

Scaling Assessment with Adaptive Comparative Judgement

Jeremy Singer
Sarah Honeychurch
Niall Barr
Steve Draper

Assessment by Pairwise Ranking

A method of ranking artefacts by making comparative judgements, rather than absolute ones.

Ranking Jokes

Your job is to pick the two best jokes by sorting the cards into order of “comedic value.”

The criterion for “comedic value” is how strongly you judge the joke would appeal to *this* audience.

Actual Jokes (for reference)

How many programmers does it take to change a lightbulb?

None – that's a hardware problem.

Don't use "beef stew" as your computer password. It's not *stroganoff*.

Teacher: was your internet research helpful for your essay assignment?

Student: Yes, I've found seven people who will write an essay for me.

How do you spot an outgoing Educational Technologist?

They look at your shoes instead of their shoes, when they talk to you.

There are 10 different kinds of educators –

those who understand digital technology and those who don't.

Teacher: Where's your homework?

Pupil: I lost it in a fight with a kid who said you weren't the best teacher in this school.

I'm not sure if they're clapping my presentation

OR because it has finished.

Your argument is sound; nothing but sound.



bareskin

Distinctive Benefits of Pollitt's ACJ Approach

- Method “**scales**”
- Compelling **naturalness**
- Can be used with **sets** of markers
- Can be used for **peer review**
- Can easily mark **cross-media** (& **multi-media**)
- Can easily be used for/with unusual, subjective, and implicit **marking criteria**
- Can be used by **matching against exemplars**
- <http://www.psy.gla.ac.uk/~steve/apr/apr.html#usp>

Adaptive Comparative Judgement

- The **software** has been built, tested, and used; and by more than one person / organisation. (Also done for conference talk refereeing at UofG.)
- A major **experiment** has been done and published, using professional markers; supporting the key claims (Pollitt, 2012).
- This paper additionally reports an important qualitative datum: that the markers were highly sceptical (did the experiment for the money, at standard professional rates for marking) but came to see it as **better as well as faster** than their traditional way of doing marking).


Our ACJ Implementation: the software

- A simple IMS LTI application that can be linked from Moodle, FutureLearn or any other LTI host.
- Submissions can be text, source code, PDFs, images or YouTube URLs.
- Submissions can be added by staff for a review only exercise, or by each student.
- Like *Moodle Workshop* and *Aropä*, it has separate submission and review phases

Our ACJ Implementation: the algorithm

- Sorting done in 'rounds'
- New pairing allocated at start of each round
- Three different phases, each with a different 'scoring' method as sort improves
- A simulation (using random errors in comparison) was used to refine the algorithm


Phase 1: Random Initial Order, Neighbour Comparison, Quartile Bins



40	39	41	42	44	43	38	37	32	31	33	34	36	35	45	46	55	54	56	57	59	58	53	52	48	47	49	50	51	30	29	9	8	10	11	13	12	7	6	2	1	3	4	5	14	15	24	23	25	26	28	27	22	21	17	16	18	19	20	0
32	38	40	41	31	27	22	24	26	43	44	55	57	59	54	53	46	48	52	20	29	10	13	9	5	3	14	12	19	17	49	50	7	8	47	51	23	58	1	2	56	45	4	6	25	33	34	16	15	0	30	35	36	42	28	11	21	37	39	18
24	26	20	17	14	27	32	53	57	46	44	41	13	29	5	43	4	40	39	35	12	38	3	47	56	1	59	55	54	48	49	52	34	36	22	23	21	10	11	19	25	9	31	0	7	18	2	16	15	58	37	6	50	45	8	33	28	42	30	51
24	17	26	44	53	14	29	5	34	4	22	21	25	3	47	31	48	46	27	20	35	57	10	9	41	38	13	1	32	54	2	43	50	56	59	55	52	40	51	49	33	19	39	16	6	11	12	28	23	0	7	36	42	58	8	37	15	30	18	45

round #

Phase 2: Using Earlier Judgments to Select New Comparisons



round #

40	39	41	42	44	43	38	37	32	31	33	34	36	35	45	46	55	54	56	57	59	58	53	52	48	47	49	50	51	30	29	9	8	10	11	13	12	7	6	2	1	3	4	5	14	15	24	23	25	26	28	27	22	21	17	16	18	19	20	0
32	38	40	41	31	27	22	24	26	43	44	55	57	59	54	53	46	48	52	20	29	10	13	9	5	3	14	12	19	17	49	50	7	8	47	51	23	58	1	2	56	45	4	6	25	33	34	16	15	0	30	35	36	42	28	11	21	37	39	18
24	26	20	17	14	27	32	53	57	46	44	41	13	29	5	43	4	40	39	35	12	38	3	47	56	1	59	55	54	48	49	52	34	36	22	23	21	10	11	19	25	9	31	0	7	18	2	16	15	58	37	6	50	45	8	33	28	42	30	51
24	17	26	44	53	14	29	5	34	4	22	21	25	3	47	31	48	46	27	20	35	57	10	9	41	38	13	1	32	54	2	43	50	56	59	55	52	40	51	49	33	19	39	16	6	11	12	28	23	0	7	36	42	58	8	37	15	30	18	45
26	5	17	24	32	25	20	22	34	41	53	57	47	46	44	38	29	3	14	13	39	4	16	1	35	55	48	52	54	51	9	2	36	21	23	12	27	0	19	31	10	49	50	56	15	18	11	6	28	59	30	40	7	33	43	42	58	8	37	45
17	24	5	13	22	20	3	14	47	26	46	25	29	34	44	32	38	53	41	16	4	12	57	21	48	54	2	9	1	0	31	52	27	43	39	19	49	59	51	55	10	35	23	36	6	18	33	7	11	40	30	15	50	56	42	37	28	58	8	45
17	5	26	22	13	24	47	34	46	14	29	53	20	3	25	32	57	41	54	12	21	44	4	38	1	39	31	27	48	16	9	51	0	23	10	52	49	6	2	19	7	43	35	11	15	36	55	59	33	18	30	37	28	50	40	56	8	42	58	45
17	22	24	26	13	5	29	47	53	20	25	46	14	34	21	54	3	31	44	41	12	27	16	32	57	39	4	38	0	23	2	48	1	9	6	7	19	10	35	51	52	59	49	43	11	33	18	30	55	36	15	40	42	50	8	28	37	56	58	45

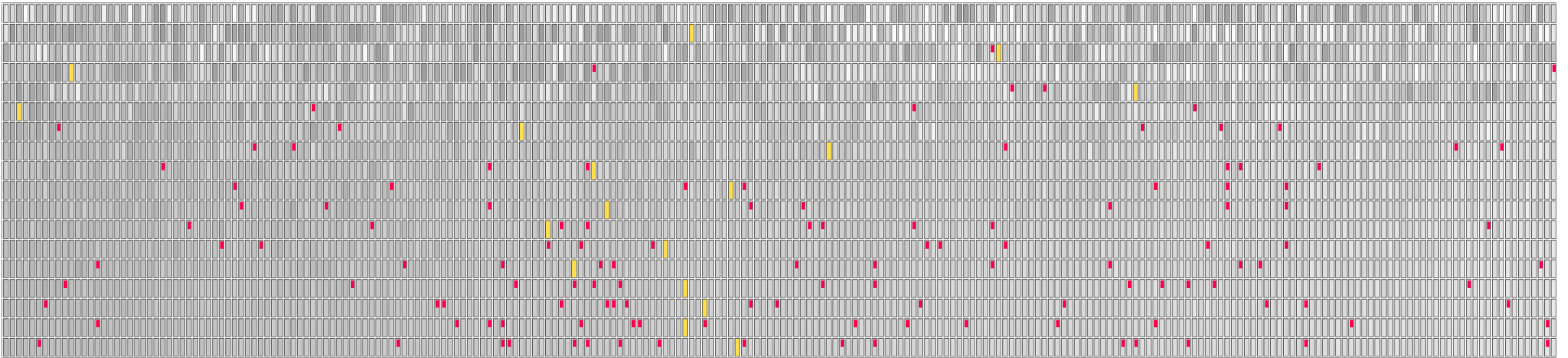
Phase 3: More Refined Comparison with Near Neighbours

40	39	41	42	44	43	38	37	32	31	33	34	36	35	45	46	55	54	56	57	59	58	53	52	48	47	49	50	51	30	29	9	8	10	11	13	12	7	6	2	1	3	4	5	14	15	24	23	25	26	28	27	22	21	17	16	18	19	20	0
32	38	40	41	31	27	22	24	26	43	44	55	57	59	54	53	46	48	52	20	29	10	13	9	5	3	14	12	19	17	49	50	7	8	47	51	23	58	1	2	56	45	4	6	25	33	34	16	15	0	30	35	36	42	28	11	21	37	39	18
24	26	20	17	14	27	32	53	57	46	44	41	13	29	5	43	4	40	39	35	12	38	3	47	56	1	59	55	54	48	49	52	34	36	22	23	21	10	11	19	25	9	31	0	7	18	2	16	15	58	37	6	50	45	8	33	28	42	30	51
24	17	26	44	53	14	29	5	34	4	22	21	25	3	47	31	48	46	27	20	35	57	10	9	41	38	13	1	32	54	2	43	50	56	59	55	52	40	51	49	33	19	39	16	6	11	12	28	23	0	7	36	42	58	8	37	15	30	18	45
26	5	17	24	32	25	20	22	34	41	53	57	47	46	44	38	29	3	14	13	39	4	16	1	35	55	48	52	54	51	9	2	36	21	23	12	27	0	19	31	10	49	50	56	15	18	11	6	28	59	30	40	7	33	43	42	58	8	37	45
17	24	5	13	22	20	3	14	47	26	46	25	29	34	44	32	38	53	41	16	4	12	57	21	48	54	2	9	1	0	31	52	27	43	39	19	49	59	51	55	10	35	23	36	6	18	33	7	11	40	30	15	50	56	42	37	28	58	8	45
17	5	26	22	13	24	47	34	46	14	29	53	20	3	25	32	57	41	54	12	21	44	4	38	1	39	31	27	48	16	9	51	0	23	10	52	49	6	2	19	7	43	35	11	15	36	55	59	33	18	30	37	28	50	40	56	8	42	58	45
17	22	24	26	13	5	29	47	53	20	25	46	14	34	21	54	3	31	44	41	12	27	16	32	57	39	4	38	0	23	2	48	1	9	6	7	19	10	35	51	52	59	49	43	11	33	18	30	55	36	15	40	42	50	8	28	37	56	58	45
17	26	24	22	53	20	46	13	5	21	34	47	54	29	25	27	14	41	57	32	31	3	44	38	12	16	23	39	9	19	0	4	48	10	35	6	7	2	52	1	49	33	18	51	36	11	15	59	43	30	28	37	55	40	50	8	56	42	58	45
17	26	24	22	53	20	46	5	13	21	47	34	29	25	54	14	27	41	57	32	3	31	38	44	9	0	12	39	16	23	4	6	10	48	1	2	19	33	49	7	11	35	15	51	52	28	18	43	59	42	36	30	40	37	55	50	8	56	58	45
17	26	24	22	53	46	20	5	13	47	21	29	25	34	54	14	41	27	31	32	57	3	44	38	9	39	12	23	1	0	4	16	10	2	7	48	19	35	49	6	52	18	33	43	59	11	36	51	15	55	40	28	8	30	50	42	37	56	58	45
17	26	24	22	53	46	5	20	13	47	29	21	25	14	34	54	27	41	31	3	38	32	12	39	57	44	4	23	2	9	0	1	10	48	19	16	6	49	35	7	33	59	51	43	15	52	11	28	30	18	55	40	42	37	36	50	56	8	45	58
17	26	24	22	53	5	46	13	20	29	47	21	25	14	34	54	31	27	3	41	44	32	23	38	57	39	9	0	2	48	4	16	49	10	1	7	19	43	6	18	52	35	59	51	33	15	36	11	8	40	28	55	30	42	37	56	50	45	58	
17	26	24	22	5	53	46	13	20	29	47	21	25	14	34	54	27	41	31	3	44	38	23	57	12	32	39	9	4	2	0	48	19	10	49	1	16	43	52	6	7	35	11	15	18	55	33	51	59	40	30	28	36	50	8	37	42	56	45	58
17	26	24	22	5	53	13	46	20	29	47	21	25	34	14	54	41	27	31	3	38	12	44	23	57	32	39	0	9	4	48	2	10	1	43	49	16	6	19	7	35	18	59	52	33	15	28	51	11	40	37	55	42	36	30	8	56	50	45	58
17	26	24	22	5	13	53	46	20	29	47	21	25	34	14	54	41	27	3	31	38	44	12	23	32	39	57	9	48	4	1	0	19	49	10	2	16	43	35	52	6	15	7	51	30	33	59	55	11	18	40	28	36	8	42	56	37	50	45	58
17	26	24	22	5	13	53	46	20	29	47	21	25	34	14	54	31	41	27	3	12	38	23	44	57	32	39	0	9	48	4	10	1	6	49	16	2	43	19	7	52	33	35	15	18	59	28	51	55	30	40	11	42	36	50	37	8	45	56	58
17	26	24	22	5	13	53	46	20	29	47	21	25	34	14	54	31	41	3	12	27	23	44	38	32	39	9	57	4	48	0	1	10	49	19	2	6	52	35	16	43	7	59	51	15	11	33	18	55	40	28	36	50	37	8	50	42	45	58	

round #

Demonstration of Scaling

- The same simulation with 600 'artefacts'
- After 17 round sorting is very good
- (Image shows middle $\sim 1/3$ with one 'artefact' highlighted)



Adaptive Comparative Judgement

<https://learn.gla.ac.uk/acjdemo/>

This demonstration lets you try out ACJ by comparing photographs of wildlife and flowers. (It uses a development version of the software that doesn't require a login)

Case Study

Functional Programming in Haskell: Supercharge Your Coding



- Futurelearn MOOC (n=1000)
- COMPSCI4021 (n=80)



Case Study Continued

In the Haskell MOOC, we asked students to peer assess using ACJ.

They see other students' solutions to the exercise after they have submitted their own solution.



Students received:

1. Problem spec (to implement)
2. Quality guidelines as judgment criterion
3. Peers' solutions (to compare)
4. Ranking of their own work (quartile bin)
5. A sample solution

Sample solution

```
-- | The 'wordPhrase' function spells out an individual word
-- For example, "a is for apple"
wordPhrase :: String -> String
wordPhrase x = (head x) : " is for " ++ x

-- | The 'speller' function generates text for a spelling book
-- from a list of words
speller :: [String] -> String
speller [] = []
speller [x] = wordPhrase x
speller [x,y] = xPhrase ++ ", and " ++ yPhrase
  where
    xPhrase = wordPhrase x
    yPhrase = wordPhrase y
speller (x:xs) = wordPhrase x ++ ", " ++ speller xs
```

Student comments

- I can see different ways of thinking and I try to understand which one is better(more efficient) and I hope that I will be able to make my own codes more efficient in the future.
- The approach forces you to think differently. This can only be trained by doing it.
- Being able to compare your own work against lots of others lets you see roughly how well/poorly you are progressing in the course compared to your classmates as a whole.
- I think that it is a very useful exercise (both writing a code and comparing the codes of other students) and it is organised in a great way. I would like to thank the course educators.
- As you start comparing you can see the different approaches students started using and everything could be compared faster.

Student comments about their experience of ACJ

I think that it is a very useful exercise (both writing a code and comparing the codes of other students) and it is organised in a great way. I would like to thank the course educators.

forces you
This
by

Being able
work against
you see rough
you are prog
compared to your
a whole.

As you start comparing you can see the different approaches students started using and everything could be compared faster.

Where next?

- Still a development / pilot tool
 - Further refinement possible
- Could this be useful in your teaching?
 - Scholarship / research
 - Not a 'Service'

ACJ Software Components

Display User Interface

Windows, panes for multiple docs, etc

Input User Interface

Note taking, feedback
private bank of comments

Stats Calculations

Estimates of order & distances

Decision Making

Select pairs for comparison

Minimise Incidental Complexity of Marking

- Several colleagues use ad-hoc software lash-ups to increase marking speed and quality
- e.g. pen input, PDF annotation, cloud storage, multi-window/screen views
- Plimmer (2006) systematically explores the possibilities for paperless, pen-based marking
- Our proposed UI components would be informed by their experiences

References

- Dale, V.H.M., et al, 2019. Learner experiences of a blended course incorporating a MOOC on Haskell functional programming. Accepted for publication in *Research in Learning Technology*
- Plimmer, B. & Mason, P., 2006. A pen-based paperless environment for annotating and marking student assignments. *7th Australasian User Interface Conference*.
- Pollitt, A., 2012. The method of Adaptive Comparative Judgement. *Assessment in Education: Principles, Policy & Practice*, 19(3), pp.281–300.
- Thurstone, L. L., 1927. A law of comparative judgment. *Psychological Review*, 34(4), pp.273-286.
<http://dx.doi.org/10.1037/h0070288>

Pointers

Steve.Draper@glasgow.ac.uk

Sarah.Honeychurch@glasgow.ac.uk

Niall.Barr@glasgow.ac.uk

Jeremy.Singer@glasgow.ac.uk

Source code: <https://github.com/niallb/ACJ-LTI>

Further notes: <http://www.psy.gla.ac.uk/~steve/apr/apr.html>