

## EMOTIONS LINKED TO CERTAINTY-BASED SELF-ASSESSMENT: A MULTIDISCIPLINE EXPERIENCE

Ana Remesal<sup>1</sup>, María José Corral<sup>2</sup>, Iria SanMiguel<sup>2</sup>, Patricio García-Minguez<sup>3</sup>,  
Tomas Macsotay<sup>4</sup>, & Ernesto Suárez<sup>5</sup>

<sup>1</sup>*Department of Cognition, Developmental and Educational Psychology, University of Barcelona (Spain)*

<sup>2</sup>*Department of Psychobiology, University of Barcelona (Spain)*

<sup>3</sup>*Department of Economics, University of Barcelona (Spain)*

<sup>4</sup>*Department of Humanities, University Pompeu Fabra (Spain)*

<sup>5</sup>*Department of Law, University Pompeu Fabra (Spain)*

### Abstract

This exploratory study presents data from various curricular areas in Higher Education: Primary Teacher Education, Secondary Teacher Education, Psychobiology, Microeconomics, Art History and Law. In all these areas, the instructors implement the same instructional program during a one-semester module to improve students' self-regulated learning behaviour. The project had two main goals: (1) enhancing students' learning engagement through certainty-based self-assessment and (2) researching the metacognitive and affective (motivational and emotional) implications of such instructional program in students' learning behaviour. Over 1000 students participate in these courses, responding to three progressive (discipline-related) multiple-choice online learning tests with a knowledge-certainty component at three different moments during one semester, incorporating the certainty-based marking (CBM). Certainty declaration activates deep metacognitive and emotional processes. We implemented the testing system in these courses only with a diagnostic and formative purpose. After each testing point, students responded to an accompanying reflection questionnaire that allowed the collection of their emotional reactions in a Likert-scale form. The students manifested positive/negative and activating/deactivating retrospective emotions (joy, pride, sadness, shame, anger) and prospective emotions (hope, boredom, indifference, fear, anxiety) in relation to the certainty-based self-assessment experience. Results indicate that the certainty-based self-assessment program triggered all of these emotions in different amounts. These results are positive news concerning the students' engagement in learning and particularly motivated self-regulated learning. This paper presents preliminary results concerning the students' emotional experience with this innovative diagnostic self-assessment program. Differences and similarities between areas and educational levels are exposed. The main result points to similarities outweighing differences.

**Keywords:** *Self-assessment, certainty-based marking, formative assessment, students' emotions, higher education.*

---

### 1. Introduction

This exploratory study presents a first approach in a multidisciplinary context. This multidisciplinary context is urgently needed in higher education to compensate for the over-exposure of education students (Quinlan et al., 2013). In our study, students from Primary Teacher Education, Secondary Teacher Education, Psychobiology, Microeconomics, Art History and Law were invited to participate in an instructional experience grounded on the so-called certainty-based marking (CBM) (Gardner-Medwin, 2019). Leclercq proposed this assessment strategy underlying CBM long before (Leclercq, 1982, 2003) to endorse metacognition during learning assessment processes (Chamberland & St-Onge, 2013). The CBM algorithm alters the habitual grading scale (1 to 10 points); thus, both students and instructors need to readapt to the new scale ranging between [-6\*items] and [+3\*items]. We implemented the testing system in our project only with a diagnostic and formative purpose. This knowledge-certainty component consists of a declaration of the subjective level of certainty/doubt about the correctness of one's answer to each question item. This certainty declaration activates deep metacognitive and emotional processes. A special algorithm inserted in the virtual campus adjusts

resulting grades to the connection of each of three levels of declared certainty (low, middle, high) and the correctness of response (error, hit). Up to now, research in higher education revolving around this specific assessment strategy is present in a variety of areas, such as the medical area (Smrkolj, Bancov & Smrkolj, 2022) or engineering (Yuen-Reed & Reed, 2015), among many others.

Nevertheless, all these previous studies present two critical areas of shortage. First, they are mono-disciplinary and only provide results on each particular area under different instructional conditions, making comparison impossible in practice. Second, they present merely CBM results but disconnected from other likely related psychological processes and constructs to understand the effects of such an assessment strategy. In this sense, we present a pioneer study. Our project pursued two main goals: (1) enhancing students' learning engagement through certainty-based self-assessment and (2) researching the metacognitive and affective (motivational and emotional) implications of such instructional program in students' learning behaviour. Particularly, we were interested in the students' emotional reactions taking positive-negative but also activating-deactivating emotions into account since they all affect the learning behaviour (Pekrun & Linnenbrink-Garcia, 2012).

## 2. Research questions and design

Despite the project aiming to analyse a longitudinal process, for this paper, we focus on the students' first experience with CBM-based self-assessment. Thus, we concentrate on the following research questions:

- RQ1: What is students' satisfaction with this first encounter with CBM?
- RQ2: What emotional reactions do students report at their first experience with CBM-self-assessment?
  - RQ2a: Are there differences in students' emotional reports regarding *educational level*?
  - RQ2b: Are there differences in students' emotional reports regarding *disciplinary areas*?

## 3. Method

The same instructional design took place in all participating courses. Students responded to a series of knowledge tests, including ten multiple-choice questions and a reflection questionnaire to report emotions triggered by the response experience under CBM conditions. The CBM algorithm implies the challenge of the habitual marking scheme. Marks are no longer between 0-10 points (as traditionally in the participating institutions) but move in a range between [-60, +30]. For the students' personal, qualitative interpretation of results, four quality levels of results were suggested to them: level 0 (-60 up to 0 points), level 1 (1-10 points), level 2 (11-20 points), level 3 (21-30 points). Further, students reported their satisfaction with this first experience (1-10 points) and their emotions (Likert scale 0-5 points). For the sake of space limit, in this paper, we will focus only on *retrospective* emotions, both positive (joy, pride, relief) and negative (shame, sadness, anger), as an individual evaluation of the CBM-based results. We applied a non-parametric test (Kruskal-Wallis') according to the ordinal nature of our data and the very unequal but also meagre response rate of two groups of students (Economics and Art History).

### 3.1. Sample

In the six course modules, 1492 students were enrolled; 1161 (77.81%) voluntarily responded to the first knowledge test in their respective course. After applying exclusion criteria (cases that were not first attempt, had missing data or used less than 20% of given time for response), 988 valid responses remained. However, when invited to contribute further to the study by responding to additional demographic and reflection questionnaires, the response fell to a final sample of 411 students. Table 1 presents descriptive data from the final participating sample of students. The majority were women (69.8%), with an average age of nearly 25 years. The response ratio of valid responses was 41.6%; however, it was very different depending on the disciplinary area, to the extreme that no students of Law responded to the reflection questionnaire. Consequently, this disciplinary area had to be discarded from further empirical analysis.

## 4. Preliminary results

Table 2 presents descriptive results about the students' first encounter with the CBM-based knowledge tests in their respective courses. The response time ratio results show that students had sufficient time for response (either 10 or 15 minutes, depending on each instructor's decision for each particular course). Having enough time for response is a condition for stress control, increasing the

probability that an additional feeling of time pressure does not worsen emotional reactions. In that table of results, we can observe that Psychobiology students had the best learning results, followed by prospective primary and secondary teachers. In contrast, Art History and Economics students had the lowest results; unsurprisingly, these latter students showed a minor participation ratio in the voluntary reporting of emotions (see Table 1).

Table 3 exposes results on the report of retrospective emotions as a reaction to the CBM-based self-assessment experience. Remarkably, the highest mean value for positive and negative emotions is 2.3 (joy, pride and shame), which is still low. In coherence with the direct CBM results presented in Table 2, in Table 3, we see that the students with higher results report more positive retrospective emotions (joy, pride, relief) than the students of Art History and Economics, who report more on negative retrospective emotions (shame, sadness, anger).

#### 4.1. RQ1: Students' reported satisfaction

Altogether, students report a relatively moderate to low satisfaction ( $M = 5.3$ ;  $SD = 2.5$ ) at this first encounter with the CBM system. When looking deeper into the two variables considered in this study:

- *Satisfaction. Educational level.* We found significant differences regarding subjects' satisfaction with the CBM results ( $\chi^2(1, 411) = 9.056$ ,  $p = .0026$ ); however, with a minimal effect size.
- *Satisfaction. Disciplinary area.* We found a significant difference regarding subjects' satisfaction with the CBM results ( $\chi^2(4, 411) = 20.323$ ,  $p = .0004$ ). In addition, Psychobiology students were more satisfied (6.5) with a Cohen's  $d$  of 0.48; the same effect size appeared for Economics students in their lack of satisfaction (4.1).

#### 4.2. RQ2a: Students reported emotional reactions regarding educational level

- *Emotional value (positive/negative).* We found significant differences for *joy* ( $\chi^2(1, 411) = 7.063$ ,  $p = .0079$ ), with very small effect size. There was no difference for *pride* or *relief*. We found no differences regarding negative emotions between students at Bachelor and Master degree.
- *Potential emotional effect (activating/deactivating).* At Master's level, positive emotions with activating effect (*joy/pride*) present a significant difference over deactivating *relief* at  $\alpha = 0.05$  ( $\chi^2(2, 305) = 7.34$ ,  $p = .025$ ), while negative emotions with activating effect present a more significant difference ( $\chi^2(2, 305) = 35.63$ ,  $p = 0$ ). At Bachelor's level, we do not identify any significant difference in activating (*shame/anger*) or deactivating emotions (*sadness*).

#### 4.3. RQ2b: Students reported emotional reactions regarding disciplinary areas

- *Emotional value (positive/negative).* We found significant differences for *joy* ( $\chi^2(4, 411) = 24.61$ ,  $p = .0001$ ), *pride* ( $\chi^2(4, 411) = 24.781$ ,  $p = .0001$ ), and *relief* ( $\chi^2(4, 411) = 14.313$ ,  $p = .0064$ ). Primary Teacher students' results had a small effect size regarding higher *joy* ( $M = 2.3$ ,  $d = 0.46$ ) and a medium effect size for more *pride* ( $M = 2.3$ ,  $d = 0.62$ ). In contrast, Art History showed a small effect size for low *joy* ( $M = 1.1$ ,  $d = 0.46$ ), and Economics students reported the lowest positive emotions with a medium effect size ( $M_{\text{joy}} = 0.9$ ,  $d = 0.62$ ;  $M_{\text{pride}} = 0.7$ ,  $d = 0.62$ ;  $M_{\text{relief}} = 0.4$ ,  $d = 0.77$ ). In coherence with the previous results on positive emotions, students of Art History and Economics reported the highest values of negative emotions. However, we found significant differences only for *shame* ( $\chi^2(4, 411) = 19.578$ ,  $p = .0006$ ). We found no difference in the report of *sadness* or *anger*. In the case of shame, Primary Teacher students had the lowest report with a medium effect size ( $M = 0.6$ ,  $d = 0.54$ ), while Art History and Economics had the highest, with medium and small effect sizes respectively ( $M_{\text{AH}} = 2.3$ ,  $d = 0.77$ ;  $M_{\text{E}} = 1.9$ ,  $d = 0.46$ ).
- *Potential emotional effect (activating/deactivating).* We do not identify any significant difference in activating (*shame/anger*) or deactivating emotions (*sadness*) across areas.

## 5. Discussion and conclusions

In this paper, we address only a tiny fraction of the project that we are currently undertaking. In contrast with previous research on emotions related to assessment, our study presents evidence of actual emotions in a natural educational setting, as opposed to laboratory, abstract experiments or off-context surveying (Bieg, Goetz, & Lipnevich, 2014). We have yet to find any previous literature to compare our results with, as previous publications on CBM do not inquire about related psychological

processes or constructs in connection with this alternative assessment experience. Hence, this section will offer our reflections on the presented results. These preliminary results lead us to several considerations:

First, there are low, perhaps restrained, emotional reactions, both positive and negative value, activating and deactivating. It is important to note that the testing system had merely a formative purpose, having the resulting marks no accreditation value for the students. We hypothesise stronger emotional reactions (both positive and negative) if the CBM marks would have a summative effect, based on literature on achievement emotions (Pekrun & Linnenbrink-Garcia, 2012).

Secondly, we found little difference in our results regarding educational level. Students at Bachelor's degrees and Master's degrees reacted similarly to this first encounter with the CBM-self-assessment system. Therefore, in that sense, this instructional strategy is equally valid for both levels. Considering the activating/deactivating potential of perceived emotions in the learning process, master students reported a more significant difference between activating and deactivating emotions, both positive and negative, than bachelor students. We need further investigation regarding the detail of the subject's age. In our sample, the mean age of master students was about 27 years, and the mean age of bachelor students was close to 20 years. However, there is one-third of students between 30-50 years of age in our sample. In the subsequent analysis, we shall pay attention to chronological age rather than educational level to identify possible differences between late adolescents and young to mid-adults' perceptions (Novacek, 2013).

Finally, we need to make some remarks on the low participation of students in some of the disciplinary areas. Notably, the two courses presenting the higher participation rate are Primary Teacher (70%) and Secondary Teacher (67%) education, both courses directly addressing in their programs learning contents related to learning processes themselves; meanwhile, the disciplinary areas different from education present a much lower response. In other words, students from the education area are much more prone to participate in the study. In contrast, students off of education might lack a reflection culture in relation with their learning process. Indeed, we need much more research into this; we need to learn more about the inherent complexity of teaching and learning in higher education and the specificity of content areas (Gotez et al. 2006; Quinlan, 2016). In our project, deeper analysis is still pending to consider a variety of individual variables regarding demographic aspects (age, workload, family responsibilities, personal tendency to risk-taking) and educational aspects (conceptions of assessment, mathematical competence regarding probability situations). In addition, we need to explore prospective emotions, as well as calibration and metacognitive reactions (attribution of results and plans for improvement or maintenance) to determine how the CBM assessment strategy might endorse self-regulated learning.

Table 1. Participants' description by disciplinary area.

Disciplinary area	Within area % response (decreasing)	Sample across area % response	Female %	Male %	Age M(SD)	Educational level
Prim. Teacher (n=37)	67.3	9.0	75.7	24.3	19.9(2.4)	Bachelor
Sec. Teacher (n=305)	67.2	74.2	66.5	33.5	26.7(6.6)	Master
Art History (n=12)	14.1	2.9	83.3	16.7	21.7(3.9)	Bachelor
Psychobiology (n=47)	7.7	11.4	89.4	10.6	20.1(5.4)	Bachelor
Economics (n=12)	5.7	2.4	40	60	20.8(3.3)	Bachelor
Law	0					Bachelor
TOTAL (n=411)	41.6	100	69.8	30.2	24.8(6.7)	

Table 2. Results on CBM (first description).

Disciplinary area	CBM result (decreasing) M(SD)median	Response time (%) M(SD)	Points [-60, +30]		Declared level of result (%)			
			Min.	Max.	Level 0	Level 1	Level 2	Level 3
Psychobiology	9.2(7.7)11	52.8(21.9)	-17	+22	10.6	36.2	34.0	19.1
Prim. Teacher	6.9(6.3)8	64.3(19.5)	-13	+19	13.5	51.3	29.7	5.4
Sec. Teacher	4.7(9.0)6	57.4(19.3)	-27	+24	25.6	46.2	22.3	5.9
Art History	4.1(4.1)4	59.5(13.2)	-4	+11	--	91.7	8.3	--
Economics	0.9(16.1)5	66.6(23.4)	-33	+21	30	50	10	10
Bachelor (n=106)	7.1(8.4)8	58.9(20.9)	-33	+22	12.3	49.0	27.4	11.3
Master (n=305)	4.7(9.0)6	57.4(19.3)	-27	+24	25.6	46.2	22.3	5.9
TOTAL	5.5(8.9)6	57.8(19.7)	-33	+24	22.1	46.9	23.6	7.3

Table 3. Results on reported satisfaction and emotions.

Disciplinary area	Satisfaction (decreasing) M(SD)median	Retrospective emotions, positive M(SD) 6 point Likert (0-5)			Retrospective emotions, negative M(SD) 6 point Likert (0-5)		
		Joy	Pride	Relief	Shame	Anger	Sadness
Psychobiology	6.5(2.6)7	2.2(1.3)	2.0(1.3)	1.7(1.3)	1.1(1.3)	0.6(1.2)	0.7(1.2)
Prim. Teacher	6.1(1.9)6	2.3(1.1)	2.3(1.3)	1.9(1.4)	0.6(1.2)	0.8(1.0)	0.7(1.1)
Sec. Teacher	5.0(2.5)5	1.6(1.3)	1.4(1.2)	1.3(1.3)	1.3(1.3)	0.7(1.0)	0.8(1.0)
Art History	4.5(2.1)4	1.1(1.6)	1.1(1.8)	1.2(1.4)	2.3(1.8)	1.3(1.4)	1.6(1.3)
Economics	4.1(2.9)3.5	0.9(1.1)	0.7(1.1)	0.4(0.7)	1.9(1.7)	1.3(1.2)	0.9(1.2)
Bachelor (n=106)	5.9(2.5)6	2.0(1.3)	1.9(1.4)	1.6(1.4)	1.2(1.5)	0.8(1.2)	0.8(1.2)
Master (n=305)	5.0(2.5)5	1.6(1.2)	1.4(1.3)	1.3(1.3)	1.3(1.3)	0.7(1.0)	0.8(1.1)
TOTAL	5.3(2.5)5	1.7(1.3)	1.5(1.3)	1.4(1.3)	1.3(1.3)	0.7(1.1)	0.8(1.1)

### Acknowledgement

This study was funded by the University of Barcelona. Research grant REDICE22-3060.

### References

- Bieg, M., Goetz, T., & Lipnevich, A. A. (2014). What students think they feel differs from what they really feel—academic self-concept moderates the discrepancy between students' trait and state emotional self-reports. *PLoS one*, 9(3), e92563.
- Chamberland, M., & St-Onge, C. (2013). Back to basics: keeping students cognitively active between the classroom and the examination. *BMC Medical Education*, 47(7), 641-643. doi:10.1186/medu.12218
- Gardner-Medwin, T. (2019). Certainty-Based marking: stimulating thinking and improving objective tests. In C. Bryan, & K. Clegg (Eds), *Innovative Assessment in Higher Education* (pp. 141-150). NY: Routledge.
- Leclercq, D. (1982). Confidence marking, its use in testing. *Evaluation in Education*, 6(2), 163-287. doi:10.1016/0191-765X(82)90011-8
- Leclercq, D. (2003). *Diagnostic cognitif et métacognitif au seuil de l'université. Le projet MOHICAN mené par les 9 universités de la Communauté Française Wallonie Bruxelles*. Liège: Editions de l'université de Liège. <http://hdl.handle.net/2268/17837>
- Novacek, P. (2013). Confidence-based assessments within an adult learning environment. In *IADIS International Conference on Cognition and Exploratory Learning in Digital Age (CELDA 2013)*, (pp. 403-406).
- Pekrun, R., & Linnenbrink-Garcia, L. (2012). Academic emotions and student engagement. In S. L. Christenson, A. L. Reschly, & C. Wylie (Eds.), *Handbook of research on student engagement* (pp. 259-282). New York, NY: Springer.
- Quinlan, K. M. (2016). How emotion matters in four key relationships in teaching and learning in higher education. *College Teaching*, 64(3), 101-111.
- Quinlan, K. M., Male, S., Baillie, C., Stamboulis, A., Fill, J., & Jaffer, Z. (2013). Methodological challenges in researching threshold concepts: a comparative analysis of three projects. *Higher Education*, 66, 585-601.
- Smrkolj, S., Bancov, E., & Smrkoli, V. (2022). The reliability and medical students' appreciation of certainty-based marking. *International Journal of Environmental Research in Public Health*, 19, 1706. doi:10.3390/ijerph19031706
- Yuen-Reed, G., & Reed, K. B. (2015). Engineering Student Self-Assessment through Confidence-Based Scoring. *Advances in Engineering Education*, 4(4).