - Boosted Arithmetic Student Scoring

- Step 1: Calculation of tutor’s group score as arithmetic mean of weighted-criteria product scores
- Step 2: Calculation of the peer ratings as arithmetic means of weighted-criteria process ratings for each student by his or her peers
- Step 3: Calculation of student ratings as arithmetic means of the peer ratings received by a student
- Step 4: Calculation of the mean student rating as the weighted arithmetic mean over all student ratings
- Step 5: Calculation of student score as a weighted arithmetic mean of group score and student rating

BASS - Boosted Arithmetic Student Scoring

- Tolerance (z) parameter (default value is 2): This denotes the extent to which student score deviates from the group (product) score.
- Impact (p) parameter (default value is 1): This denotes the effect/impact of peer assessment on the student score. If p = 0, student score = group (product) score.

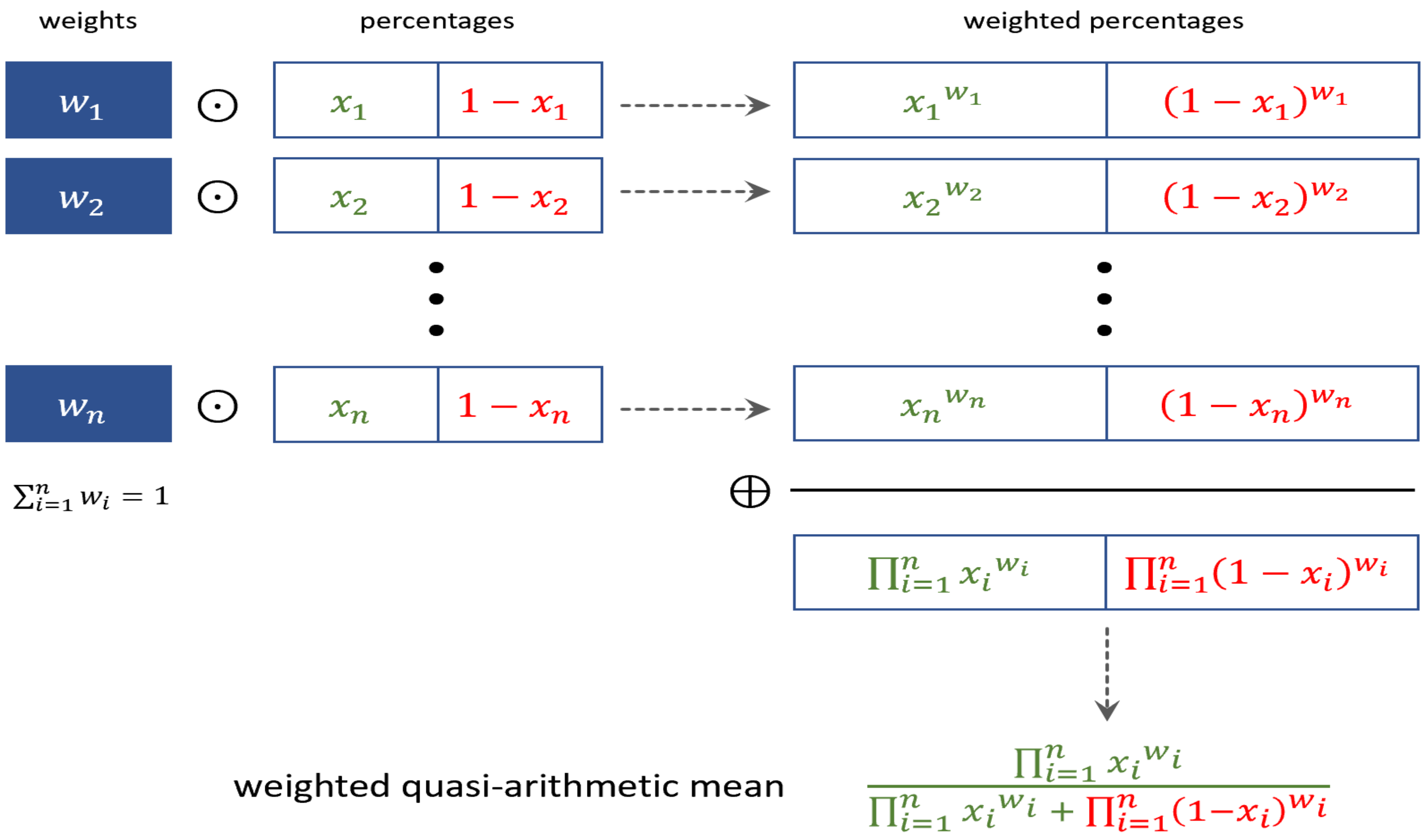


- Note: The above parameters have been introduced to counteract what many people see as a <u>*weak point of peer assessment*</u>: namely, that students may try to <u>game the system</u> by giving systematic high(er) peer ratings to their peers in order to pull up their scores.

QASS - Quasi Arithmetic Student Scoring

Rationale: The product assessment (tutor score) and process assessment (peer score) usually measure broadly distinct competences with very little overlap. In traditional arithmetic models (including BASS), we take the average/mean of both the scores. This average measures competences that are common to both product and process assessments. We leave out all evidence of student's competences that are unique to either product or process assessment. In other words, average is sensible when it is only applied to two identical measurements. Hence, we really need to calculate quasi-arithmetic mean.

QASS - Quasi Arithmetic Student Scoring

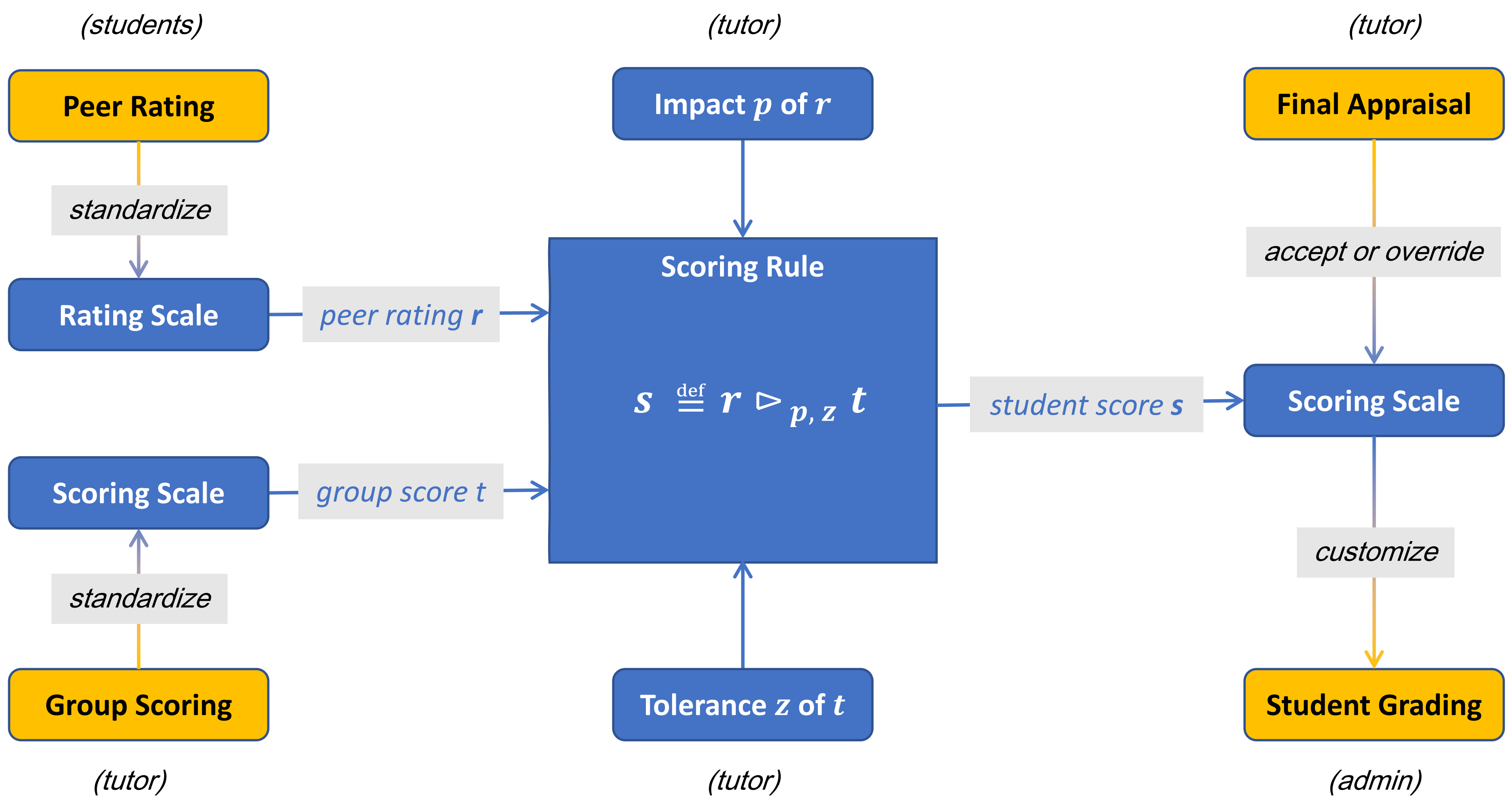


QASS - Quasi Arithmetic Student Scoring

- The Split-Join-Invariance principle (SJI) requires that if we take the average (quasi) of the student scores based on peer assessment, we will recover the original group score
- This is automatically satisfied by using QASS.
- Impact parameter (p) can still be used to moderate the students' ratings by the tutor for any excessive bias or spread.
- Tolerance parameter (z) can still be used to set the extent to which student score can deviate from group (product) score. For example, the assessment policy of university may specify that no more than 35% individual deviation from tutor's group (product) score.
- Further information:

(2020) Vossen, P.H. and Ajit, S.: Towards a Paradigm Change in Group and Peer Assessment in Software Engineering Education. In Proceedings of the 32nd IEEE International Conference on Software Engineering Education and Training (CSEE&T 2020), 9-12 November 2020, Berlin,Germany. (Accepted)

General Model



DEMO

Thank You

Contacts:

Dr Suraj Ajit suraj.ajit@northampton.ac.uk

Paul Vossen p.h.vossen@googlemail.com