

# **Students' Experience of Comparison-Based Learning in Self-Assessment: A Phenomenographic Study in Higher Education**

Marie Hoffelinck<sup>1a</sup>, Pascal Detroz<sup>a</sup>, Dominique Verpoorten<sup>a</sup>

<sup>a</sup>IFRES, University of Liege, Liege, Belgium

The Version of Record (VoR) of this manuscript has been published and is available in *Higher Education Research & Development*,

<https://www.tandfonline.com/doi/abs/10.1080/07294360.2025.2463521>

---

<sup>1</sup> Corresponding author. Contact: marie.hoffelinck@uliege.be

# **Students' Experience of Comparison-Based Learning in Self-Assessment: A Phenomenographic Study in Higher Education**

This paper explores, through a phenomenographic analysis, how higher education students experience learning-through-comparison processes embedded in self-assessment activities. By combining observations and interviews, the paper offers a threefold contribution: (1) it provides two in-depth narratives of self-assessment activities; (2) it develops a typology that distinguishes between explicit versus implicit and analogical versus analytical comparisons; and (3) it proposes an outcome space to illustrate how students experience these contrasting assignments. The analysis detects a specific value of analogical comparisons for self-generated learning and of analytical comparisons for a better understanding of the course stakes. In both cases, making comparisons explicit is a way to foster cognitive and metacognitive benefits. Additionally, the findings indicate that comparisons with peer productions should receive special attention as students are less familiar with them than with traditional instructor feedback. Lastly, results suggest that consistency in self-marking should not always be interpreted as an indicator of powerful self-assessment. These insights hold implications for researchers and educators concerned with the ins and outs of self-assessment practice.

Keywords: learning by comparing; postsecondary education; self-assessment; outcome space

Word count: 8127

## **Introduction**

The purpose of this paper is to better understand comparison-based learning processes (CLPs) within the context of self-assessment (SA) teaching practices, understood as pedagogical activities designed by teachers to prompt students to self-assess. The importance of comparative processes for learning have been given a growing attention in recent years (see notably: Alfieri et al., 2013; Bartholomew & Jones, 2022; Hartell & Buckley, 2021; Nicol, 2021) and there is a need to further understand how they occur in different situations, including SA. Strengthening the connection between these two

fields offers a mutually reinforcing dynamic: comparative processes naturally thrive within the context of SA, while SA practices can be enriched through the structured incorporation of CLPs. This synergy may not only advance instructional design but also support the development of more autonomous and reflective learners.

### ***Self-assessment practices***

SA is a broad concept. Even though there have been several attempts to propose an overarching definition (notably: Andrade, 2019; Panadero et al., 2016), prominent authors in the field keep deploring the lack of consensus on the matter (Andrade, 2019; Boud & Brew, 1995; Yan et al., 2023). A salient ambiguity resides in the polarisation between a conception of SA as an internal process (akin to ‘feedback’, Andrade, 2019) versus a more practical standpoint that regards SA as an ‘act’ (akin to ‘techniques’, Brown & Harris, 2013). Anyhow, mostly due to the globally positive impact it appears to have on academic performance (Brown & Harris, 2013; Yan, 2020; Yan et al., 2022), research on the matter is prolific. Notwithstanding, Andrade (2019) posits that there needs to be more studies dedicated to unpacking what happens when students self-assess: “there is an even more pressing need for investigations into the internal mechanisms experienced by students engaged in assessing their own learning. Angela Lui and [Heidi Andrade] call this the *next black box* (Lui, 2017).” (Andrade, 2019, p. 10). This is exactly what this study intends to contribute to. To do so, it takes naturalistic ‘SA practices’ (acts) as a starting point, to inform ensuing ‘SA processes’.

### ***Comparative processes***

A key aspect of any process of SA is that it inherently entails a cognitive operation of comparison by the student to evaluate the quality of their own production/performance.

This aligns with what a prominent author in feedback literature (Nicol, 2021) has recently stressed: measuring the relative quality of an answer against a yardstick implies a process of comparison. The importance of comparative mental processes to support students' evaluative skills is well-acknowledged in comparative judgement research (Bartholomew et al., 2022; Dewit et al., 2021; Jones & Alcock, 2014; Jones & Wheadon, 2015; Potter et al., 2017), however it remains globally under-studied (Nicol, 2021). And indeed, in SA research, while the comparative nature of the process is often recognised (Brown & Harris, 2013; Panadero et al., 2012, 2016; Ruchniewicz & Barzel, 2019; Sargeant et al., 2010; Taras, 2010; Yan & Brown, 2017), this dimension still needs to be better understood in order to shed light inside the 'black box'. More specifically, this paper intends to contribute to the understanding of students' engagement with external sources of feedback (such as teachers' comments, marking criteria or peer productions) in the context of SA.

### ***Phenomenographic approach***

Besides Nicol's (2021) conceptual stance, there is only a modicum of research inspecting the impact of specific comparative teaching activities on learning in higher education (for example: Bartholomew et al., 2022; Bransford & Schwartz, 1999; Naud et al., 2023; Sidney et al., 2015). Overall, the question of experience—that is the way in which the comparative learning opportunities are *lived* by students—remains unexplored. To address this issue, this paper adopts a phenomenographic orientation (Åkerlind, 2005, 2023; Marton, 2015; Svensson, 1997) to delineate different qualitative ways of experiencing the comparative dimensions of SA.

Traditionally, phenomenographic research, as presented by Marton (2015), aims at discerning different 'ways of seeing' amongst learners in order to identify the critical aspects of an *object of learning* (i.e. what is to be learned). Thus, the recourse to

Marton's theory here is unusual in the sense that it is not applied to a contextualised object of learning (a '*what*') but rather to specific learning opportunities or processes ('*hows*')—comparative in nature—embedded in a specific type of pedagogical activity. What is the focus of analysis here are therefore the mental processes—understood as lived experiences of SA—that are crucial for students to change how they see either an object of learning or their own performance. When phenomenographic research is generally oriented towards the unpacking of cognitive 'critical aspects' of an object of learning (Marton, 2015), the focus of this paper can be understood as '*processual critical aspects*' of SA.

In line with Marton (2015) and Marton & Booth (1997), it is considered here that there are 'more powerful' ways of experiencing something. A 'more powerful experience' consists of more complex ways of experiencing, meaning that more different aspects of a phenomenon are experienced simultaneously by the learner (Marton & Booth, 1997, pp. 176–177). Hence the second objective of this paper: articulating students' accounts of SA CLPs experiences in an outcome space, delineating less and more powerful ways of experiencing the phenomenon (Åkerlind, 2005; Marton, 2015; Marton & Booth, 1997).

This paper addresses therefore two research questions:

- *What does characterise CLPs presented to students in the context of SA?*
- *How are CLPs experienced by students in the context of SA?*

## **Material and Methods**

To deepen the understanding of how students live the comparative aspects of SA, two real-world academic activities were investigated.

## ***Contexts***

Both contexts share the common feature of being instances of self-marking practices (i.e. formative activities during which teachers ask their students to evaluate their work under the form of a numbered score). These practices were identified during discussions amongst ten participants in a teacher professional development programme (*Formasup*, n.d.). These educational activities were observed on-site. Upon completion, attending students were invited to participate in semi-structured interviews.

Coincidentally, the two targeted pedagogical activities were third-level Bachelor's courses organised by the Applied Science Faculty of the University of Liege. One took place within a computer science course and the other one within a thermodynamics one. The computer science class is mandatory for students in informatics and both courses are optional for students in applied sciences (depending on the specialty field they chose). Approximately 20 students participated in the computer science class observed, from which five accepted to be interviewed and 50 to the thermodynamics one, from which three accepted to be interviewed. The material used for this research is therefore constituted of two filmed observations and eight interviews with students.

## ***Empirical material sources and instruments***

### ***Observations***

The two pedagogical activities have been filmed to keep track of the precise sequences of events. The recordings were used to trace and qualify (see below, vignettes and tables 1 & 2) the different CLPs afforded to students during the pedagogical activities. Drawing on Nicol's (2021) analytical frame for comparison processes as a labelling instrument, CLPs are characterised by four different features. They can be *implicit*, when the comparison opportunity exists but no formal action is required from

the students (e.g. they hear about their peers' answers) or *explicit*, when teachers ask students to act with respect to the comparison (e.g. self-mark). Furthermore, CLPs may be of two different natures: analytical or analogical. An *analogical* comparison occurs between several analogous elements that share common fundamental features (e.g. comparing the student's production with the standard solution) whereas *analytical* comparisons consist of the confrontation of an element with a reference that is different in its nature (e.g. an essay vs. a rubric).

### *Interviews*

The interviews took place within a maximum of two weeks after the observed session. They were scheduled to suit the students' preferences in terms of time and location, with six conducted at the university cafeteria, one at the first author's office, and one online. As per recommended by Svensson (1997), an open explorative collection of data approach was favoured because it combines prepared questions of the researcher with openness to emergent elaborations by interviewees in order to let any important fact come up and not only the ones preconceived by the researcher. Following Entwistle's advice (1997), attention was also paid to move from very concrete actions (such as what students recalled writing on their sheets) to more abstract ones (how students apprehended the differences between their sheets and the standard solution). To ensure qualitatively dense material, the tone of the interview was intentionally casual and attention was paid to make students comfortable sharing their experiences (Kaufmann, 2016). All interviews were carried in French (first language of all participants and authors) and excerpts reproduced here have been translated by the first author.

### ***Phenomenographic method***

The aim of this research is to explore ‘naturalistic’ pedagogical activities in order to categorise and articulate different *ways of experiencing*<sup>2</sup> (Åkerlind, 2023; Svensson, 1997) the CLPs embedded in them. For each type of CLP (see below, tables 1 & 2), the various ways they were experienced by students in both pedagogical activities are described using gerunds (Charmaz, 2006, p. 46), thus refining categories (Lejeune, 2019; Marton, 1981). Adopting a second-order perspective (Marton, 1981), these ways of experiencing are apprehended through the analysis of self-reported students (re)actions to the instructional activity they attended. By understanding students’ acts, the observer can apprehend their way of experiencing the assignment (Marton, 2015, pp. 110–113) or, in this specific case, how they seize (or not) the learning opportunities offered to them. If some actions were prompted by the teachers’ instructions, this does not mean that there is not a ‘lived’/‘individual’ way of experiencing/reacting to it: there are different ways to interpret teachers’ guidelines.

---

<sup>2</sup> The word ‘conception’ has often been used in phenomenographic writings to signify what this research orientation aims at describing, especially in the earlier published works (Marton & Booth, 1997, p. 185). However, ‘conception’ and ‘way of experiencing’ together with ‘meaning’, ‘ways of seeing’ or ‘lived objects of learning’ may be understood as synonyms in the context of phenomenography (Marton, 2015, p. 106; Marton & Booth, 1997, p. 185). Here, the expression ‘ways of experiencing’ is favoured because it is better suited to account for the students’ *processual* experience, that can be understood as a medium to change their *conceptions* of objects of learning.

## Results

### *Comparative dimensions of self-assessment in two instructional designs*

The first research output presented here are detailed accounts of the two SA activities. Both include some ‘self-marking’ and teacher feedback on the expected answers but vary regarding other features. They are presented below as *vignettes* and followed by two tables, listing the CLPs observed. In the vignettes, Greek letters and bold characters are used to identify the CLPs later described in the tables.

AUTHOR POSTPRINT

### Vignette 1

It is the fourth week of the second semester. The teacher has invited the students to an extra session, during lunch break, to take part to a ‘training test’. In total, there will be four training tests organised during the semester. Participation is voluntary.

Unfortunately, the teacher cannot be physically present during the activity—he has tested positive for Covid—therefore, a teaching assistant is there to facilitate the practical aspects of the session while the teacher is giving his instructions through a videoconference call.

Approximately 20 students are present. When they are all ready, the teacher presents them with an exercise (on the screen) that they ought to resolve in 25 minutes, simulating an exam: they cannot use their notes nor interact with each other. The exercise consists of the drawing of an entity-relationship diagram. Before starting to work, students are informed that once the exercise is completed, they will be given correction criteria to help them assess the quality of their work, similar to the ones that will be used to grade their exams. After that, they will be invited to share their score through an online platform.

When the time is up, the teacher asks the students to stop working and immediately proceeds to the correction by showing a slide describing the number of points allocated to the different elements of the solution, amounting to a total of 15. After that, **the teacher shows the expected answer to the first part of the exercise ( $\alpha$ ), comments it ( $\beta$ ) and explains how to apply the assessment criteria: ‘this error results in X-point deduction’ ( $\gamma$ ).** The teacher also **describes an alternative solution that is acceptable ( $\alpha$ ).** Then, he asks whether anyone has any question or if they would like to share any other alternative solution. He then proceeds the same way for the next parts of the exercise.

Different students asked a few questions. On several occasions, **students shared alternative solutions they found ( $\delta$ ).** Each time, **the teacher commented on the proposed solution ( $\beta$ ),** explaining why it would be acceptable or not at the exam, often pointing out the implications such an answer would have on other parts of the diagram to maintain coherence throughout. The teacher also commented on the relative quality of different solutions. For example, on one occasion, the teacher admitted the solution suggested by a student ‘would technically work’ but is not ‘elegant’, meaning that it could generate further problems should it be implemented in the “real world”.

The following chain of CLPs may be observed in the computer science class (see table 1).

The order in which they are presented below follows a logico-chronological order coherent with the vignette described above. However, it is important to note that students do not necessarily cognitively perform the comparison processes in that order, more so they describe going back and forth to different steps several times during the session.

Moreover, not all students seize (all) comparative learning opportunities, as will be developed upon further below.

| Reference in the vignette | Implicit/Explicit | Comparison process | Object                        | Comparator              |
|---------------------------|-------------------|--------------------|-------------------------------|-------------------------|
| $\alpha$                  | Explicit          | Analogical         | Student's own production (SP) | Standard response (SR)  |
| $\beta$                   | Implicit          | Analytical         | SP                            | Teacher's comments (TC) |
| $\gamma$                  | Explicit          | Analytical         | SP                            | Marking criteria (MC)   |
| $\delta$                  | Implicit          | Analogical         | SP                            | Peer's production (PP)  |

Table 1. Comparison-based learning processes (computer science training test)

*Thermodynamics end-of-class recap test*

Similarly, vignette 2 and table 2 (see below) give an account of how the thermodynamics end-of-class test unfolded and the CLPs embedded in the pedagogical activity.

## Vignette 2

It is half-way through the second semester. On Monday afternoons, students have a two-hour thermodynamics theory class immediately followed by a two-hour thermodynamics practical exercises session. Both classes are highly interrelated but taught by different teachers. Every week, approximately the last half-hour of the practical exercise session is dedicated to an ‘end-of-class recap test’, made of one or a few questions on the most important elements of that day’s lesson.

When the time comes, the teacher announces that the test is about to start. The test questions are projected on the screen and the students receive a printed version to work on. A teaching assistant, not present earlier, joins the class. The teacher gives a few precisions on the questions and invites the students to start working and reminds them that—as usual—they are welcome to ask questions, to her or to the teaching assistant.

Immediately, students (roughly 50 of them are present) start working, occasionally **whispering with one of their classmates ( $\epsilon$ )**, **pointing at their sheet ( $\zeta$ )**, **looking at their course notes** or rising their hand to **request assistance ( $\beta$ )**.

After about 20 minutes, the teacher declares that the correction is about to start and invites the students to take a coloured pen (different than the one they have used). She asks students to **estimate their score out of twenty points ( $\eta$ )**. In the upper right corner of the test sheet they have received, some space is dedicated to three different scores, as follows:

- *A priori* estimated score:     /20
- Obtained score:             /20
- Peer-validated score:         /20

Then, she proceeds to **give the correct answers ( $\alpha$ )**, resolving herself the equations using a digital pen on her tablet screen, projected on the classroom white screen and **explaining briefly the answers as she goes ( $\beta$ )**. After having answered all the questions, the teacher **indicates how many points are allocated to each expected answer: each correct answer results to X-many points, asks the students to calculate their obtained score ( $\gamma$ )** and reminds them to **swap test sheets with a peer to have them correct their sheets too ( $\delta, \theta, \iota$ )**. Students proceed to pack their stuff and, leaving the classroom, put their test sheets on the teacher’s desk.

| Reference in the vignette | Implicit/Explicit | Comparison process | Object                          | Comparator         |
|---------------------------|-------------------|--------------------|---------------------------------|--------------------|
| $\epsilon$                | Implicit          | Analytical         | SP                              | Peer comments (PC) |
| $\zeta$                   | Implicit          | Analytical         | SP                              | Course notes (CN)  |
| $\beta$                   | Implicit          | Analytical         | SP                              | TC                 |
| $\delta$                  | Implicit          | Analogical         | SP                              | PP                 |
| $\eta$                    | Explicit          | Analogical         | Overall performance on the test | Past performance   |

|          |          |            |    |                    |
|----------|----------|------------|----|--------------------|
| $\alpha$ | Explicit | Analogical | SP | SR                 |
| $\gamma$ | Explicit | Analytical | SP | MC (point balance) |
| $\theta$ | Explicit | Analogical | PP | SR                 |
| $\iota$  | Explicit | Analytical | PP | MC (point balance) |

Table 2. Comparison-based learning processes (thermodynamics end-of-class recap test)

***Six typical occurrences of comparison-based learning processes embedded in self-assessment***

Resorting to the framework identified by Nicol (2021), six occurrences, similar in both activities, covering four different comparison types pinpointed by the latter are the targets of the forthcoming analysis.

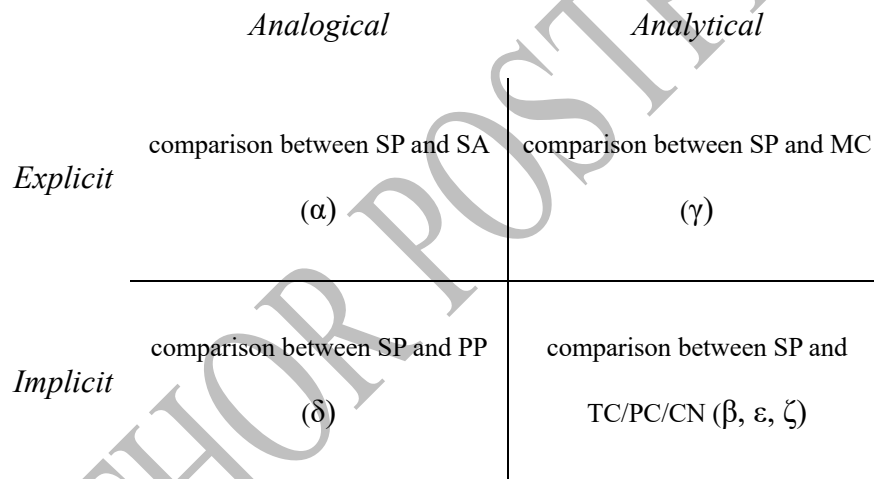


Fig. 1 Four different types of CLPs embedded in the SA activities.

Some observed CLPs are not referred to in the next section. It is the case of ( $\theta$ ) and ( $\iota$ ), because interviewed students admitted not having performed them, even though

they were explicitly requested to<sup>3</sup>. It is also the case of ( $\eta$ ), because the accounts students gave of this moment never included any reference to external sources of information nor had anything to do with what they have or could have learned. Rather—similarly to what Yan & Brown (2017) refer to as ‘internal feedback’ that is guided by subjective experiences or intuition—, they described marking themselves according to their representations of past performance. For example, Alan said: ‘18/20 because I often make inattention mistakes’.

### ***Different ways of experiencing comparison-based learning processes***

In this section, different categories representing qualitative different ways of experiencing CLPs by students are presented: (A), (B), (C), (D), etc. and then articulated in an outcome space (see below). For each CLP, different ways of experiencing it are presented, from the least powerful to the most powerful one. As per recommended by Entwistle (1997), various *verbatim* accounts extracted from the interviews are reproduced to exemplify the categories. Some of them were the same in both classes. However, as more complex levels of experience are reached, divergences appear. Both commonalities and variations between the two studied SA activities will feed discussion.

---

<sup>3</sup> It is interesting to note that these two also share the specificity of not including the SP as comparator and as such were potentially not perceived as an opportunity for SA on the students’ part; whether they could be and how remains an open question that will not be addressed further here.

### *Explicit analogical comparison-based learning processes (SP vs SR)*

These moments correspond to when the standard response (SR) is ‘revealed’. They are crucial in students’ experiences as they amount to the discovery of their ‘failure’ or ‘success’. Their immediate reaction consists of deliberately indicating on their sheet of paper any divergence between their solution (SP) and the SR, either by using a coloured pen or by crossing wrong answers. At first, their intention is not (yet) to try to understand anything but somehow to keep track of the differences between what they did and what was expected from them. This corresponds to the first category of way of experiencing explicit analogical CLPs:

#### (A) Highlighting the differences

Amira<sup>4</sup>: ‘If it’s correct, I put a little “tick”, if it’s wrong, I write the correct answer, well, I cross it and I write the correct answer’.

Tristan: ‘And I was comparing “I was right/I was wrong” and I wasn’t really trying to think about what was written’.

It is crucial to note that this comparison takes some time, variable from one student to the other, and, most importantly it is dependent on the number of differences (mistakes) they made. Moreover, while they are performing this first explicit CLP, students are not really cognitively available for anything else:

Sophie: “I just looked. I didn’t listen to what [the teacher] said. [...] My first intention was to look and then, afterward, I started to listen to him, see if there were any precisions.”

---

<sup>4</sup> All names are pseudonyms.

After this initial step, students reported clearly identifying their mistakes but fell short of finding an explanation to them (which occurs later on, as described below).

(B) Pointing the mistake(s)

Stéphane: 'If I want to improve myself, I must see exactly where I got it wrong.'

Tristan: 'More importantly, after, being able to tell myself: "yes, okay, now I see where I made a mistake, I'm not going to make it anymore."'

Going further, students from the thermodynamics class (try to) analyse 'how' different their answer is from the correct one to understand where the mistake(s) came from.

Contrariwise, this level of experiencing was not observed in the computer science class at this stage. This may intrinsically be linked to the nature of the expected answer but another explaining element may also lie in the way the answers were communicated to students: in the computer class, the correct answers were reproduced on a slideshow prepared by the teacher beforehand and therefore suddenly appeared on screen. On the opposite, the thermodynamics teacher was resolving equations in front of the class. Even if students indicate comparing the answer first and looking at the reasoning afterwards and only if their answer differed from the teacher's, they directly have a cognitive access to the reasoning behind the standard response that offers a basis for reflection; unlike the situation in the computer science class.

(C) Searching for an explanation for the identified mistake(s) — only in the thermodynamics class

Alan: 'If I made a mistake, I wonder, well, where does it come from? I think it is important to know, because it is not only congratulating oneself for success but also look at why one made a mistake'.

Amira: 'What is often misleading when we are correcting, it's approximation mistakes. So, often, we don't have exactly the same value [as the teacher] but we see that the calculation is correct, so we think: 'Oh, no, it's not the same value' but then we look at the calculation and so eventually we think: "oh, well, it's alright"'.

#### *Explicit analytical comparison processes (SP vs MC)*

When comparing their work with the marking criteria (MC), students' first intention corresponds to assessing themselves the way the assessor would have done so. They recognise a part of partiality to it. So, more than being 'fair', what they describe doing is trying to put themselves in their assessors' shoes.

#### **(D) Grading oneself the same way a 'real assessor' would have**

Alan: 'At the very beginning of the test, [the teacher] asked that we convert the pressure from bar to pascal [...], well, I, I didn't do it. I did it, each time, in my calculations: I multiplied by 10 exponent blablabla and, well, listen, she withdrew a point for the answer at the very beginning, so I graded myself at 24/25 [even though all my answers were correct]'.

Lola: 'We don't all judge ourselves the same way. When it's only one teacher who grades, they are already a little bit more impartial in the way they allocate grades. But, is [the teacher] mean? Does [the teacher] give away marks easily or not?'

Beyond that, some students mentioned adopting a position of severity towards themselves, aimed at stimulating them to do even better next time.

(E) Being as severe as possible with oneself

Alan: 'I'd say I'm a bit harsher with myself than with my classmates because for myself I really want to be full on'.

Quite to the opposite, students recount specific occurrences when they decided to intentionally be 'gentler' with themselves because of the perceived origin of the mistake: they knew the correct answer or the correct way to get there but for some circumstantial reasons they did not write it down properly. This attitude shows a more powerful way of experiencing the CLP as it triggers a reflection about the reason(s) why they made the mistake. In this case, they distance themselves from the 'real assessor' posture to adopt a metacognitive one. They do not assess their real answer but their perceived ability to later be able to find the correct one in a similar situation: comparing their answer to the MC allows them to attribute a status to their mistake (e.g. an oversight vs. an intellectual fault in Tristan's words).

(F) Using marking criteria with leeway depending on the attributed cause of mistake

Tristan: 'I personally withdrew myself a half mark instead of one because it was just an oversight and not an intellectual fault'.

Amira: 'Sometimes there are justifications that we don't necessarily think about writing down and that [the teacher] requests [...]. So, that's a mark for which we don't know how to assess ourselves because we just didn't think about writing it down. [Interviewer:] Then, when you are grading yourself, you think: 'that, I could have written it down, it's not that I didn't know it'? [Amira:] Yes, that's it'.

Beyond these ways of experiencing the explicit analytical CPOs ((D), (E) and (F)), that

are directly linked with the concrete action requested (grading their tests), the two different activities result in significant differences regarding more powerful ways of experiencing CLPs. Indeed, students in the computer class show a deeper reflection on the meaning of the MC and on the elements that are constituent of quality in their field (see (G) and (H)).

(G) Identifying the core of the concept(s) — only in the computer science class

Nicolas: ‘I remember that [the teacher] had said: “if you make that mistake, you withdraw 2 marks because it is a very bad mistake”. Well, then, [...] I realised something that I hadn’t really understood when the teacher had explained it during the lecture.’

(H) Assimilating the underlying message of the course — only in the computer science class

Sophie: ‘The marking grid, it leads me to think, [that] it is important to understand that in the course because it is important for the subject, not just important for the teacher but fundamental to understand the subject because, well, later, when I will use it [in a professional context], well, it is something that will be useful for me to know: okay, it’s good to know how to make beautiful diagrams and else but what is important too is to communicate why we did it this way.’

An explanation to the fact that more powerful ways of experiencing explicit analytical CLPs are observed in the computer science class may lie in the fact that, in the thermodynamics one, no ‘mark’ was attributed to the method or the reasoning leading to the correct answer. The marking grid consisted of a broad statement: ‘this answer is worth that many points’; without any distinction regarding the quality of the answer and conveying a simple dualistic view: correct vs incorrect. Whereas, in the computer science class, the marking grid consisted of a series of criteria that, if not met, implied a

deduction. For example: ‘attribute not represented as “key” -> -0.5’, which drew students’ attention to several distinct elements constituent of quality.

#### *Implicit analogical comparison processes (SP vs PP)*

As implicit, these CLPs do not correspond to a concrete action requested by the teachers. And, indeed, some students completely disregarded it:

Nicolas: ‘I didn’t..., I forgot them [the peers’ solutions]. I didn’t pay much attention... [...] Well, I think I spent this time kind of daydreaming, that’s possible. [...] I think unconsciously I thought: “the teacher’s answer is the best. I’m not going to start spending time trying to confront what I did with other solutions potentially wrong or less good”’.

Tristan: ‘Well, other people’s answers, [it is less interesting] because how other people think, it’s not going to help me.’

For students who do seize the opportunity, only one level of experience was accounted for: drawing on differences they identify, they take this occasion to reflect on *themselves*, their methods, their own reasoning and on how to improve them.

#### (I) Reflecting on one’s own way of working

Alan: ‘If I see things that the person sitting next to me has done and that I haven’t, maybe I’ll remember it for myself: [...] ‘he drew a really nice table, I should make one too because it’s a little bit messy on my sheet’.

Lola: ‘Others said another [solution] than the teacher’s one and it wasn’t mine either so I thought: “That’s because they had misunderstood that thing” [...]. And I thought: “why did I choose to do it this way and not another way?”’

#### *Implicit analytical comparison processes (SP vs TC/PC/CN)*

In both classes, various opportunities for implicit analytical CLPs were offered to students. In the computer science class, most opportunities came from the teacher’s

explanations and feedback. In the thermodynamics class, the sources were more diverse, as students were authorised to collaborate, ask questions to the teacher/assistant, look at their notes, etc. Only one way of experiencing this type of CLP was observed: even though students did not consider all the sources as equal—a teacher’s comment is always ‘worth more’ than a peer’s explanation—they all serve the same function, namely they are ‘helpful’ to understand better.

(J) Understanding better one’s mistake(s) thanks to comments / explanations / feedback

Alan: ‘It allows us to discuss in groups and sometimes we can explain things among ourselves because the assistants, they’re not infinitely available.’  
Stéphane: ‘When the teacher explained, it seemed clearer to me’.

***An outcome space of the ways of experiencing comparison-based learning processes in the context of self-assessment***

Using the different categories above, an outcome space was drawn (see fig. 2), representing the different ways of experiencing CLPs as well as their hierarchical structure. The categories are ordered in such a way that more powerful ways to experience CLPs are located upper in the outcome space. Moreover, the outcome space also represents the hierarchical inclusivity (Åkerlind, 2005; Marton, 2015) of the different ways of experiencing CLPs: more powerful categories build on less powerful ones. This means that an upper/more powerful way of experiencing cannot be experienced without having experienced the lower/less powerful one it builds on, such as represented by arrows (see fig.2).

No distinction is made between the types of CLPs at this stage: this corresponds to the way students experienced the SA activities, going back and forth between the

different types of CLPs, building on each other to generate learning; such as they were lived during the pedagogical activity.

Furthermore, as it is shown in the outcome space, the different ways of experiencing can be characterised by their realm of development, namely metacognition and cognition (see fig. 2). The cognition realm encompasses mental processes that allow an experience of increased understanding of the object of learning; whereas the metacognition realm is concerned with experiences of judgements that students make about their performance (Berger & Büchel, 2012; Lai, 2011; Leclercq & Poumay, 2004; St-Pierre, 2004).

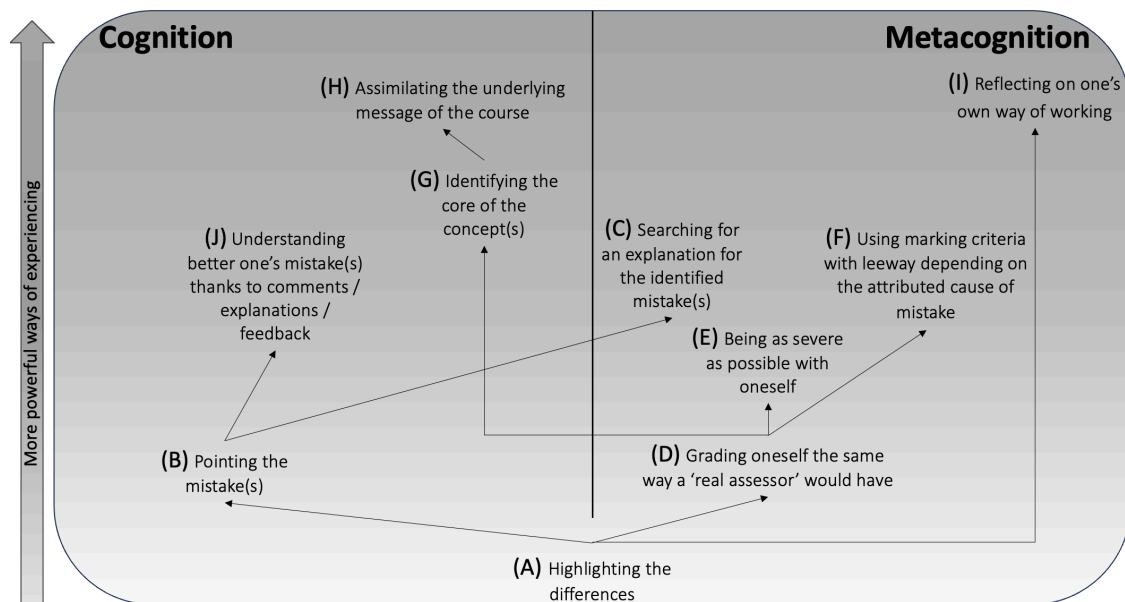


Fig. 2 Outcome space of the nested ways of experiencing comparison learning processes

## Discussion

### *On the distinction between explicit and implicit comparisons*

Implicit CLPs do not lead to as many different ways of experiencing—in this case, only two: (I) and (J)—as explicit ones do—here, eight: (A), (B), (C), (D), (E), (F), (G) and

(H). One crucial element here is the fact that when a comparison is implicit, (almost) no time frame is allocated to it; no concrete action is needed on the part of students, therefore there is not enough time for very powerful ways of experiencing it to form and the experience remains at a ‘basic stage’, similar for all students. In line with Nicol's (2021) call for greater attention to comparison processes in the learning context, this finding highlights the importance of making CLPs explicit.

### ***On the distinction between analogical and analytical comparisons***

When comparing the types of learning experiences generated by analogical comparisons—i.e. (A), (B), (C) and (I)—as opposed to the ones offered by analytical comparisons—i.e. (D), (E), (F), (G), (H), (J)—it is to be noted that analogical comparisons, when they prompt more powerful ways of experiencing, result in a self-generated understanding of the object of learning while it is not the case with analytical ones. Indeed, in the explicit analogical comparison in the thermodynamics class, it is by looking at the nature of the gap between their answer and the correct one that students have understood the origin of their mistake. Similarly, the students’ ways of experiencing the implicit analogical CLPs with their peers’ answers lies in their own ability to reflect on the cause or origin of an identified difference between what they did—or the way they did it—and what their peer did. Henceforth, one can argue that analogical comparisons are learning processes propitious to self-regulation (Butler & Winne, 1995), one desired effect of SA (Andrade, 2019; Brown & Harris, 2013; Lin-Siegler et al., 2015). On the opposite, analytical comparisons seem to promote a deeper understanding of the course stakes (see (G), (H) and (J)). These results tend to contradict with Panadero & Romero (2014) who posited (though in a different context) that rubrics are better tools than no-rubric to foster self-regulated learning. Indeed, what

the results of the present study indicate is that rubric (analytical comparisons) vs. no-rubric (analogical comparisons) seem to yield *different types of learning*: profound cognitively speaking or self-generated. This new finding should prompt further research to broaden Panadero & Romero's (2014) answer to the question: 'to rubric or not to rubric?', also taking into account the possible positive impact of resorting to multiple analogical comparisons with productions of various qualities (similarly to what was experimented by Lin-Siegler et al., 2015 and what permit comparative judgement digital tools when used in the context of peer assessment, see for example: Bartholomew et al., 2022; Jones & Alcock, 2014; Potter et al., 2017).

### ***On the disciplinary contexts***

The two observations carried took place in what is commonly called 'hard science' courses, and one may wonder whether the same types of CLPs could occur in other disciplinary contexts, where traditionally less positivist approaches are adopted. As shown above, the two observed SA practices were based on different assumptions: in the computer science class an answer is considered 'correct' as long as it 'works' (but can be of various quality, depending on its clarity, simplicity, etc.) whereas in the thermodynamics one, the conception of correctness is closer to a strictly positivist (or dualistic) view: correct vs incorrect (even though there is some leeway: e.g. 'approximation mistakes' are tolerated). The findings reported here show that in both SA activities, all four types of CLPs were present (implicit/explicit, analogical/analytical; see tables 1 & 2). This seems to indicate that they are pertinent to analyse SA practices regardless of the underlying conception of what constitutes 'quality' or 'correctness' in the field, hence in varied disciplinary contexts. However, it is still to be investigated whether the prevalence of each type of CLPs differs depending

on the disciplinary epistemological principles.

***On the distinction between comparisons with a peer production and with a teacher production***

If students seem to be accustomed to and quite naturally engage with CLPs that involve a comparator emanating from their teacher, it is not the case for the ones originating from their peers. Indeed, both explicit and implicit CLPs involving peers' production have been overlooked by some students (see above). This is congruent with research on peer feedback (see for example : Cho & MacArthur, 2011 or Kaufman & Schunn, 2011). However, when experienced, these types of CLPs lead to the highest level of metacognitive thinking recorded in the study (I). A greater attention should be given to these CLPs to prompt students to engage with them, notably by making them explicit.

***On self-assessment consistency***

Andrade (2019) regrets that much of recent SA research still questions consistency (i.e. the correlation between students' evaluations and experts' ones) even though, according to her, there is no need for further inquiries on the matter. However, the present study sheds light on an often-disregarded side of consistency—the students' perspective—and, by doing so, raises novel questions.

Indeed, interestingly, none of the ways of experiencing self-marking (an explicit analytical CLP) categorised in this study denoted an aim to 'fairness', 'accuracy' or 'objectivity' of grading. More so, three postures were identified: 'grading oneself according to what a real assessor would have' (D), 'being as severe as possible with oneself' (E) and 'using marking criteria with leeway depending on the attributed cause of mistake' (F). The first one (D) directly echoes Andrade (2019) calls for the use of the

word ‘consistency’ rather than ‘accuracy’ in the context of SA: students are well-aware of the lack of what researchers call ‘inter-rater reliability’ and try to fit in their assessors’ shoes. Rather than pursuing an ‘objective’ marking: they aim for consistency with *their* teachers’ grading. In this sense, this posture is oriented toward a strategic goal: assessing how far (or close) they are to *passing their exam*, not how well they master a skill or a subject.

The two other postures reveal different goals. While ‘being as severe as possible...’ (E) denotes a will to push oneself to always do better, no matter what, ‘using marking criteria with leeway...’ (F) considers the perceived cause of mistake and aims at reflecting an accurate evaluation of the level of performance the student currently is at. If the purpose of SA practices “is to generate feedback that promotes learning and improvements in performance” (Andrade, 2019, p. 2), it is probably the two last postures that seem the most likely to make it happen. Concomitantly, they also are the *least likely to generate consistent self-marking*. Nonetheless this counter-intuitive and somewhat provocative statement, these findings throw doubts on a widespread idea that consistent self-marking equates powerful SA and that should warrant further inquiries on the matter.

## Conclusion

In this paper, the analysis of two naturalistic SA practices has shown that different types of CLPs are embedded in such practices. These can be analytically divided in four types: explicit analogical, explicit analytical, implicit analogical and implicit analytical (Nicol, 2021). Furthermore, through a phenomenographic analysis, different ways of experiencing these various CLPs were structured in an outcome space outlining their hierarchical structure (fig. 2). Together and in line with the aims of phenomenography

(Marton & Booth, 1997, pp. 218–220), these reveal some key takeaways that denote the deep, (meta)cognitive processes students must undertake to engage with CLPs in SA activities:

- Explicit CLPs tend to foster more powerful learning experiences for students. Therefore, it seems of great importance to prompt students to take action with regard to CLPs (making them explicit) and allocating them sufficient time for students to engage in more complex levels of thinking;
- Students tend to overlook CLPs involving their peers' productions, therefore making them explicit is even more important;
- Analytical CLPs (e.g. the use of rubrics) can generate a profound understanding of the course stakes while analogical CLPs (e.g. with a standard answer) favour understanding that originates from the self and therefore seem an appropriate method to foster autonomous learning;
- Consistency in self-marking should not be considered as a sole indicator of powerful SA.

These elements, by taking into account the different forms of CLPs put forward by teachers in their SA practices, contribute to enrich former SA process models (such as the ones put forward by Sargeant et al., 2010 or Yan & Brown, 2017) by describing different forms of students' engagement with external sources of feedback.

## **Acknowledgements**

This research has been approved by the ethical committee of the Faculty of Psychology, Logopaedics and Education Science of the University of Liege (reference 2223-039).

The authors would like to warmly thank the two teachers and eight students who accepted to take part in this study, which, matter-of-factly, could not have taken shape without their dedicated contribution.

The authors would also like to thank the two conscientious reviewers who have considerably contributed to broaden the reach of this paper with their remarkably pertinent comments.

### **Declaration of interest statement**

The authors report there are no competing interests to declare.

### **References**

- Åkerlind, G. S. (2005). Variation and commonality in phenomenographic research methods. *Higher Education Research & Development*, 24(4), 321–334.  
<https://doi.org/10.1080/07294360500284672>
- Åkerlind, G. S. (2023). Common misunderstandings of phenomenographic research in higher education. *Higher Education Research & Development*, 0(0), 1–16.  
<https://doi.org/10.1080/07294360.2023.2218804>
- Alfieri, L., Nokes-Malach, T. J., & Schunn, C. D. (2013). Learning Through Case Comparisons: A Meta-Analytic Review. *Educational Psychologist*, 48(2), 87–113. <https://doi.org/10.1080/00461520.2013.775712>
- Andrade, H. L. (2019). A Critical Review of Research on Student Self-Assessment. *Frontiers in Education*, 4.
- Bartholomew, S. R., & Jones, M. D. (2022). A systematized review of research with adaptive comparative judgment (ACJ) in higher education. *International Journal of Technology and Design Education*, 32(2), 1159–1190.  
<https://doi.org/10.1007/s10798-020-09642-6>
- 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>

- Berger, J.-L., & Büchel, F. (2012). Métacognition et croyances motivationnelles: Un mariage de raison. *Revue française de pédagogie. Recherches en éducation*, 179, Article 179. <https://doi.org/10.4000/rfp.3705>
- Boud, D., & Brew, A. (1995). Developing a typology for learner self-assessment practices. *Research and Development in Higher Education*, 18, 130–135.
- Bransford, J. D., & Schwartz, D. L. (1999). Rethinking Transfer: A Simple Proposal with Multiple Implications. *Review of Research in Education*, 24, 61–100. <https://doi.org/10.2307/1167267>
- Brown, G. T. L., & Harris, L. R. (2013). Student Self-Assessment. In J. H. McMillan (Ed.), *Sage handbook of research on classroom assessment* (pp. 367–393). Sage Publications.
- Butler, D. L., & Winne, P. H. (1995). Feedback and Self-Regulated Learning: A Theoretical Synthesis. *Review of Educational Research*, 65(3), 245. <https://doi.org/10.2307/1170684>
- Charmaz, K. (2006). *Constructing Grounded Theory. A practical Guide Through Qualitative Analysis* (Reprinted 2009). SAGE.
- Cho, K., & MacArthur, C. (2011). Learning by reviewing. *Journal of Educational Psychology*, 103(1), 73–84. <https://doi.org/10.1037/a0021950>
- Dewit, I., Rohaert, S., & Corradi, D. (2021). How can comparative judgement become an effective means toward providing clear formative feedback to students to improve their learning process during their product-service-system design project? *Design and Technology Education: An International Journal*, 26(3), 276–293.
- Entwistle, N. (1997). Introduction: Phenomenography in Higher Education. *Higher Education Research & Development*, 16(2), 127–134. <https://doi.org/10.1080/0729436970160202>
- Formasup. (n.d.). Retrieved 8 January 2024, from <http://formasup.uliege.be/>
- Hartell, E., & Buckley, J. (2021). Comparative Judgment: An Overview. In A. Marcus-Quinn & T. Hourigan (Eds.), *Handbook for Online Learning Contexts: Digital, Mobile and Open* (pp. 289–307). Springer International Publishing. [https://doi.org/10.1007/978-3-030-67349-9\\_20](https://doi.org/10.1007/978-3-030-67349-9_20)

- Jones, I., & Alcock, L. (2014). Peer assessment without assessment criteria. *Studies in Higher Education*, 39(10), 1774–1787.  
<https://doi.org/10.1080/03075079.2013.821974>
- Jones, I., & Wheadon, C. (2015). Peer assessment using comparative and absolute judgement. *Studies in Educational Evaluation*, 47, 93–101.  
<https://doi.org/10.1016/j.stueduc.2015.09.004>
- Kaufman, J. H., & Schunn, C. D. (2011). Students' perceptions about peer assessment for writing: Their origin and impact on revision work. *Instructional Science*, 39(3), 387–406.
- Kaufmann, J.-C. (2016). Commencer le travail: Rapidité, souplesse, empathie. In *L'entretien compréhensif* (4ème édition, p. pp.33-58). Armand Colin.
- Lai, E. R. (2011). *Metacognition: A Literature Review* (p. 40) [Research Report]. Pearson.
- Leclercq, D., & Poumay, M. (2004). *Une défintion opérationnelle de la métacognition et ses mises en oeuvre*. L'AIPU : 20 ans de Recherches et d'Actions Pédagogiques ; bilans et perspectives. 21ème congrès de l'AIPU, Marrakech, Maroc.
- Lejeune, C. (2019). *Manuel d'analyse qualitative*. De Boeck.
- Lin-Siegler, X., Shaenfield, D., & Elder, A. D. (2015). Contrasting case instruction can improve self-assessment of writing. *Educational Technology Research and Development*, 63(4), 517–537. <https://doi.org/10.1007/s11423-015-9390-9>
- Marton, F. (1981). Phenomenography—Describing conceptions of the world around us. *Instructional Science*, 10(2), 177–200. <https://doi.org/10.1007/BF00132516>
- Marton, F. (2015). *Necessary conditions of learning*. Routledge.
- Marton, F., & Booth, S. (1997). *Learning and awareness*. L. Erlbaum Associates.
- Naud, S., Sander, E., & Benetos, K. (2023). Usage pédagogique des analogies dans l'enseignement supérieur. *Revue internationale de pédagogie de l'enseignement supérieur*, 39(2). <https://doi.org/10.4000/ripes.4866>
- Nicol, D. (2021). The power of internal feedback: Exploiting natural comparison processes. *Assessment & Evaluation in Higher Education*, 46(5), 756–778.  
<https://doi.org/10.1080/02602938.2020.1823314>
- Panadero, E., Brown, G. T. L., & Strijbos, J.-W. (2016). The Future of Student Self-Assessment: A Review of Known Unknowns and Potential Directions.

- Educational Psychology Review*, 28(4), 803–830.  
<https://doi.org/10.1007/s10648-015-9350-2>
- Panadero, E., & Romero, M. (2014). To rubric or not to rubric? The effects of self-assessment on self-regulation, performance and self-efficacy. *Assessment in Education: Principles, Policy & Practice*, 21(2), 133–148.  
<https://doi.org/10.1080/0969594X.2013.877872>
- Panadero, E., Tapia, J. A., & Huertas, J. A. (2012). Rubrics and self-assessment scripts effects on self-regulation, learning and self-efficacy in secondary education. *Learning and Individual Differences*, 22(6), 806–813.  
<https://doi.org/10.1016/j.lindif.2012.04.007>
- Potter, T., Englund, L., Charbonneau, J., MacLean, M. T., Newell, J., & Roll, I. (2017). ComPAIR: A New Online Tool Using Adaptive Comparative Judgement to Support Learning with Peer Feedback. *Teaching & Learning Inquiry*, 5(2), 89–113.
- Ruchniewicz, H., & Barzel, B. (2019). Technology Supporting Student Self-Assessment in the Field of Functions—A Design-Based Research Study. In G. Aldon & J. Trgalová (Eds.), *Technology in Mathematics Teaching* (Vol. 13, pp. 49–74). Springer International Publishing. [https://doi.org/10.1007/978-3-030-19741-4\\_3](https://doi.org/10.1007/978-3-030-19741-4_3)
- Sargeant, J., Armson, H., Chesluk, B., Dornan, T., Eva, K., Holmboe, E., Lockyer, J., Loney, E., Mann, K., & Van Der Vleuten, C. (2010). The Processes and Dimensions of Informed Self-Assessment: A Conceptual Model: *Academic Medicine*, 85(7), 1212–1220. <https://doi.org/10.1097/ACM.0b013e3181d85a4e>
- Sidney, P. G., Hattikudur, S., & Alibali, M. W. (2015). How do contrasting cases and self-explanation promote learning? Evidence from fraction division. *Learning and Instruction*, 40, 29–38. <https://doi.org/10.1016/j.learninstruc.2015.07.006>
- St-Pierre, L. (2004). L’habileté d’autoévaluation: Pourquoi et comment la développer ? *Pédagogie collégiale*, 18(1), 33–38.
- Svensson, L. (1997). Theoretical Foundations of Phenomenography. *Higher Education Research & Development*, 16(2), 159–171.  
<https://doi.org/10.1080/0729436970160204>
- Taras, M. (2010). Student self-assessment: Processes and consequences. *Teaching in Higher Education*, 15(2), 199–209. <https://doi.org/10.1080/13562511003620027>

- Yan, Z. (2020). Self-assessment in the process of self-regulated learning and its relationship with academic achievement. *Assessment & Evaluation in Higher Education*, 45(2), 224–238. <https://doi.org/10.1080/02602938.2019.1629390>
- Yan, Z., & Brown, G. T. L. (2017). A cyclical self-assessment process: Towards a model of how students engage in self-assessment. *Assessment & Evaluation in Higher Education*, 42(8), 1247–1262. <https://doi.org/10.1080/02602938.2016.1260091>
- Yan, Z., Lao, H., Panadero, E., Fernández-Castilla, B., Yang, L., & Yang, M. (2022). Effects of self-assessment and peer-assessment interventions on academic performance: A meta-analysis. *Educational Research Review*, 37, 100484. <https://doi.org/10.1016/j.edurev.2022.100484>
- Yan, Z., Panadero, E., Wang, X., & Zhan, Y. (2023). A Systematic Review on Students' Perceptions of Self-Assessment: Usefulness and Factors Influencing Implementation. *Educational Psychology Review*, 35(3), 81. <https://doi.org/10.1007/s10648-023-09799-1>