

*Third CJ Users' meeting
University of Birmingham
17th January 2025*



Exploring students' experience of CJ: a qualitative perspective on the act of ordering

Marie Hoffelinck
University of Liège, Belgium

INTRODUCTION



- ▶ *Unfolding* rather than outcomes of CJ sessions
- ▶ « Learning by comparing in Higher Education »
- ▶ Holistic judgements >< Analytic judgements (Canty, 2012; Verhavert et al., 2019, Sadler, 2009)
- ▶ Criteria guiding individual decisions (Jones & Sirl, 2017; Palisse et al., 2022)
- ▶ *Sorting Things Out* (Bowker & Star, 2023 [1999])

How are decisions made in CJ sessions? What does it say about what CJ is?



CONTEXT

Questions and practice of young child education (PEDA-0069)

- ▶ Spring semester 2024
- ▶ MA in Education
- ▶ Mid-semester formative evaluation (3p., 1st draft of a section of the final essay)
- ▶ Peer feedback activity
- ▶ Comproved
- ▶ 5 CJ + 4 feedback/student
- ▶ 6 quality criteria (e.g. ‘a good essay should demonstrate use of specific professional and scientific vocabulary’)
- ▶ 4 Longitudinal interviews with 8 self-selected students (**before making CJ – after CJ sessions** - after feedback reception – after final assignment completion) + recording of the ‘CJ initiation’ class

The screenshot shows a peer feedback form with two columns of text. The left column contains a paragraph about Jérôme, and the right column contains a paragraph about Anais. Below the text are two sets of feedback criteria: 'Forme et lisibilité (fil conducteur, vocabulaire, posture compréhensive)', 'Rigueur (ressources théoriques, pratiques)', and 'Forme et lisibilité (fil conducteur, vocabulaire, posture compréhensive)'. A red arrow points from the 'Feedback criteria' label to the 'Forme et lisibilité' criteria.

The screenshot shows a 'Feedback criteria' window with a list of examples for feedback. The criteria are: 'Le fil conducteur est absent / Le fil conducteur est présent au sein de chaque point mais pas globalement pour tout le travail / Le fil conducteur est présent autant au sein de chaque point qu'entre les points.', 'Utilisation de termes non spécifiques donnant l'impression que le cours n'a pas été suivi / Utilisation de termes parfois spécifiques au domaine parfois non / Utilisation approprié du vocabulaire professionnel et scientifique spécifique au domaine', and 'Usage régulier de de termes déontiques (falloir, devoir, ...) accompagné de nombreuses prescriptions / Usage de certains termes déontiques avec prescriptions, mais également des propositions interprétatives/ Usage réduit des termes déontiques avec prescriptions davantage de propositions interprétatives.'

FINDINGS



(1) Using explicit criteria alleviates suspicions towards CJ

Teacher: “The first time I was presented [CJ], **I was sceptical**, and I said: ‘I don’t want anything normative at all. I don’t want to be told here’s the 1st, the 2nd, the 3rd ... **I want something criterion referenced.**’ And when you dig further into [the method], the criterion element is there indeed. We don’t want to put you in a hit parade. What we want is that you make the assessment criteria your own.” [...]

Student: “I still wonder... because it is called ‘comparative *judgement*’. I mean, well, as an educational psychologist, it makes me feel uncomfortable... [...]”

Teacher: “In French, the word ‘**judgement**’ is often associated with something arbitrary. **But when you use criteria, it is not.** I said no to the idea of you seeing a ranking. [...] The aim is not ranking, exclusion or disqualification. The aim is that you make your own a series of criteria. Actually, it is adopting a referentialisation approach.”



(2) Easy vs. hard decisions

- *Deciding easily when...*

- *Having trouble deciding when...*

- *...one item is inappropriate*

‘Some essays did not respect the instructions regarding the assignment at all so it was easy to decide’ (Arthur)

- *...both productions are of high quality*

‘Sometimes, decisions were hard to make, because there were good things in both items.’ (Anne-Lise)
‘It was harder to compare when both essays were of high quality’. (Léa)

See Jones & Sirl, (2017)

- *...the topic of both item was similar*

‘It was easier to compare when both essays were dealing with the same topic’ (Léa)

- *...several criteria have to be taken into account simultaneously*

‘On some occasions, it was hard to make a decision. I’d have preferred making decisions based on each individual criterion because sometimes two criteria were best in one item and two other ones in the other one.’ (Claire)

(3) How hard decisions are made



- *Hierarchising criteria: form over content*

‘In one case, the comparison was harder because both texts had different flaws. It was hard to determine which criterion was more important than the others. So, I read the texts over and over several times and finally I chose the one that had a better structure’ (Catherine)

‘What was decisive was the clarity of the point made, the articulation between ideas. The form. What was “*un peu gros*” [most salient]’ (Sophie)

‘I did privilege form’ (Claire)

- *Beyond criteria: ‘heart’, ‘subjectivity’, proximity and ‘being programmed’*

‘Sometimes decisions were harder to make, so it was about making **a choice from the heart**. It was maybe not the most objective but I chose the essay that **spoke to me the most**, maybe because the points that were made were **closer to the ones I had myself developed.**’ (Manoël)

‘Maybe I favoured essays that were dealing with the **same subject as mine...**’ (Léa)

Arthur: ‘When one judges, there’s always a part of **subjectivity** to it.’ —Marie: ‘What subjectivity? As here you don’t know who wrote the essay...’ —Arthur: ‘For example, when [the layout] is not justified, one has already a bad opinion from the start even though one hasn’t read anything yet’.

‘I realised spelling was super important to me. An essay with many spelling mistakes meant that it did not deserve to be chosen as the best. But maybe actually the content was better but there was something “qui a *buggé*” [that went wrong] with form. [...] But, in the end what is it with spelling? Maybe it could be a foreign student who doesn’t master the language... But... nowadays writing with spelling mistakes... I mean, there are means to get it checked... [...] What value do we give to spelling as teachers, right? **We’re programmed**, aren’t we?’ (Anne-Lise)



(4) Progressively factoring in the assessment criteria

‘The first decision I really made it **“au feeling” [with my guts]** because I didn’t have the criteria in mind.’
(Claire)

‘I think **subjective elements were more important at the beginning**. I think they had less impact on the last decisions because by that time I had the criteria fresh in mind ‘cause I had to use them to write feedback.’ (Arthur)

‘When I had to check the criteria to write feedback, sometimes I doubted the choice I made.’ (Anne-Lise)

‘At the beginning, I really felt an impostor syndrome for the first comparisons. And, then I gained confidence, and it seemed easier’ (Catherine)

See Bartholomew et al. (2022)

DISCUSSION



- ▶ Using pre-determined criteria...
 - ▶ Makes decisions harder
 - ▶ Does not preclude other considerations from impacting judgements

- ▶ CJ as a black box
 - ▶ Standards internalised by communities of practices (Bowker & Star, 2023 [1999]; Wenger, 1998)

REFERENCES



- Bartholomew, S. R., Mentzer, N., Jones, M., Sherman, D., & Baniya, S. (2022). Learning by evaluating (LbE) through adaptive comparative judgment. *International Journal of Technology and Design Education*, 32(2), 1191–1205. <https://doi.org/10.1007/s10798-020-09639-1>
- Bowker, G., & Star, S. L. (2023). *Arranger les choses: Des conséquences de la classification [Sorting Things Out. Classification and its Consequences]* (H. Boullier, V. Cardon, É. Dagiral, B. Kotras, A. Peerbaye, V. Zabban, V. Cardon, É. Dagiral, & A. Peerbaye, Trans.). Éditions de l'EHESS.
- Canty, D. (2012). *The Impact of Holistic Assessment using Adaptive Comparative Judgement on Student Learning*.
- Jones, I., & Sirl, D. (2017). Peer assessment of mathematical understanding using comparative judgement. *NOMAD Nordic Studies in Mathematics Education*, 22(4). <https://doi.org/10.7146/nomad.v22i4.148924>
- Palisse, J., King, D. M., & MacLean, M. (2022). Comparative judgement and the hierarchy of students' choice criteria. *International Journal of Mathematical Education in Science and Technology*, 53(1), 206–226. <https://doi.org/10.1080/0020739X.2021.1962553>
- Sadler, D. R. (2009). Transforming Holistic Assessment and Grading into a Vehicle for Complex Learning. In G. Joughin (Ed.), *Assessment, Learning and Judgement in Higher Education* (pp. 1–19). Springer Netherlands. https://doi.org/10.1007/978-1-4020-8905-3_4
- Verhavert, S., Bouwer, R., Donche, V., & De Maeyer, S. (2019). A meta-analysis on the reliability of comparative judgement. *Assessment in Education: Principles, Policy & Practice*, 26(5), 541–562. <https://doi.org/10.1080/0969594X.2019.1602027>
- Wenger, E. (1998). *Communities of Practice: Learning, Meaning, and Identity* (1st ed.). Cambridge University Press. <https://doi.org/10.1017/CBO9780511803932>