Semi-automatisch verbeteren in groep

School <3 Onderzoek

Dries Vrijsen & Filip Moons

fwo

Wist je?

49%

... van de leraars binnen de Europese Unie klaagt over teveel verbetertaken!

... tweede grootste ergernis

(Eurydice, 2021)

Date

81% B- 13/16

−3

2. $63 \div s = 7$

$s = \underline{9}$

4. $32 + p = 40$

$p = \underline{8}$

6. $4c + 2 = 18$

$c = \underline{(3)}$

8. $12 - (10 - 3) = w$

$w = \underline{5}$

Show y

Te weinig feedback

Validiteitsproblemen met volledig digitale opgaven

PEN-EN-PAPIER

PP

VOLLEDIG AUTOMATISCH

FA

PEN-EN-PAPIER

PP
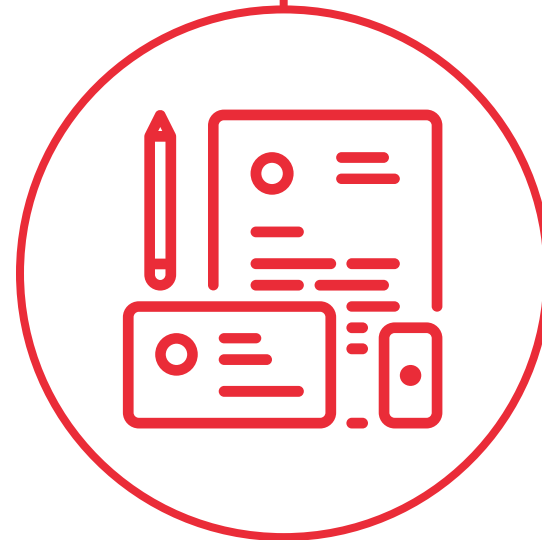
SEMI-AUTOMATISCH VERBETEREN

SA

VOLLEDIG AUTOMATISCH

FA

University of Antwerp

Eerste studie

Eerste studie

Toets op lineaire vergelijking

- 60 studenten

- 45 leraars

- *Focus op individuele leraars die feedback geven*

University of Antwerp

Classic versus atomic feedback

Manipulate the formula

$$A = 2\pi rh + 2\pi r^2 \quad \text{to } h$$

Classic feedback

Mind the fact that the dominant operation in the right-hand side of the equation is an addition! It is impossible to divide the left-hand side by 2πr because, in the first step, it is not handled as the common factor of the right-hand side. Your final answer is right, but written this way, it seems as coincidence. Going from the first to the second step, normally you would substract 2πr² from both sides, meaning that it shouldn't be placed in the nominator. It is unclear of this is an additional mistake or a compensation of the previous mistake.

Grade: 3/10

Classic versus atomic feedback

Classic feedback

Mind the fact that the dominant operation in the right-hand side of the equation is an addition! It is impossible to divide the left-hand side by 2πr because, in the first step, it is not handled as the common factor of the right-hand side. Your final answer is right, but written this way, it seems as coincidence. Going from the first to the second step, normally you would substract 2πr² from both sides, meaning that it shouldn't be placed in the nominator. It is unclear of this is an additional mistake or a compensation of the previous mistake.

Grade: 3/10

Atomic feedback

- **First step**
  - Mind the fact that the dominant operation in the right-hand side of the equation is an addition! Threshold: Max 5 out of 10 points
  - It is impossible to divide the left-hand side by 2πr because, in the first step, it is not handled as the common factor of the right-hand side.
- **Second step**
  - Your final answer is right, but:
    - Going from the first to the second step, you should substract 2πr² from both sides. -2 points
    - 2πr² shouldn't be placed in the nominator.
    - It is unclear of this is an additional mistake or a compensation of the previous mistake.

Grade: 3/10

University of Antwerp

First paper: Research questions

RQ1 Does the semi-automatic approach with reusable feedback (SA) lead to significant time savings compared to paper-based feedback (PP), and does the amount of feedback differ between the two conditions?

RQ2 Can we distinguish atomic feedback from non-atomic feedback? Can we find patterns in the reasons why the non-atomic items are non-atomic?

RQ3 How reusable is atomic feedback?

First paper

Test on lineair equations

- 60 students

- Containing
  - **1 lineair equation** *easy procedural*
  - **1 manipulation of a formula** *complex procedural*
  - **1 word problem** *problem-solving*

First paper: Methodology – Crossover experiment

Condition	Teacher 1	Teacher 2	Teacher 3	...	Teacher 45
SA	S1 S2 S3	S4 S5 S3	S4 S2 S6		S1 S5 S6
PP	S4 S5 S6	S1 S2 S6	S1 S5 S3		S4 S2 S3

University of Antwerp

15

First paper: Methodology & Results

RQ1

No significant time difference between PP and SA
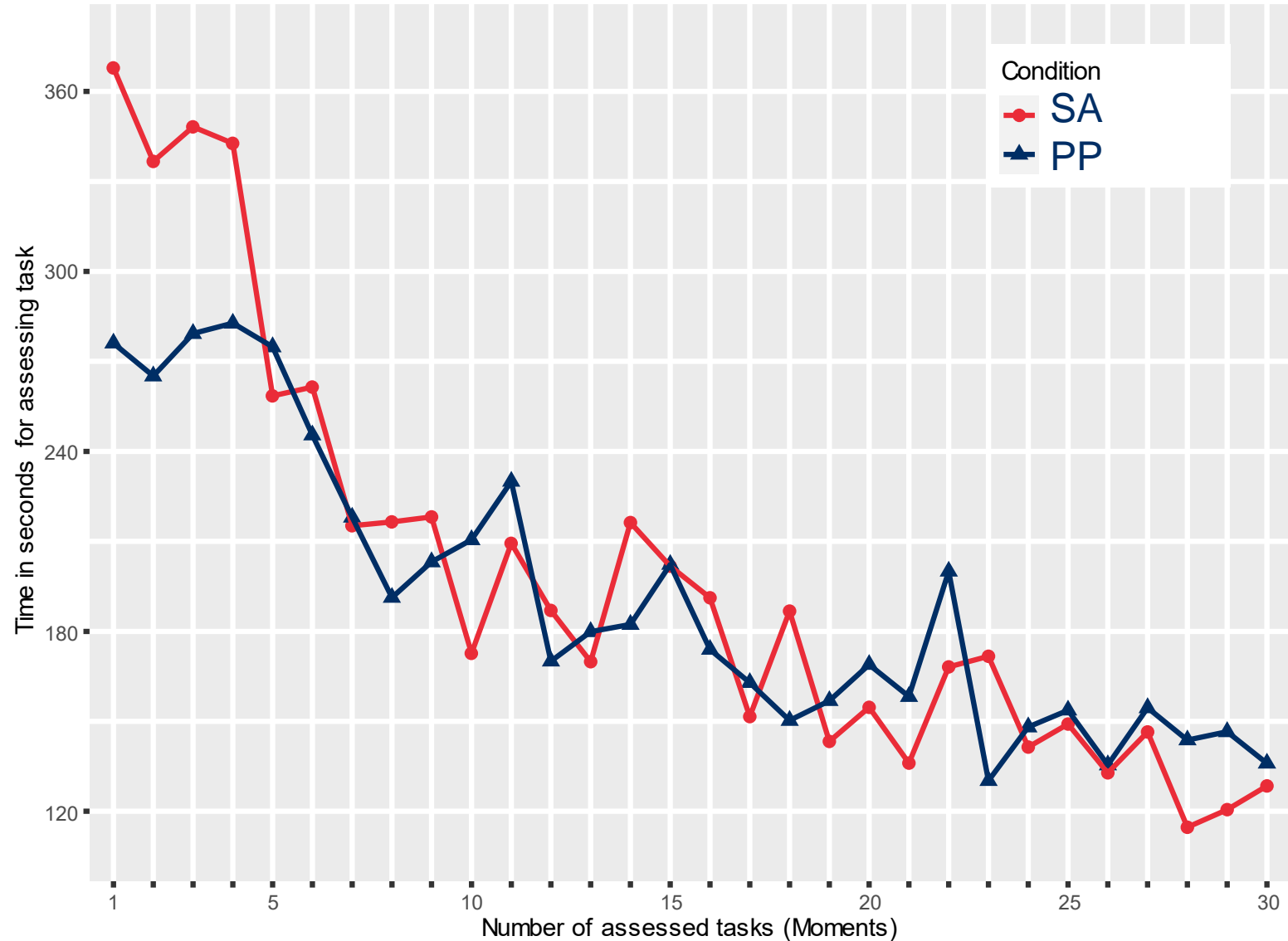
First paper: Methodology & Results

RQ1

Significant difference in amount of feedback, medium effect size ($d = 0.41$)

First paper: Methodology & Results

RQ2

- More qualitative approach
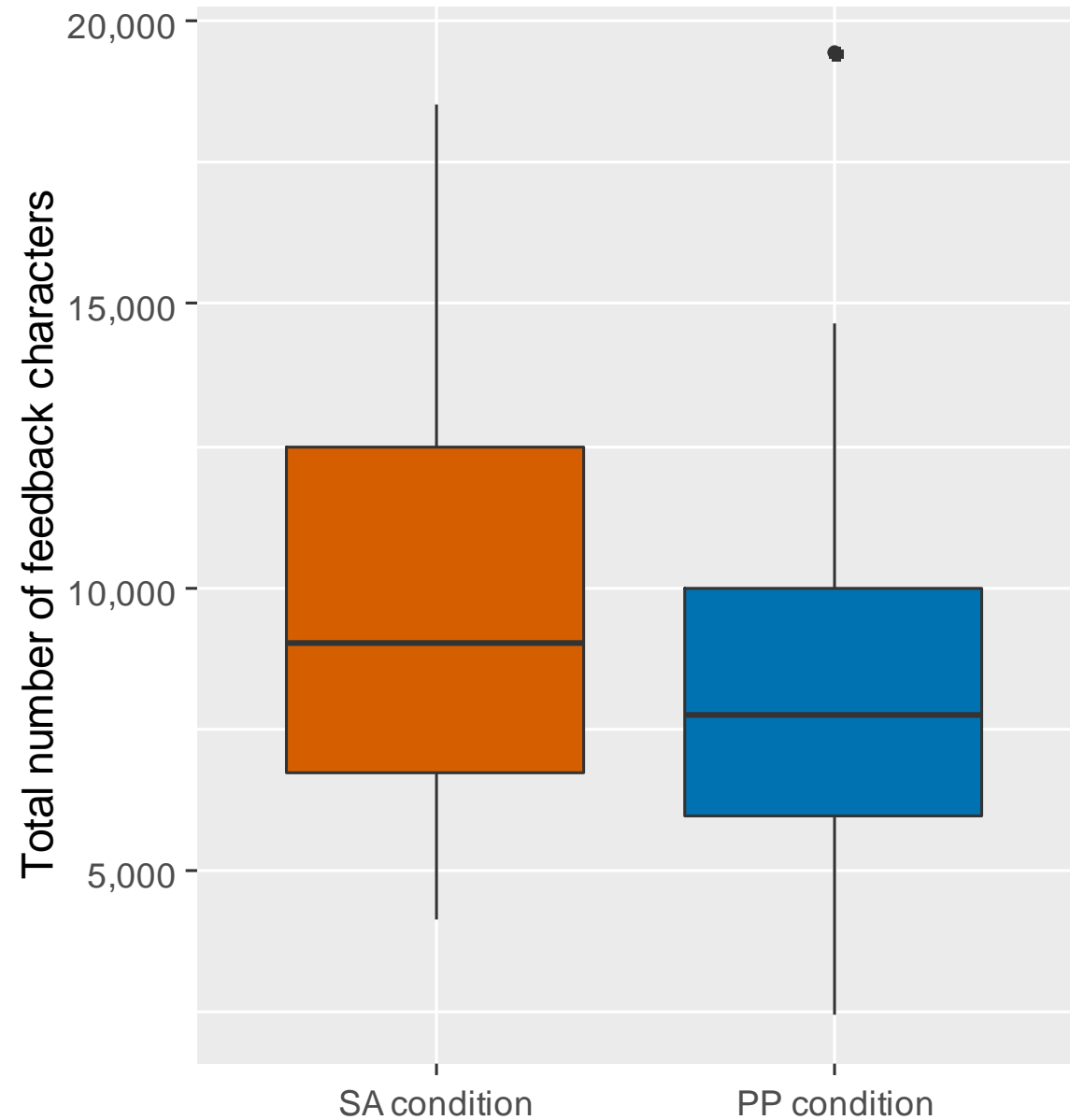- Coded every feedback item from the participating teachers as 'atomic' or 'not-atomic'
- Two coders (me & student worker)
- Codebook is on the Google Drive

	Cohen's κ	% agreement	Level of agreement	Coded items
1	0.53	81.2	Moderate	First 1000 items
2	0.70	88.0	Substantial	100 random items
3	0.66	86.8	Substantial	Full dataset (2,591 items)
4	0.84	93.5	Almost perfect	Full dataset (2,591 items)

University of Antwerp

First paper: Methodology & Results

RQ3

Items classified as atomic were significantly more reused than the non-atomic items (odds ratio = 2.6).

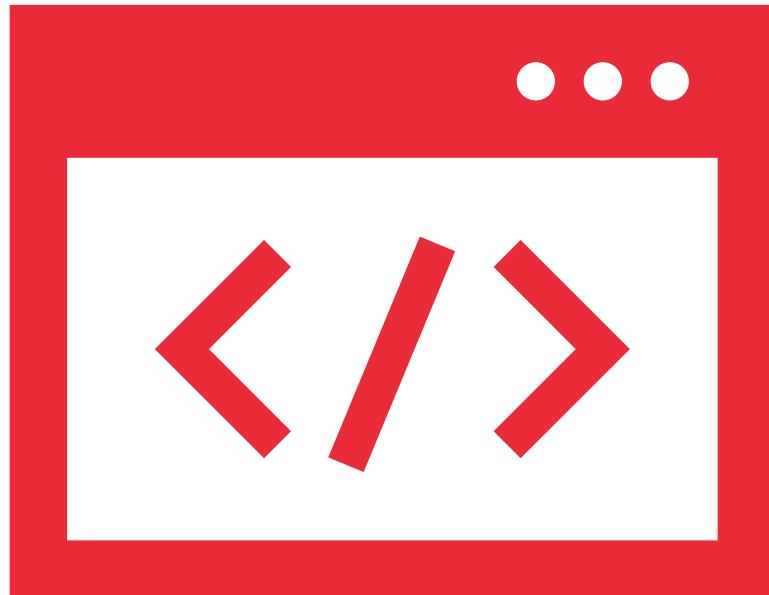| Atomicness	Reusability		Total
	Reused	Not reused	
Atomic item	731 (40.9%)	1,055 (59.1%)	1,786 (73.7%)
Non-atomic item	131 (20.5%)	507 (79.5%)	638 (26.3%)
Total	862 (36.6%)	1,562 (64.4%)	2,424 (100%)

Tweede studie

Tweede studie

- Kan dit ook werken in groep?
- Dynamisch verbetermodel, gedeeld tussen correctoren
- Feedback naar kandidaten en verbeterstijlen

Aan jullie!

Ontwikkelde Moodle plug-in

Demo

Navigeer naar

- www.mathsa.uantwerpen.be/moodle

Credentials

- Username: **[voornaam][achternaam]**
- Password: **Welkom!123**

University of Antwerp

Onderzoeksvragen

- **Correctoren**
  1. Is er een verschil in tijd?
  2. Is er een verschil in interbeoordelaarsbetrouwbaarheid?
  3. Hoe ervaarden correctoren het systeem?

Nieuw   **Traditionele verbetersysteem**

Blind   **Zichtbaar**

Onderzoeksvragen

- **Studenten**
  1. Begrijpen studenten de feedbackfiches?
  2. Hoe wenselijk vinden zijn feedbackfiches?

Methodologie

Flemish Examination Commission

3 exam designers

7 external assessors
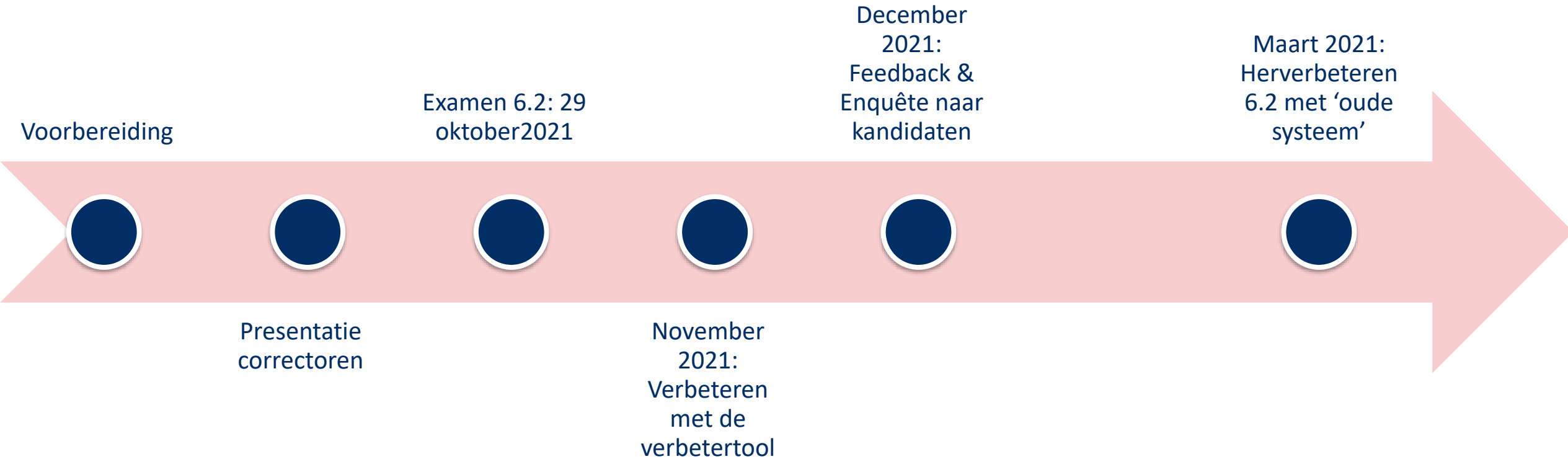
60 students

Methodologie

Voorbereiding

Presentatie correctoren

Examen 6.2: 29 oktober2021

November 2021: Verbeteren met de verbetertool

December 2021: Feedback & Enquête naar kandidaten

Maart 2021: Herverbeteren 6.2 met 'oude systeem'

Methodologie

1. Designing the system

2. Training of assessors

3. Exam day (29th Oct 2021)
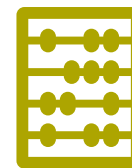Mix of digital and handwritten questions

4. Assessors asses exams
30 exams in common

4. Assessors fill in a survey
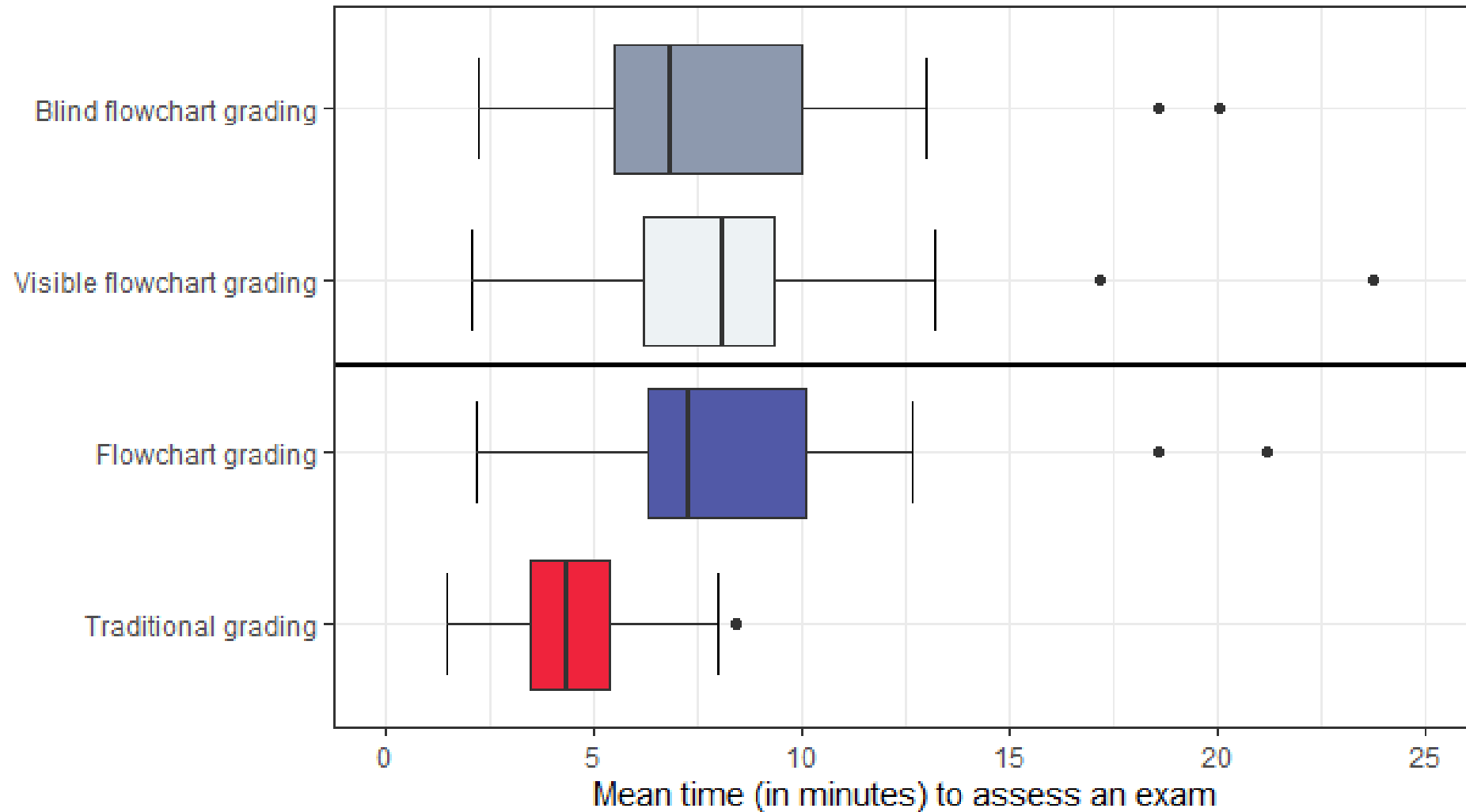
5. Students get their grades

6. Students surveys and focus interviews

7. (Mar 2022) **Re-assessing with traditional correction scheme**

University of Antwerp

27

Onderzoeksresultaten
Correctoren - Is er een verschil in tijd?

Onderzoeksresultaten

Correctoren – Is er een verschil in interbeoordelaarsbetrouwbaarheid?

Question	Overall κ	Blind κ	Visible κ	*p*-value
Q1	0.803	0.833	0.767	.198
Q2	0.641	0.812	0.687	.045*
Q3	0.490	0.520	0.420	.009**
Q4	0.785	0.723	0.873	.004**
Q5	0.835	0.909	0.760	.039*
Q6	0.473	0.394	0.586	.052
Q7	0.847	0.825	0.892	.337
Q8	0.759	0.685	0.652	.574
Q9	0.735	0.748	0.733	.828
Q10	0.862	0.901	0.829	.177
WHOLE EXAM	0.710	0.722	0.698	0.24

Onderzoeksresultaten

Correctoren – Is er een verschil in interbeoordelaarsbetrouwbaarheid?

Question	SA κ	Traditionl κ
Q1	0. 788	0.788
Q2	0.835	0.858
Q3	0.423	0.682
Q4	0.834	0.758
Q5	0. 920	0.964
Q6	0.639	0.837
Q7	0.935	0.897
Q8	0.667	0.654
Q9	0.751	0.672
Q10	0.808	0.764
WHOLE EXAM	0.683	0.632

University of Antwerp

Onderzoeksresultaten
Correctoren – Hoe ervaarden zij het systeem?

Scales	Visible SA grading M±SD	Blind SA grading M±SD
1. Perceived Usefulness	5.7 ± 0.7	4.6 ± 1.5
2. Perceived Ease of Use	5.4 ± 1.0	4.5 ± 1.4
3. Anxiety	2.5 ± 1.1	3.6 ± 1.7
4. Attitude Towards Using	6.1 ± 0.8	4.4 ± 1.7
5. Behavioral Intention to Use	5.6 ± 1.2	4.4 ±1.7

University of Antwerp

Onderzoeksresultaten

Studenten – Hoe ervaarden zij het systeem?

Students' survey item	M±SD
My feedback was too uninformative or brief to be helpful	3.6 ± 1.9
My feedback encouraged me to improve	4.7 ± 1.7
I will make even better exams based on my personal feedback	4.9 ± 1.6
This personal feedback helps me to reflect on what I have learned	5.0 ± 1.3
My feedback indicated clearly how my scores were calculated	5.5 ± 1.1
I understand most of my feedback	5.3 ± 1.4
It would be great if the Examination Commission always gave this type of feedback	6.3 ± 0.7
I feel demoralized or angry after reading my feedback	2.8 ± 1.8
The relationship between the feedback and the score is clear	5.2 ± 1.2

Dankjewel!

University of Antwerp