SAM (Short Answer Method)

I want YOU

To give ME the correct answer

Not the other way around

Projectleiding:
Theo van Mourik T.j.van.Mourik@rug.nl
Else Havik

Overview

› Why we started it
› What we did
› The results
› How you could start

Multiple choice

› Large group of students means exam of 100% MC
  - 1^e year 50 students: essay (tralala)
  - 2^e year 100 students: essay (pff)
  - 3^e year 200 students: essay (aargh)
  - 4^e year 250 students: MC (tralala)
› MC-questions are <u>often</u> good questions
› Test recognition, not reproduction

| I wrote a book on it | I know this | It's on the tip of my toungue | Now you mention it | Sounds familiar | Is that a type of shoe? |

Knowledgeability

Multiple choice

> Large group of students means exam of 100% MC
  - 1^e year 50 students: essay (tralala)
  - 2^e year 100 students: essay (pff)
  - 3^e year 200 students: essay (aargh)
  - 4^e year 250 students: MC (tralala)
> MC-questions are <u>often</u> good questions
> Test recognition, not reproduction
> Bad students can guess, good ones are corrected for guessing (=unfair)
> Paradoxal demand of answer alternatives
  - convincing yet indisputably incorrect
> How often are MC-options ambiguous? This happens…
  A. a lot
  B. often
  C. regularly

Goal

› Students

› NSE

› Teachers

› Policy and strategic plans

› More varied and challenging way of testing

› Not just MC

What we did

› Short Answer Method

› 1-3 words

› Qualitative and efficient addition to MC questions

- No recognition but reproduction

- No correction for guessing

- Direct feedback

› Automatic grading based on answerkey

› Student can suggest answers

Student gives a suggestion

The next day Anne is playing in the sand again. Another child passes on his bike and rings loudly. Anne looks expectantly at her father. Due to the learning mechanism [A] Anne expects an icecream cone again.

YOUR ANSWER

Answer

Given answer
claasical conditioning
Score

Current score

0.0

ANSWER GROUPS

Current answer key

Correct Answers - classical conditioni...	100.0
Incorrect Answers -	0.0

Title
Correct Answers

Answers
classical conditioning, classic conditioning
Feedback

MAKE A SUGGESTION

Reason

Your suggestion

○ No suggestion
◉ Only spelling differs
○ Only word choice differs
○ Other

Send in suggestion   Make Suggestion

In comes SAM

> Short Answer Method
> 1-3 words
> Qualitative and efficient addition to MC questions
  - No correction for guessing
  - No recognition but reproduction
  - Direct feedback
> Automatic grading based on answerkey
> Student can suggest answers
> Teacher judges suggestions only

Clustered answers to mark

43%

57%

- Suggested
- Not Suggested
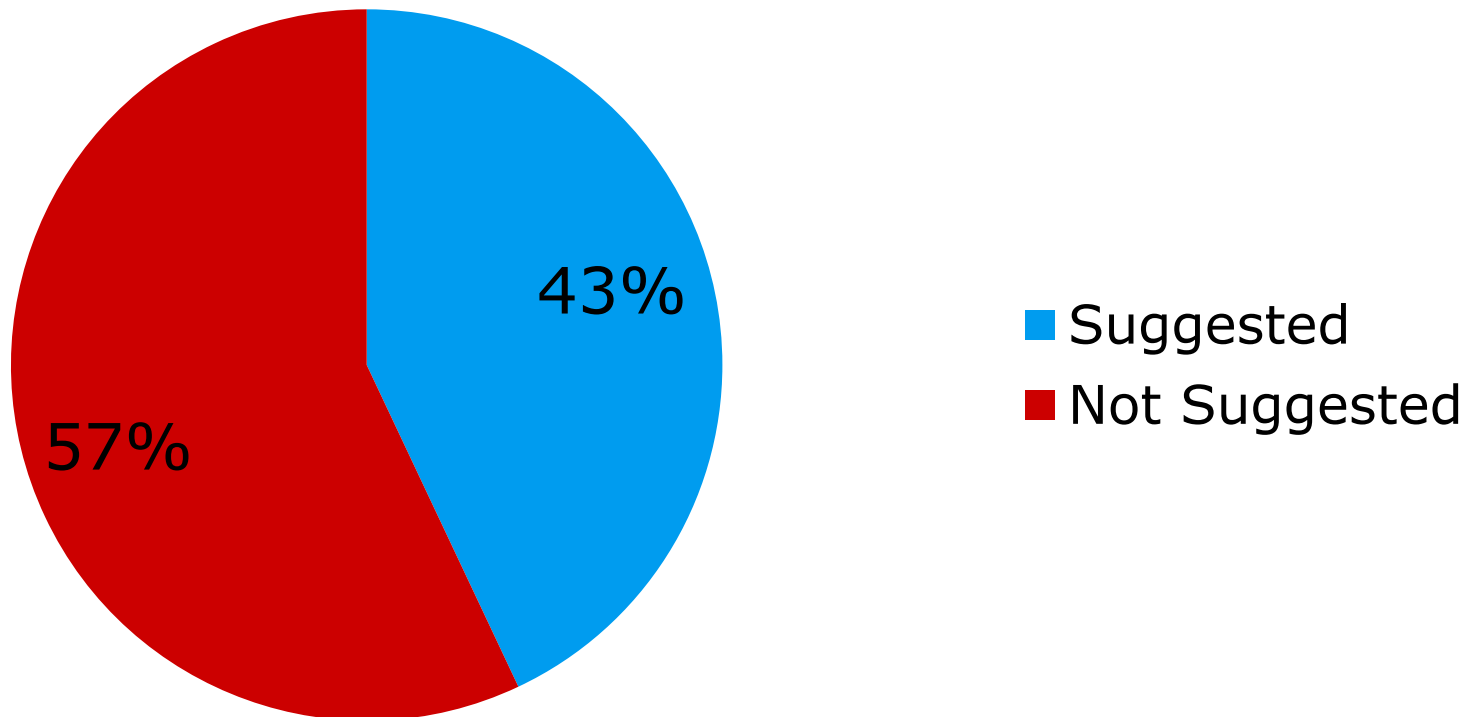
› (2 exams, 219 students)

Teacher can accept a suggestion

SUGGESTIONS

Save suggestions

		Add to:	
Only word choice differs	effect size	Action Required	0
Only word choice differs	something	Action Required	0
Only word choice differs	numbers of observations	1 (1)	1
Only word choice differs	participants	Action Required / 1 (1)	1
Only word choice differs	sample sizes	1 (1)	1
I typed in statistical power, because I thought that power might be inadequate because ther won't be 20 participants per condition.	statistical power	1 (1)	1
Not enough observations per cell lead to inadequate statistical power			
Only word choice differs			
It does have an inadequate statistical power if the sample size is low			
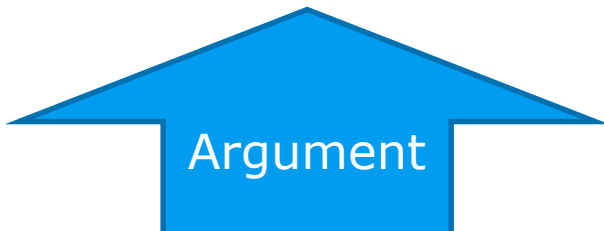
Argument

Answer

Question formats

1. The education of (___A___) is the best of the UG
2. Anne (4) expects an icecream when she hears a bell. Apparently she is (___A___) conditioned where the bell is the (___B___) stimulus and the icecream is the (___C___) stimulus.
3. To measure creativity, you can best measure the (___A___), (___B___), (___C___) and (___D___).
4. There are four levels of measurement, in order they are (___A___), (___B___), (___C___) and (___D___)
5. This video shows a boy with developmental disorder (___A___), you can recognize this by (___B___).

Question formats

1. The education of (___A___) is the best of the UG
2. Anne (4) expects an icecream when she hears a bell. Apparently she is (___A___) conditioned where the bell is the (___B___) stimulus and the icecream is the (___C___) stimulus.
3. To measure creativity, you can best measure the (___A___), (___B___), (___C___) and (___D___).
4. There are four levels of measurement, in order they are (___A___), (___B___), (___C___) and (___D___)
5. This video shows a boy with developmental disorder (___A___), you can recognize this by (___B___).

The project '18-'19

› 4 courses with exams (5/8 exams done)

› 5 courses preparing questions, 2 workshops, manual in progress

› Blackboard implementation (basic functions done)

› Evaluation research (analyzing 4th and 5th exam, writing report on cheating experiment)

The results

> Psychometrics

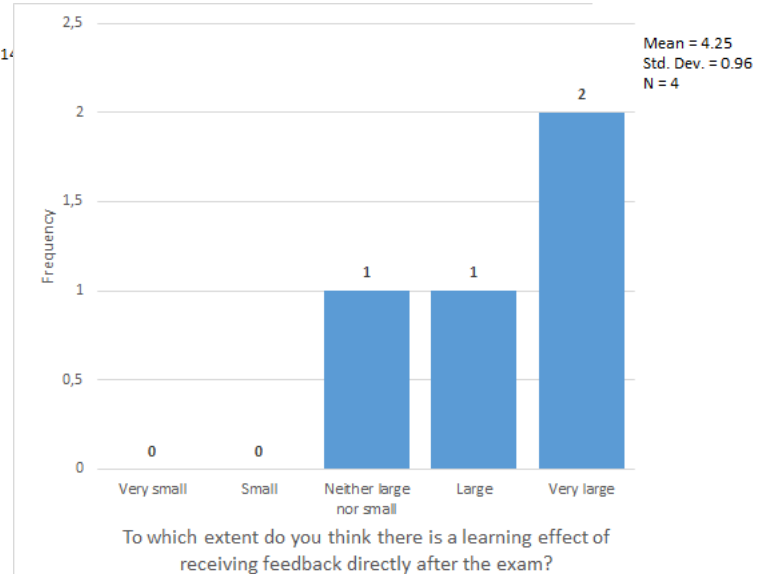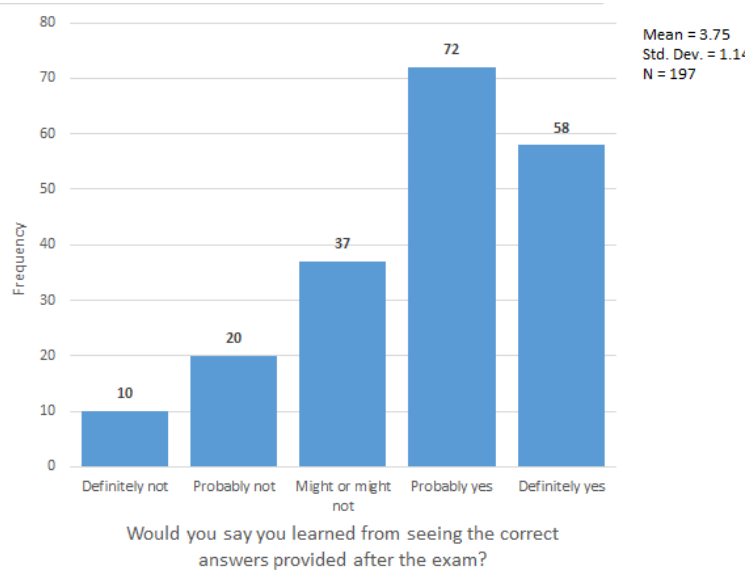> Learning effect

> Student evaluation

> Time for 'marking'

Psychometrics comparison

	PSMIN11 7-11-'18	PSMIN11 Resit 30-1-'19	PSMOB-3 25-1-'19	PSMOB-3 Resit 11-4-'19	PSBE2-08 28-3-'19
#students	173	46	55	17	269
#MC	29	31			50
#SAQ	20	18	28	30	10
#Essay			7	6	
Difficulty					
M PC MC	0,67	0,67			0,75
M P SAQ	0,63	0,52	0,54	0,50	0,80
M P Essay			0,66	TBA	
Discriminatory value					
M RIR MC	0,20 (RIT 0,30)	0,12			(RIT 0,30)
M RIR SAQ	0,44	0,41	0,28	0,20	0,47
Reliability					
MC	0,64	0,49			0,79
SAQ	0,86	0,82	0,76	0,61	0,63
Essay			TBA	TBA	
Correlation					
MC & SAQ	0,69	0,65			TBA
Essay & SAQ			0,66	TBA	

Learning effect

> Students learn more due to direct feedback
> SAM suggestions ensure students read the feedback
> Students and teachers expect a learning effect

Mean = 3.75
Std. Dev. = 1.14
N = 197

Would you say you learned from seeing the correct answers provided after the exam?

Mean = 4.25
Std. Dev. = 0.96
N = 4

To which extent do you think there is a learning effect of receiving feedback directly after the exam?

Student evaluation

SAQ, compared to MC	BEHAVE N=16	Exams N=199
General opinion on SAM 1= negative 5 = positive	3,4	3,1
Like direct feedback 1= no 5 = yes		4,2
SAQ improve exam quality 1= disagree 5 = agree	3,9	3,3

Judging suggestions

› Short answers (classic)
  - 173 students
  - 2 min (with exceptions of 8 min)
› Short answers, but discouraged suggestions
  - 269 students
  - 11 min
› Long answers up to 10 words
  - 55 students
  - 5:30 min

Short? Answer Method

› Ideally advanced SAM formats approach essay
› 2 SAQ went essay
  – Grading took 15 min per question
  – Equivalent in SAM took 2 min
› 1-3 words for 'automatic' grading
› 4-10 words still faster than with Essaymodule

› Discovering the border between SAQ and essay
› Disclaimer:
  · SAM is faster
  · Essay is usually better (for higher order questions)

How you can start

1. Be booked for a digital exam
2. Mail me T.j.van.Mourik@rug.nl
3. Consult ESI DigitaalToetsen@rug.nl
4. Create SAQ. Start small or go big
   - Just a few with the manual
   - An itempool with the manual, coaching or a workshop
5. Inform your students, demo-course, practice exam
6. Administer exam
7. Judge suggestions
8. Evaluate

Disclaimer: Support arrangement needs to be discussed

The End

Questions and discussion