



OPEN ACCESS

EDITED BY

Gavin T. L. Brown,
The University of Auckland,
New Zealand

REVIEWED BY

Zi Yan,
The Education University of Hong Kong,
Hong Kong SAR, China
Elena Tikhonova,
Peoples' Friendship University of Russia,
Russia

*CORRESPONDENCE

Ana Remesal
✉ aremesal@ub.edu

SPECIALTY SECTION

This article was submitted to
Assessment,
Testing and Applied Measurement,
a section of the journal
Frontiers in Education

RECEIVED 03 December 2022

ACCEPTED 23 February 2023

PUBLISHED 30 March 2023

CITATION

Remesal A and Estrada FG (2023) Synchronous
self-assessment: First experience for higher
education instructors.
Front. Educ. 8:1115259.
doi: 10.3389/educ.2023.1115259

COPYRIGHT

© 2023 Remesal and Estrada. This is an open-
access article distributed under the terms of
the [Creative Commons Attribution License
\(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction
in other forums is permitted, provided the
original author(s) and the copyright owner(s)
are credited and that the original publication in
this journal is cited, in accordance with
accepted academic practice. No use,
distribution or reproduction is permitted which
does not comply with these terms.

Synchronous self-assessment: First experience for higher education instructors

Ana Remesal* and Flor G. Estrada

Departamento de Cognición, Desarrollo y Psicología de la Educación, Facultad de Psicología,
Universidad de Barcelona, Barcelona, Spain

We present a qualitative study of four cases of university instructors (teacher educators) implementing synchronous self-assessment (SSA). SSA consists of an innovative assessment strategy during written exam situations, which highlights the students' voice and agency, giving it greater weight in the power balance traditionally established between instructors and students in classroom assessment practices. In this article, we focus on the effects and pedagogical potential of this assessment strategy from instructors' point of view. In our study, three instructors were novices in implementing this strategy; the fourth instructor had several years of experience with it. The four instructors agreed on basic design features for an end-of-semester exam offered in four groups of first-year students of the same shared program at a Bachelor's degree for Kindergarten Educator and Primary School Teacher. The instructors were individually interviewed after the assessment session in their course and the exams were gathered for analysis. Content and discursive analysis was carried out on the data. Results show substantial differences in the evaluative artefacts (instructors' exams) in terms of cognitive demand and formative assessment potential, and point to noticeable needs for professional development in pursuit of assessment literacy in Higher Education.

KEYWORDS

self-assessment, higher education, assessment literacy, teachers' conceptions of assessment, innovation, exam design

1. Introduction

Learning assessment is still today one of the most significant challenges for instructors, regardless of educational level. In an international context of curricular renewal focused on the development of competencies, new challenges arise for teaching and assessment. In this article, we want to present an innovative proposal that leads us to analyse the potential for improvement of particular exam situations as a whole (Hernández Nodarse, 2007), unlike other previous proposals that refer to the nature of specific assessment activities. (e.g., Villarroel et al., 2021).

Multiple-choice standardised exams are frequent in higher education, especially in non-humanistic areas (e.g., López Espinosa et al., 2014; Roméu and Díaz Quiñones, 2015; Herrero and Medina, 2019; Imbulpitiya et al., 2021). Interest in test design is relatively recent in the university context. Up to now, literature on classroom assessment does not recognise a test's unitary value as an interactive classroom experience (Hernández Nodarse, 2007) but analyse assessment tasks in a non-contextual and isolated way. In our approach, however, exams are precisely taken as a unitary interactive experience in the teaching-learning process, as we will explain later.

Xu and Brown's (2016) most recent review of teacher education for assessment literacy is the first to highlight the importance of instructors' beliefs and conceptions about assessment itself, as well as their emotional experience. The 13 key points highlighted by Popham 2 years earlier (2014) did not yet include them. University instructors' assessment literacy still needs to be improved. Many instructors lack specific initial training for assessing unless their field of research, or their individual motivation, leads them to explore the teaching and learning processes (López Espinosa et al., 2014). Thus, we face a deficient field of knowledge as a starting point.

In this article, we will first present the essence of our proposed evaluative strategy. Second, we report the results of a quadruple case study in the context of higher education; finally, we will provide a set of theoretical reflections and proposals for open lines of research.

1.1. What is synchronous self-assessment?

Synchronous Self-Assessment (SSA) is an innovative classroom assessment strategy (Remesal et al., 2019; Remesal, 2021) that allows the students to make crucial and impact-full decisions within their assessment process, recognising and valuing their learning, thus emphasising their agency. Ideally, this strategy comes to life in written exam situations. Unlike other assessment situations or activities, the exam has particular characteristics that make it ideal for SSA. First, it is an explicit assessment situation, where all the participants are aware of a series of rules that aim at exposing the learning generated or elaborated during a specific time. Second, it happens under a certain time pressure in which students know that their best version of the learning effort made is actually at stake. Therefore, the quality of synchronicity does not refer to technological aspects on this occasion. It indicates the simultaneity of the hetero –or external–assessment processes as led by the instructor and the self –or internal–assessment processes led by the students in the shared classroom interactive space and concerning the same assessment activities.

The students' active role comes to the foreground through two decisions they must take. First, they must select a series of activities from a total to solve. The number of activities to choose and solve will depend on the educational level and the time available for completion. Secondly, they must also choose a weighted grading for their solved activities (equal scores, maximum or minimum difference). The grading options may also vary regarding the educational level and duration of the exam. Unlike the proposals about self-grading (Crowell, 2015), SSA provokes a qualitative comparison of personal performance in a short series of solved assessment activities immediately after resolution. SSA relates to recent conceptual proposals such as the evaluative judgement (Tai et al., 2018). Since this concept is not associated with any specific pedagogical measure, we propose that SSA would offer a concrete way to educate towards this evaluative judgement, as students make a value judgement about their performance in a short series of activities in a comparative manner, applying personal quality criteria that they may construct throughout the learning process.

SSA rebalances the power relation between instructor and student in the assessment situation, with the student assuming much more responsibility in strategic aspects such as “By which activities am I going to demonstrate the learning I have achieved?” and “How will my performance be valued?” With this innovative proposal, students actively participate in their learning assessment. SSA

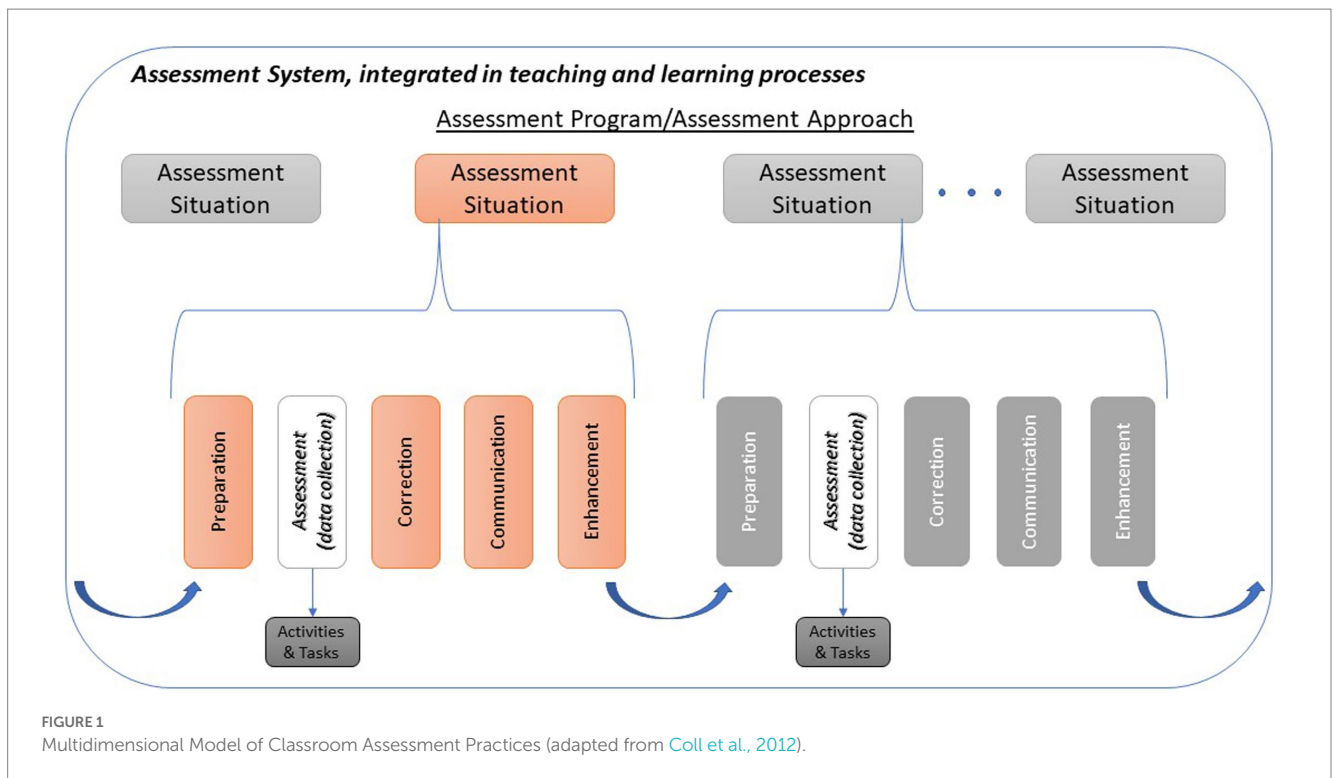
launches deep metacognitive processes that potentially lead the students to a new and greater awareness of everything learned. Thus, facilitating internal self-assessment in a natural way (Nicol, 2021) as a subjective phenomenon that accompanies the entire learning process and, of course, its evaluation (Yan and Brown, 2017; Yan, 2020), opening up new learning opportunities (Yan and Boud, 2021) and, eventually, promoting self-awareness and emotional self-management. Thus, we propose that the exam situation, as a moment of purposeful assessment, explicitly shared between the participants, with the goal of external demonstration of the maximum learning achieved, and all this under a certain time pressure (Remesal et al., 2022), is an ideal opportunity to encourage the student's agency.

1.2. The multidimensional model of classroom assessment practices

According to the multidimensional model of classroom assessment practices (MMCAP) (Coll et al., 2012), we distinguish five different moments or segments of interactivity in classrooms related to learning assessment: (1) preparation, (2) assessment *de facto* or data collection, (3) correction, evaluation or grading, (4) feedback, and (5) posthoc pedagogical enhancement. Each of these segments constitutes links in a chain with particular actions, roles and contingent compromises by the protagonists of the educational process, teacher and learner. A whole set of these five links constitute an *assessment situation*. Figure 1 presents this model. A preliminary phase of instructional design is necessary. It encompasses all five steps, and more than those, the bigger picture of all the assessment situations within a course (a term, a unit, etc., whatever pedagogical unit we may refer to), which Coll et al. define as *assessment program*.

A series of design decisions must be taken concerning the whole assessment situation. For example: What is the object of assessment? What are the learning goals relative to this assessment occasion? How are we assessing? With what activities, instruments, and resources? What purposes do we follow? When is this assessment situation taking place? Referring to which evaluation criteria and thresholds will performance be evaluated? What rules shall regulate students' participation? What kind of feedback shall students receive? Subsequently, the instructor implements this evaluative situation in its different parts:

- In the (1) *preparation segment*, the instructor aims to facilitate the students' best possible performance. Typically, contents will be recapitulated in the preparation segment, and assessment rules, learning strategies, and coping strategies will be shared. The location in time of the preparation segment might occur some minutes before the data collection segment, or a week before, for instance.
- Next, there is (2) *the data collection segment*. In this central segment, traditionally called ‘assessment’ *de facto*, students and teachers are both aware of the purpose of data collection, namely, to gather evidence of learning.
- Afterwards, the instructor evaluates the collected data in (3) *the correction segment*, applying the previously established assessment criteria.
- In fourth place, the teacher offers (4) *feedback to the students*, according to the previous decision on how, what, when, etc.,



- Eventually, teachers will make certain decisions as a consequence of the learning results in (5) *the posthoc enhancement segment*. Whether these decisions are pedagogical or formative (like adjusting teaching resources, looking for extra activities) or summative (like merely handing out a set of grades and moving on with a course) depends on many factors, starting with the teacher's conceptions of assessment, but also related to contextual or systemic constraints (Remesal, 2011).

This multidimensional model of classroom assessment proposes a basic scheme to analyse and understand educational interaction for assessing learning. It allows researchers and practitioners to reflect on it and eventually change whatever needs improvement. As many variations as we can think of, they all could be located in this basic scheme. For example, students' co-evaluation would be a variant of the correction segment, in which students are called in. Also, students' participation in a possible co-construction of assessment activities or evaluation criteria would be a variant of the preparation segment.

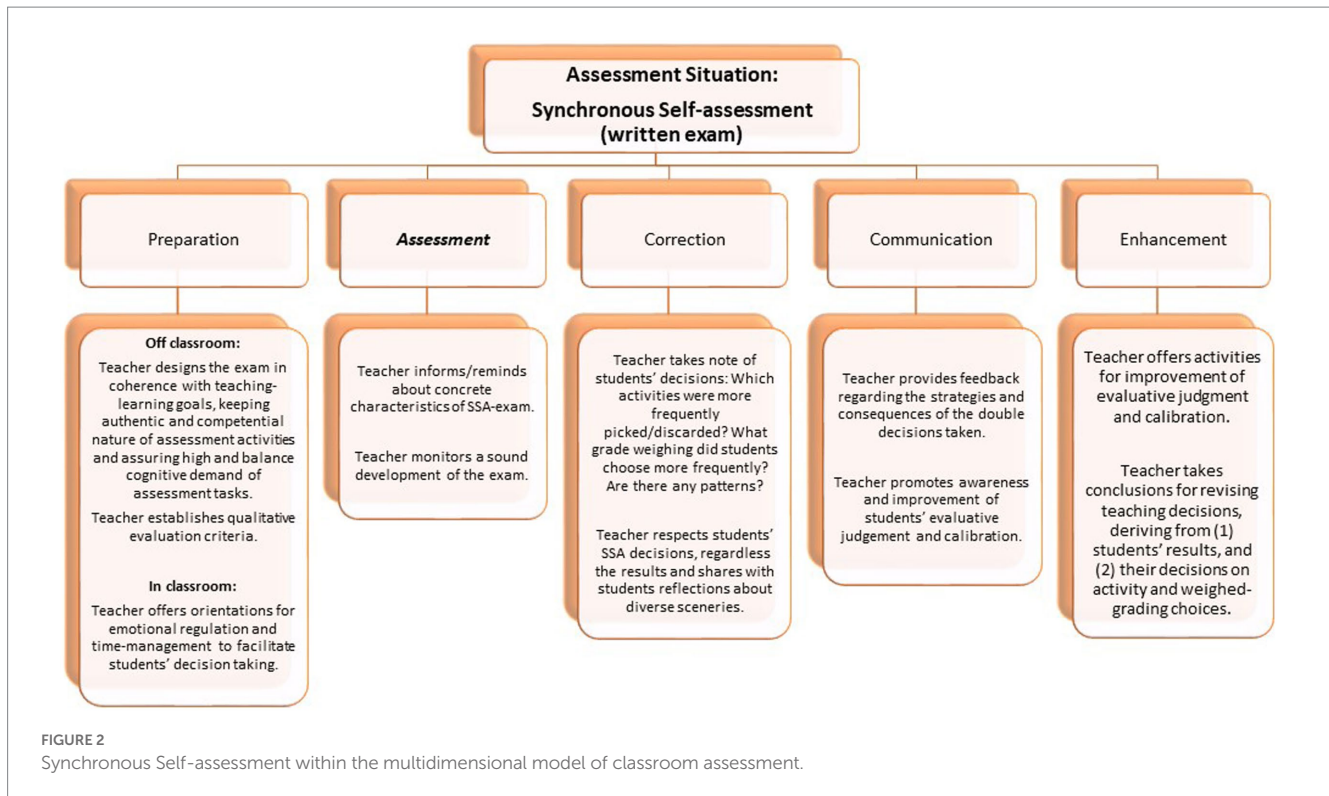
A written exam would be typically an instrument of the data collection phase; it might be part of a broader evaluation program, and it -quite naturally- would include elements of greater detail or lower level, such as assessment *activities* and *tasks*. Let us clarify the difference between these two latter elements. An *assessment activity* presents a global action request to the student, a statement or an utterance graphically identifiable, with a beginning and an ending, as separated from other action requests, clearly identifiable within the exam. At the same time, an assessment activity can contain from only one to a variable number of assessment tasks, which suppose specific and unitary cognitive demands for the student (Remesal, 2006; Remesal et al., 2022). For example, in a typical reading comprehension assessment activity the student receives a text with several associated questions. The text followed by the questions would make up the

assessment activity. However, each of the individual questions would constitute an *assessment task* that requires a detailed, independent response from the student. In turn, they offer unitary opportunities for good or bad performance. Sometimes several assessment tasks can be linked to each other so that the quality of a first response compromises subsequent responses. In any case, the assessment tasks will be contrasted one by one with the assessment and grading criteria.

1.3. Synchronous self-assessment within the MMCAP

How does synchronous self-assessment relate to the whole picture of the MMCAP? Figure 2 presents how SSA fits into this model. This strategy is best implemented during a written exam, that is, during the data collection phase. However, like any other assessment activity, it would permeate the remaining interactional segments. First, design decisions must be taken as to what form the exam should take (number of activities and tasks, their features, assessment criteria, post-hoc decisions).

The key to synchronous self-assessment lies, indeed, in the design of the exam. Throughout educational history, there is abundant literature against exams (e.g., Yu and Suen, 2005); however, we contend that there is still a high pedagogical potential in such artefacts, if analysed carefully. The exam itself is a *unitary pedagogical artefact* which demands complex design strategies, especially if aiming at assessment *for* and *as* learning (Yan and Boud, 2021). Recent experiences in diverse disciplinary areas like Law (Beca et al., 2019) and Computer Science (Rusak and Yan, 2021) bring evidence of higher education instructors' current worries concerning exam design. Following current approaches to assessment for learning and assessment as learning, an exam should present complex,



contextualised, realistic, argumentative activities and tasks. In other words, assessment activities must meet, as far as possible, the expectations of the so-called competency or authentic assessment (Villarroel et al., 2021). Authentic assessment requires complex assessment tasks. Ideally, the cognitive demand of these exams should be at least in a middle to high level.

Despite critical voices against 'exams' and in favour of other alternative assessment activities and instruments, we defend that the exam does not necessarily constrain the authenticity of each activity or task. Each exam has a *microstructure*, referring to the characteristics of each of the assessment activities and tasks included in it. Secondly, it has a *macrostructure*, referring to how these activities and tasks are distributed in time and space and how they are coherently related (or not) to each other. Moreover, finally, it has a set of *rules of interaction* that determine what each of the participants can or cannot do or is expected to do, during the development of the exam itself (for example, consulting sources or not, using the calculator, answering individually or in a group, giving spoken or written answers).

In the specific case of SSA, and more important, when the strategy is applied for the first time for students and teachers, we propose to design exams with two differentiated parts. The first part would be common and compulsory for all students to solve; the second part would offer elective activities for the students to carry out the SSA *per se*. Dividing the exam into these two parts follows a twofold objective:

- From the *student's* perspective, the common and compulsory part guarantees sufficient cognitive activation *before* undergoing SSA. Activities in this first part, hence, ideally should be designed to bring to the surface the learning needed to tackle the second part of the exam. So, in a certain way, the first common part constitutes a preparatory segment to the SSA realisation.

- From the *instructor's* side, the first common part establishes the minimum standards that all students should demonstrate, contributing both to the formative but also to assessment's accreditation and accountability purpose.

This novel strategy of synchronous self-assessment permeates all five segments of classroom interaction related to learning assessment. In Figure 2, we present the key aspects that researchers and practitioners should consider when implementing SSA: from preparation to pedagogical enhancement, if we intend to push SSA to the limit of its potential, all interactional segments ought to be considered.

We are currently engaged in a research plan for the medium-long term. Our first exploratory effort of this new strategy focused on the student. Some results have already been published concerning students' emotional experience and metacognitive engagement and management (Remesal et al., 2019; Remesal, 2021; Remesal et al., 2021). As these first studies demonstrate, SSA offers benefits in terms of a significant increase in confidence or sense of control before solving the exam. It also raises awareness and strategic resources management to increase performance. The time has arrived to look at the teaching figure. Some first results regarding the conceptions of assessment in connection with SSA have already been presented (Estrada, 2021): the positive adoption of SSA seems to be more likely amongst teachers with a richer formative conception of assessment in terms of Remesal's model of conceptions (Remesal, 2011; Brown and Remesal, 2017), that is, solid formative beliefs affecting all four dimensions (affection on teaching, on learning, on accreditation and accountability), whilst teachers with a summative conception or even just a weaker formative conception (affection of just two dimensions) are less prone to implement SSA to its whole extent.

1.4. Research goals

After exposing the conceptual basis of this novel assessment strategy, in this present study we pursue to deepen our knowledge of SSA's implications for the educational practice. Our concrete goals are:

1. To identify the characteristics of the exams designed by the participating instructors (one expert and three novels regarding SSA).
2. To explore the instructors' reflections associated with a SSA experience in their course.
3. To identify possible training needs for an adoption plan of the evaluation strategy of SSA, as well as possible lines of research that are open to us for improving assessment literacy.

2. Materials and methods

2.1. Participants

We carried out a quadruple case study (Yin, 2014) with four university instructors who implemented the SSA strategy in four groups of first-year university students at a Teacher Education Degree (two groups of Primary Education and two groups of Early Childhood Education). Table 1 presents the demographic characteristics of the four participants. Each instructor attended a group of 30 to 40 students. The selection of these four instructors responds to an intentional sampling strategy: (1) the four instructors are part of a larger teaching team and share the same course plan of Educational Psychology, (2) the four instructors also have extensive teaching experience in these grades and share the same basic formative assessment approach, at least intentionally, and (3) finally, one of the instructors could be considered an expert in SSA, whilst the other three are novels.

2.2. Data generation

This study collected data of diverse nature and origins: classroom natural artefacts and interviews. First, the participating instructors delivered their end-of-semester written exams, as they had personally designed these assessment artefacts. The instructors agreed on basic guidelines for designing and carrying out these exams:

- *Micro-structure*: the activities should aim to have particular features: they should preferably be complex activities with a high cognitive level, open argumentative resolution based on

professional actions or contexts, and as authentic as possible (Villarroel et al., 2021).

- *Macro-structure*: the exam would split into two parts: a first part, common and mandatory for all students, and a second part, with elective (SSA) activities. In the first part, students would have to solve between two and four activities. In this second part, students first would select three activities out of five offered and secondly, they would choose between a triple evaluation variant: *equitable* -all solved activities would weigh the same potential maximum value-, *maximal difference* -one solved activity would weigh for 50%, one for 30%, and the third one would account for 20% of the final grade, and *minimal difference* -two solved activities would weigh for 40%, and the last one for 20% of the final grade.
- *Interactional norms*: Students would be informed about the specific norms and innovative strategy just starting the exam. The exam would last 120 min, with parts 1 and 2 explicitly separated. Throughout the exam, students could consult doubts with the instructor.

After the students sat the exam and instructors had time to revise and grade students' responses, we conducted a semi-structured individual interview with each instructor, recorded on audio and transcribed. This interview sought to collect information about different evaluative decisions made by the instructors and their evaluation of the experience implementing the SSA strategy. The interview script was elaborated on the basis of the MMCAP.

2.3. Analysis procedure

Both authors participated equally in the analysis process; a third analyst's collaboration is acknowledged at the end of the text. To perform the analysis, we proceeded in three steps:

- First, for the analysis of the exams, the revised Bloom's taxonomy was our reference (Anderson and Krathwohl, 2001; Remesal, 2006; Villarroel et al., 2021). We started determining the quantity and internal relation of assessment activities and tasks (macrostructure). Secondly, we identified the level of cognitive demand of the assessment tasks (microstructure), considering three basic levels (*low* -to remember, to identify-, *medium* -to understand, to apply-and *high* -to evaluate, to create). Each of these levels received accordingly 1 (*low*), 2 (*medium*) or 3 (*high*) points as *cognitive demand value*.
- For the interviews, we conducted a qualitative analysis of the instructors' discourse (Braun and Clarke, 2006). The five phases, or interactional segments, of the MMCAP (Coll et al., 2012) were the point of departure of a recursive analysis procedure which advanced in a series of loops of individual analysis -by each of the three analysts-and later contrast for discussion of discrepancies until reaching a consensus. During this back-and-forth procedure of deductive and inductive analysis, emerging themes relative to the specificity of SSA were particularly in focus.
- In the final analysis step, we contrasted both data, artefacts and interview discourse, in order to identify points of coherence (or lack thereof) in each case.

TABLE 1 Demographic description of participants.

Instructor	Age (range)	Teaching experience at HE (range)	Experience with SSA
Instructor #1	Over 45y.	Over 20y.	No
Instructor #2	40–45y.	5–10y.	No
Instructor #3	Over 45y.	5–10y.	No
Instructor #4	40–45y.	15–20y.	Yes

3. Results

In order to offer a complex but also clear and panoramic picture of the results, we organise them in the following way: first, we will present the essential characteristics (micro-and macrostructure, and cognitive demand) of all the artefacts (four written exams as designed by the participating instructors); second, we will dedicate particular sections to each case addressing at once the results from the interviews and the contrasting discursive analysis of the artefacts; thirdly, we present results referring to the specificities of SSA and instructors' evaluation of this innovative experience.

3.1. Exam features: Macrostructure, microstructure and cognitive demand

Tables 2, 3 and Figures 3, 4 present the results corresponding to the exploratory, descriptive analysis of the designed exams. The cognitive demand of an assessment activity gives us clues about the performance expectations that an instructor places on students. In order to understand the results in their context, it is important to remember that all four participating instructors are members of a teaching team. As a team, they adopted a basic agreement that affected the design and development of the exam in all groups equally, as indicated in a previous section. However, as results show, differences can be seen between the analysed exams. The absolute quantity of assessment tasks that the instructors assigned to each exam part differs, oscillating in a range from 18 to 29, taking the exam as a whole, but ranging from 7 to 10 and between 11 and 20, when looking at first and second part separately (see Table 2):

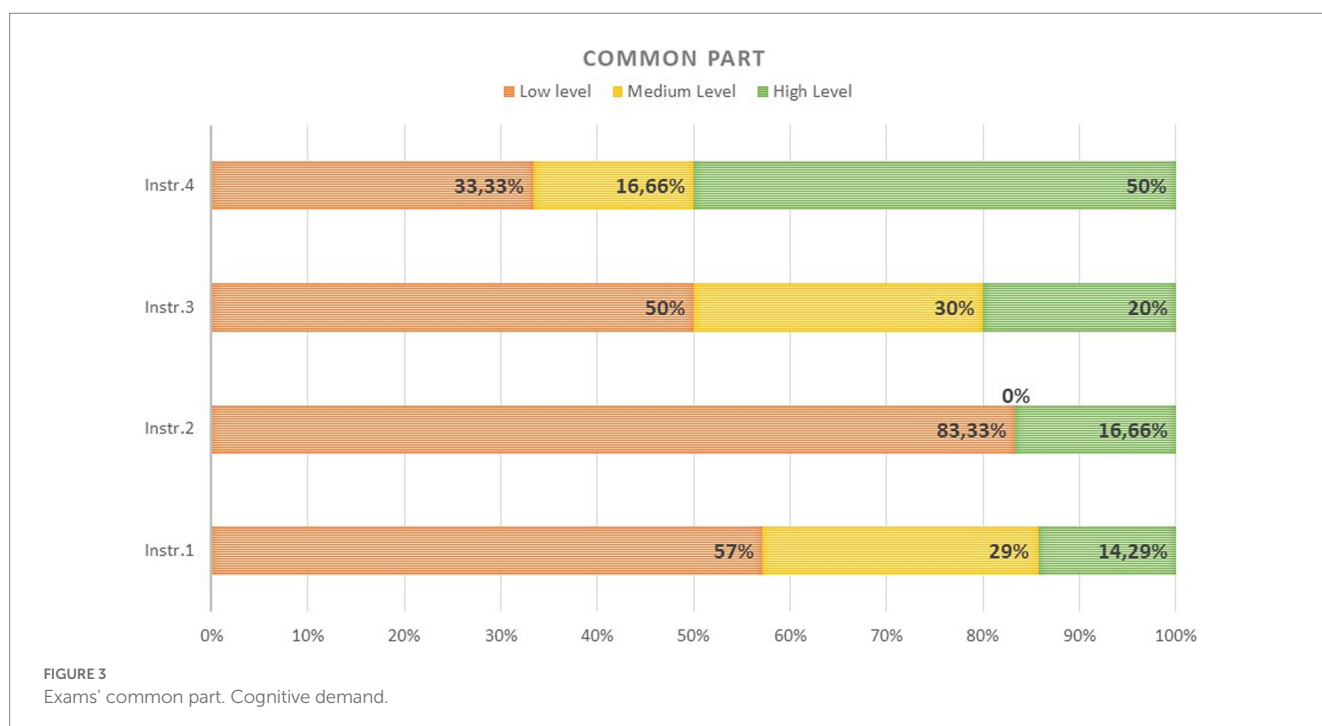
Table 3 completes the look into the four assessment artefacts designed by the participants, showing the average cognitive demand. Here we find two exams with (a) relatively low and (b) balanced cognitive demand between part 1 and part 2 of the exam (instructors

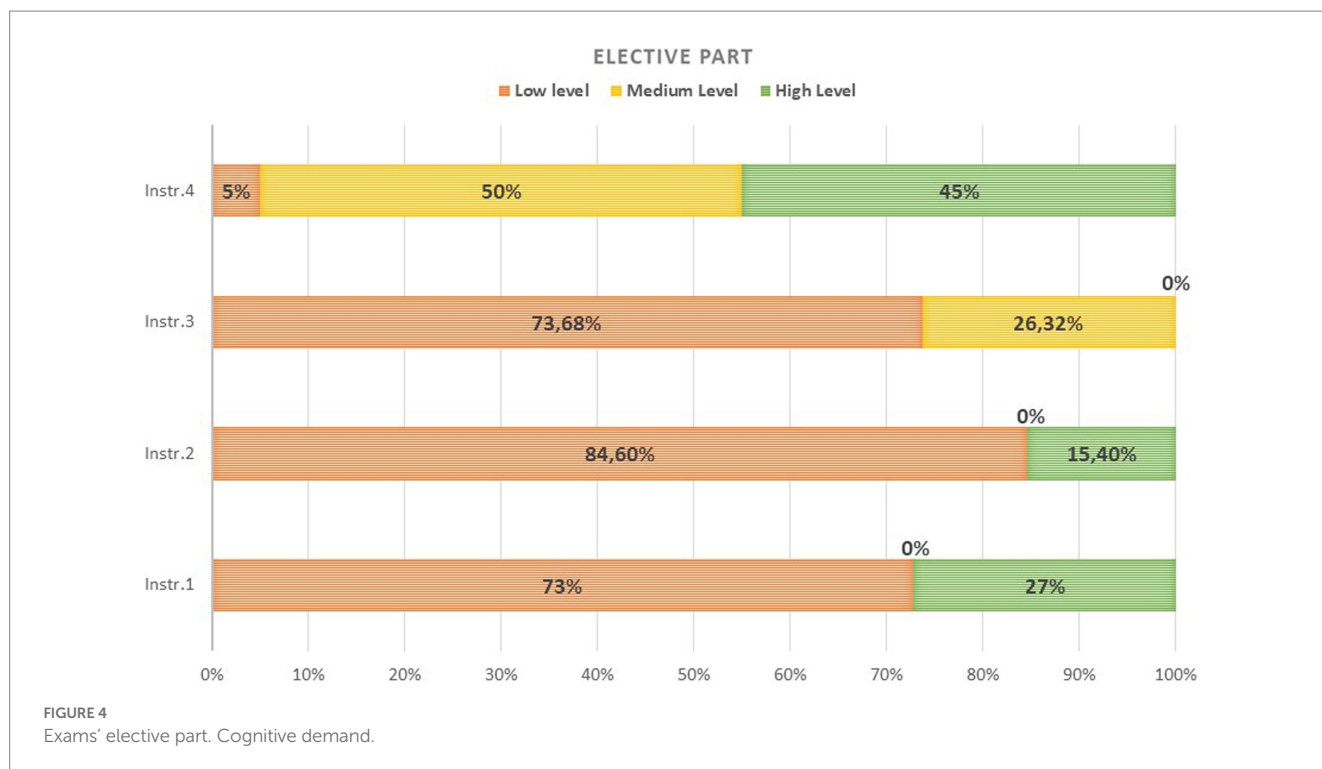
TABLE 2 Quantity of assessment tasks in exams and cognitive demand.

Cognitive demand	Exam designer			
	Instr.1	Instr.2	Instr.3	Instr.4
Activities part 1 / Activities part 2	2 / 5	3 / 5	5 / 4	2 / 5
Total of assessment tasks	18	19	29	26
Common section	7	6	10	6
Low (remember, identify)	4	5	5	2
Medium (apply, analyse)	2	0	3	1
High (create, evaluate)	1	1	2	3
Elective section	11	13	19	20
Low (remember, identify)	8	11	14	1
Medium (apply, analyse)	0	0	5	10
High (create, evaluate)	3	2	0	9

TABLE 3 Level of cognitive demand [1-low, 2-medium, 3-high]. Mean (standard deviation).

	Instr.1	Instr.2	Instr.3	Instr.4
Common section	1.57 (0.78)	1.33 (0.81)	1.70 (0.82)	2.16 (0.98)
Elective section	1.54 (0.93)	1.30 (0.75)	1.26 (0.45)	2.40 (0.59)





#1 and #2). The third artefact presented (a) low and (b) unbalanced cognitive demand with a higher demand in the first-common part (instructor #3), and the fourth artefact posed (a) medium to high cognitive demand and (b) unbalanced cognitive demand with a higher demand at the SSA-part (instructor #4).

Finally, Figures 3, 4 show these same descriptive results in percentual terms for a richer comparison. Instructor #4, who had previous experience implementing SSA, presented an exam with much higher cognitive demand to her students, in contrast with the other three instructors who implemented the innovative strategy for the first time.

3.2. Discursive analysis: Looking into details

Identifying cognitive demand is just a first look at the designed exams. The discursive analysis of the assessment activities in the exams as they were presented to the students allows us to deepen our understanding of these evaluative artefacts, shared in the following subsections. We will first report about the three novel instructors, followed by the expert one. For each case, we report results of the analysis of the assessment artefacts (exams), first and second part, and provide some excerpts from the interviews to either support to or contrast against the exams. The enunciates of the assessment activities extracted from the exams are framed to distinguish them from interview excerpts.

3.2.1. What do we learn from case #1?

As we have already seen, more than half of the tasks of the first part of exam #1 present a low-level cognitive demand ($M=1.57$; $SD=0.78$). It is mainly in the third activity when Instructor #1's caution in this first experience with SSA is most evident. This activity would initially set a high cognitive level (assess and propose improvement). Nevertheless, the cognitive demand is curtailed by the secondary instruction, exposing in advance the key action to

be carried out (*italics added*), thus reducing the task to a mere follow-up of direct orders:

Identify the errors in the following conceptual map and correct them-introducing the information missing in the relationship between concepts and modifying the one that is erroneous by the information appropriate to the same map.

In the second, elective part of the exam, all five activities (out of which the students choose three to solve) revolve around a single case narrative that brings its own contextual boundaries. Each of these activities presents a disparate level of cognitive demand ($M=1.54$; $SD=0.93$). The critical concepts sought in each activity are marked in bold by the instructor herself as a cognitive aid. A careful reading reveals that four of the five activities do not use the case as an actual trigger for a creative and genuine response but rather as an excuse or a frame for simple identification-remembrance. For example (**bold in original**):

Define meaning in learning and identify examples of three conditions for meaning attribution in the text.

Unlike the first four activities, the last one does not anticipate response hints. On the contrary, it challenges students to construct their scheme for constructing their responses. The question best collects high-level cognitive skills such as analysing, comparing, and evaluating. However, its cognitive demand is very disproportionate compared to the previous four activities. It is, altogether, a very unbalanced exam, and we can expect that very few students would choose this fifth activity (**bold in the original**):

Comment on the text to explain, in your opinion, the most relevant part of Ana's learning case using the contents of the three thematic blocks worked through the course.

Comparing part 1 and part 2, we find a balanced exam, with no difference in cognitive demand ($d=0.03$). In the interview, Instructor #1 acknowledges having designed her exam with a strong accounting purpose since she expects all students to be able to demonstrate a minimum, declarative and defining standard of knowledge in the first part common to all students. In contrast, the second part, where the SSA conditions are applied, would be the space to demonstrate knowledge in use.

“This criterion guided me as a basis, that is, I tried to ensure in the first part some conceptual knowledge that seemed relevant to me, and in the second part, well, the use of conceptual knowledge in one case, (...) how to put this knowledge into practice.” (Instr.#1)

She also adds the following idea of minimal account giving as a leading thread for the exam design:

“Very nuclear [knowledge] in terms of how we had worked on it and in terms of the importance we gave to these ideas at work and also very nuclear in the sense that if the students do not know how to answer this, well... then we have a problem” (Instr.1)

3.2.2. What do we learn from case #2?

The first part of the exam designed by Instructor #2 presents mostly tasks of low cognitive demand ($M=1.33$; $SD=0.81$) but also includes one task of high cognitive level, exemplification and argumentation (italics added):

Define self-regulation. *Why is it important teaching it at early childhood education? Give a justified example of how you would work towards self-regulation in your classroom.*

However, the second part of the exam, with SSA, presents two types of activities with a notable difference between them with regards to cognitive demand ($M=1.30$; $SD=0.75$). The first two activities refer to a single concept of the programme and ask to analyse and explain equally. These are activities that we consider to be parallel. Nevertheless, the third and fourth activities only request the identification of various concepts presented in the case narrative. The fifth activity, suddenly, raises cognitive demand (bold in original):

Analyse and explain how the three conditions of **meaningful learning** are presented in Carmen's case to learn Anatomy.

Analyse and explain how the three conditions of the **attribution of meaning** are presented in Carmen's case to learn Anatomy.

What **goals and motivational orientation** does Carmen present in her medical studies?

What kind of **learning approach, goals and motivational orientation** does Carlos suggest to Carmen?

Evaluate the **distance** between Carmen's prior knowledge and the anatomy instructor's teaching and explain the **effects** of working at that distance on the student's motivation.

The internal contrast of both exam parts presents no difference, hence it is a balanced exam ($d=0.03$) with regards to the cognitive demand. Instructor #2 states in the interview his goal that the

assessment activities of the elective SSA part maintained a consistent level of difficulty. In this way, he underlines that the student's choice of activities be based on the mastery of thematic knowledge and not on possible unbalanced diverse demands. Hence, after the analysis of the five activities, we can raise our doubts about the accomplishment of this instructor's goal regarding the balanced cognitive demand. Here too, the accountability function of the exam can be appreciated:

“I tried to ensure that they were balanced in the different issues and that they were, that they were somewhat equitable, (...) that there was not a very simple or a very complicated one (...) and I did the most complex part based on the case, if you want to excel or have a good grade, you have to go a little further”. (Instr.2)

3.2.3. What do we learn from case #3?

The exam designed by Instructor #3 presents in overall a higher cognitive demand than the two previous ones, particularly in the first part. In this third case we also find an increased amount of assessment tasks embedded in the activities, both in part 1 and part 2. An additional difference with respect to the other two novel instructors lays in the lack of balance between the common and the elective part of the exam concerning cognitive demand. Instructor #3 designs an exam with a significantly greater cognitive demand ($d=0.66$) in the common part ($M=1.70$; $SD=0.82$) and less cognitive demand in the SSA part ($M=1.26$; $SD=0.45$) to be solved by students. Additionally, she inverts the order of presentation, so that students have to respond first to the elective section (easier, according to her intended design) and in second place they respond to the common section (more difficult, from her point of view). In the elective part, she presents the students, for example, two sets of parallel activities that only differ by the alluded learning content. All of these activities are of low cognitive demand (reminder or presentation of a conceptual network) and are complementary to each other in terms of the content evaluated:

Explain the differences between collaborative and cooperative work. Briefly describe the three interpsychological mechanisms involved in the construction of knowledge amongst peers.

Explain the differences between collaborative and cooperative work. Briefly describe the three dimensions of analysis of collaboration for learning.

Considering the two mechanisms of educational influence that operate in classroom interactivity situations, explain the construction process of shared meanings.

Considering the two mechanisms of educational influence that operate in situations of classroom interactivity, explain the process of progressive transfer of control.

In the interview with Instructor #3 we learn about another crucial difference between her and her colleagues. Regarding the macrostructure, for instructors #1 and #2 the first part of the exam supposes the verification of the lowest common denominator of knowledge, hence, the basic learning goals, and the optional part entails a higher performance expectation. In contrast, for Instructor #3, the opposite happens: the optional part includes basic activities for the instructor and the first part -for compulsory response-supposes the opportunity for individual excellence. This inversion introduced by Instructor #3 also affects the rules of interaction of the exam since her students are exposed to SSA conditions when starting the exam time, unlike all the other groups. This change of rules are evident in the following excerpt:

“60% were the questions that they could choose, so they had five questions, which is what we agreed on in the teaching team, and I tried to ensure that they were balanced in the different topics (...), that there was not one that was very simple or one that was very complicated, but that they were all at a medium level, of difficulty, this is the part that they had to choose. And the part that they couldn't choose was a case, there are three questions about the case, which had a higher degree of difficulty. The truth is that I thought, 'well, here [in the first part], they have the advantage that they can choose' and well, it would be the most affordable part of the exam and the most complex part I did base on the case" (Instr.3)

3.2.4. What do we learn from case #4?

Instructor #4, with at least five-year experience in implementing SSA, designs the exam with the highest cognitive demand, comparing all four cases. Both, part 1-common part- and part 2-SSA-have over medium values (part 1, $M=2.16$; $SD=0.98$; part 2, $M=2.40$; $SD=0.59$). There is a small difference between both parts in their cognitive demand ($d=0.29$), prioritising the elective part, thus emphasising students' agency. We present an example of one activity of each of the exam parts. In the first one, a direct question requires the student to elaborate a new argument, incorporating some 'safety net' questions (Van Den Heuvel-Panhuizen, 2005) to prevent the student's mental blockage in a single option:

Is it possible to meaningfully learn something eventually wrong? If yes, what could be the consequences? If not, what do we conclude? Justify your answer and provide an example.

In the second example, from the SSA part, students must confront and combine knowledge and skills from different subject areas, such as psychology and mathematics, evaluating and appreciating the feasibility of four given options:

Look at the following images. The abscissa axis (X) represents time in all graphs, and the ordinate axis (Y) represents knowledge. If the learning process could be represented in such a simple way (something in itself impossible, as we know, because too many variables are involved), which of the following graphs would best represent a meaningful learning process? Why?

GRÁFICO 1

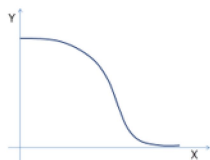


GRÁFICO 2

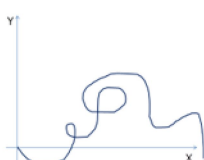


GRÁFICO 3

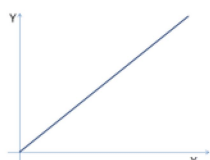
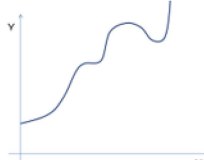


GRÁFICO 4



Instructor #4 declares in the interview that her exam aims to offer students the challenge of solving tasks of a complex nature. In both parts of the exam, assessment tasks present medium or high cognitive demand, since the activities require analysis, argumentation, and/or creative exemplification. The prompts do not request direct declarative memory responses or explicit definitions. Instead, they expect students to apply conceptual knowledge of those definitions or 'assemble' argumentative responses. This fourth exam is consistent with Instructor #4's discourse:

“In the first part, where everyone must demonstrate basic learning, I try to offer questions that go to the core of the concepts and that help the students reflect on that core [of concepts]. And in the synchronous strategy part, I try to propose them activities that are also of a competence challenge, more applied and creative”. (Instr.4)

3.3. Instructors' evaluation of the experience: Assessment literacy needs detected

One of the keys to the proper development of the SSA strategy lies in the design of the exam. As an evaluative artefact, it presents a specific microstructure, macrostructure and interactional rules that manage the student and instructor's agency space. The analysis of the presented exams, as designed by the instructors individually, show a great difference in terms of cognitive demand to the student. Comparing Instructor #4 with her colleagues, all of them novels in SSA, we find significant differences in both exam sections. Table 4 shows the varying size effects of these differences.

Through the discursive analysis of the interviews, we identified various emerging themes revealing some inconsistencies, which point to likely training needs to improve assessment literacy and assessment practices. We present below some extracts around the following aspects: emotional reaction to the experience, understanding of self-assessment and, in particular, the SSA, effects on the correction phase and the interpretation phase and use of learning outcomes for subsequent teaching decisions.

3.3.1. Emotional reactions to the first experience with SSA

Three of the instructors were novel in the implementation of the SSA strategy. In their interview, all of them manifested a positive acceptance of the challenge of this project, although they also added objections of two kinds, first of all, emotional objections. In the instructors' responses, we find concerns dealing with their own emotional response as instructors and with an empathic identification with the evaluated student.

TABLE 4 Comparing cognitive demand of the four exams (Cohen's d).

	Part of the exam	
	Common	Elective
Instr.4 versus Instr.1	0.66	1.10
Instr.4 versus Instr.2	0.92	1.63
Instr.4 versus Instr.3	0.51	2.17

Instructor #1, for example, expressed concerns about increased anxiety on the student's part:

"[I really wanted] to do it, to see what would happen, because the idea of choice specifically, that they could reflect on what score they would give each of their answers... I found it interesting, wanted to do it to see what would happen" (...) "In relation to the format, in fact, that is, deep down my concern as a teacher myself was originally that it would not affect them, aah that they would not be overwhelmed by time [pressure], by the fact of having to take their decisions". (Instr.1)

Instructor #2, on the other hand, predicted positive reactions on the students' part, since the increase in a personal agency in decision-making would also favour them:

"I thought it was a good chance to promote sense and meaning itself, in choosing the activities. Something that has always interested me; I had always thought of, the fact that the students could take some decision or choice on the activities, but I had never considered about the grading of the questions, I thought that this was more on the teacher, right? I think this was the most innovative part." (Instr.2)

Instructor #3, finally, kept a double perspective, empathic towards the student but also expressing some concern from her teaching position, anticipating an increase in the workload associated with the grading, from her point of view:

"First, I thought that the students would welcome it as something positive, a good idea for the students that would be accepted with pleasure, and that it would benefit them. And then I thought of the teaching part that I might have difficulties when correcting, because each activity would have a different weigh depending on the student's choice, which complicated the issue of evaluation for me, indeed". (Instr.3)

The emotional concern anticipated by Instructor #1 was not confirmed. Students had a positive reception in all cases, as shown in previous publications (Remesal et al., 2019, 2021):

"They were calm and they answered and they had enough time to do it, so there was not this effect that I was worried about ... at all." (Instr.1)

"(...) the students are super happy, they even asked for the next exam to have the same [form], they see it as an advantage." (Instr.3)

"In general, students are often surprised by the new rules of the exam, but they receive it well, (...) I think that above all they are happy when they hear the rules of the exam. In my experience, they've been always positive." (Instr.4)

3.3.2. SSA as a particular modality of student self-assessment

A central theme in this exploratory study is the instructors' conception of student self-assessment to understand to what extent

SSA can be perceived and appreciated as beneficial. In this sense, we highlight the reflection of Instructor #1, which reveals a conception of retrospective, let us say, traditional self-assessment. Students' self-assessment most frequently occurs as a separate metacognitive behaviour, after and outside the learning process and the demonstration/performance of learning. This conception of self-assessment of learning as an ex-post-facto action is widespread in the literature of the field, as well as in educational practice (Andrade, 2019):

"A process where the student understands what her learning process has been and how she also evaluates the results of this process, so that's it. In the process, both the work that [the student] has done is included, as well as reflecting on other elements that may affect other aspects of the educational process, it is also reflecting on like "well, I have done this, it's ok for me" or "I might need some help in this new topic"; but well, above all this, it's how the student evaluates his process and the results he has obtained". (Instr.1)

In contrast, for Instructor #2, self-assessment implies a continuous metacognitive demand. Quite a challenge if beginning students have not yet established the habit:

"for me it consists of a process of reflection on how the [learning] process has gone... and how the process is going, right? that has a great impact on the students throughout the process, on how they are learning, what difficulties they are facing, what progress they are having and well, I try to make them reflect during the process... on how they're progressing... that, on their learning". (Instr.2)

Instructor #3 refers to self-assessment as a process of shared dialogue between instructor and student (Sutton, 2009):

"Self-assessment has to be a joint exercise, the student must have all the information to be able to evaluate himself, but then there has to be --not a grade from the instructor, but a conversation, a dialogue, an exchange". (Instr.3)

Finally, for Instructor #4, the only instructor with previous experience in SSA in this study, the particular contribution would be in the multidimensionality of the self-assessment processes that underlie the moment of the exam since the students face the required decisions:

"Through this strategy, the student develops more meta-knowledge, more awareness of what they really know or do not know, or to what degree they know it... also more awareness of their cognitive and emotional resources at the time of facing the evaluation, or like knowing or recognising their nervousness, know how to manage it, know how to manage time, know how to organise themselves" (Instr.4)

Instructors new to the SSA strategy, however, were cautious in assessing the potential benefits regarding the assessment of actual complex competencies:

"I imagine that at the time of the exam they do a quick assessment of what they know and what they do not know, so what I meant is

self-assessment as a process or self-assessment as an evaluation of the results of whether this grade seems good or not good to me, or whether it is appropriate or not. Another aspect that would precisely have to do more with the learning results, not with the qualification, yeah, I kind of think so, that at that moment that helps them think "what do I know better" (...) so I think so, that, in this sense, I do think that it contributes a little more to the assessment of competencies because it precisely contributes, I understand, to this nuance of self-evaluation of what one quickly thinks what one knows and what not, and with what, what he knows and what he doesn't and with what he feels/ and-or I don't know, I don't know, with what he feels safer when answering". (Instr.1)

"I think it goes a little step further because it asks them to make a choice right in the middle of an exam... that was not expected, let's say, eh, and it goes a little step [further]". (Instr.2)

"We would be working on the competence of personal autonomy and learning to learn, so it would be a competence development (...) that they have to develop and that, normally, because these competences are not worked through or evaluated, but with this (SSA) it allows them to reflect on their learning process, on their knowledge and it would be closely related to the concept of self-assessment before, although it would not be a self-assessment with a numerical grade but rather a self-assessment of their own knowledge". (Instr.3)

3.3.3. SSA and the multidimensional model of classroom assessment practices

As we assume, SSA would indispensably affect all five evaluative segments of the interactional process of classroom assessment established by the MMCAP (Coll et al., 2012). However, according to the interviews of the three instructors who implemented it for the first time, the strategy only had a noticeable impact on two of these segments: correcting or marking the students' answers and making sense of the choices to take consequent decisions.

3.3.3.1. Impact on the marking phase

All three novices in SSA seemed to apply assessment criteria based on dichotomous absolutes (correct-incorrect answer), and they gave evidence of little reflection on the possible impact of the SSA:

"The truth is that it wasn't difficult, I mean when I corrected, I saw what was right and what was wrong and then depending on the student's choice of their score, I adjusted the score, that's it, easy, it does not really add any complexity this [SSA] do you know what? I marked their answers as usual". (Instr.1)

"The marking of each answer is based on a series of verifiers that I prepare beforehand to make the correction, so that it be a fair correction for everybody... for all the answers. So, in that sense, it has not changed, well I had to think of more indicators because there were two more activities". (Instr.2)

"I organised myself, the rubric and stuff, I was able to organise myself relatively well, yes, it is a little more challenging, because you have to be doing more calculations, but it was OK in the end". (Instr.3)

Compared to the three novices in SSA, we have Instructor #4's reflections on her more detailed correction procedure, based on more qualitative aspects, to which she adds ethical-moral factors:

"I make like a scheme of the correction criteria and what I do is collecting the concepts I hope the students introduce in their arguments and what kind of examples I am going to consider valid or non-valid. So, starting from there, I always value the example equally or even more than any conceptual definition, the definition of the concept is always implicit to me, so I always give the example as knowledge in use a more weighted score. (...) later with the SSA the student says "I want this activity to be worth 3 points, 2 or 1", so I apply that factor to what I have corrected, but it does not affect it in any other way (...) This is very important, precisely, in the sense that the instructor should not be affected by the students' choices on weighing scheme. If I was affected, I could be including another type of evaluation that could end up being unfair to the student". (Instr.4)

3.3.3.2. Impact on consecutive decision taking: potential for enhancement

SSA puts on the instructor the need to make pedagogical sense of the students' new choices, sometimes unexpected ones. Previously published results on the students' experience tell us many different reasons hidden behind students' decisions (Remesal, 2020). Some of those reasons are related to the students' conceptions of learning assessment and their traditionally passive role in the process, which might hinder them from assuming a more active role. Other reasons have to do with specific circumstances during the exam development (e.g., emotional or cognitive blockades). Other reasons are related to personal preferences or interests or to the self-awareness of personal skills. In this regard, we have also identified notable differences between novice instructors (#1, #2 and #3) and Instructor #4, with previous experience in SSA. To begin with, in the interviews, novice instructors shared representations of students as mainly passive subjects, lacking agency and strategic or reflective decision-making capacity, or they attribute decisions to non-controlled temperamental factors. For example:

"It seemed to me, at first, I thought "man, this is a bit weird", but I really suppose that they distributed a bit... randomly, I don't know (...) those who choose maximal difference... I also had the feeling that they had chosen a bit by chance, a bit miscalculating". (Instr.1)

"There were students who strategically decided to assign the greater points to the answer they thought was best formulated, and then the other two types of choices, minimal difference and equitable, were like sort of balanced... and in these cases some students had kind of more doubts when they didn't know what they had answered correctly or wrong, they strategically decided to distribute the scores equally, right?... the student who has a more strategic vision of his performance can choose the most daring options, let's say the maximal difference for example... but

then those who are more insecure or hesitant, right? They have chosen more options... of more equal distribution or either minimal difference". (Instr.2)

"(5 seconds silence) I think that students' choice may depend on various factors, it depends on their knowledge, their learning processes and then there is another factor which is personality, there are people who are much more risk-tempted and people who are less risk-bound, so to say, so I get the feeling that students who are more conservative, who are not as risky, take equal distribution. But then, there are students who are more risk-bound, I guess, and then they pick maximal difference, and then the last option, minimal difference, that one has been taken by fewer students, as far as I remember, because this one does involve a process of reflection". (Instr.3)

Instructor #4, in contrast, stated:

"Two or three students choosing an equitable grades distribution, for example, may have very different motives behind their choice. So, for me as an instructor, the fact of learning about that diversity of motivations behind the same decision also helps me learn many things from the students, about the complexity of the learning process (...) and also helps to understand misunderstandings". (Instr.4)

The positive reflection of Instructor #4 on SSA generates more introspective and self-critical reflective processes of her teaching or management of the assessment instruments to capture students' knowledge and learning process. Even assessment activities and tasks become an object of her reflection in such a way that assessment becomes truly formative, leading to regulate not only learning processes but also teaching processes (Yan and Boud, 2021):

"It adds this layer of complexity to the entire process of reflection on learning, what I still have to improve is the activities themselves, because I reckon that sometimes I'm perhaps not very explicit when it comes to proposing activities (...) So this SSA strategy forces students to read the entire exam first, take a panoramic perspective, reflect on each activity, (...) so that panoramic reading implies a different activation of knowledge. It is not sequential, (...) but it is a horizontal and panoramic reading (...) having to decide forces you to having to read all [the activities] in a row, take a step back, evaluate the situation, put all your knowledge in active working memory and strategically see better what you can solve, also always with the aim of giving your best performance. I believe that this is also a matter of justice or fairness, right? To put the student before his best possible self, well give him the option to show his best selfie, sort of, not the students portrait the instructor is looking for, from his definition of the goals and assessment criteria in the course, what we decide would be a good student, but the best possible self-portrait by the student himself". (Instr.4)

4. Discussion of results

In this article, we present the results of a first approximation to SSA from the teachers' perspective. As an innovative proposal, it is

challenging to discuss the results in a regular way due to the absence of previous specific literature to contrast with. Therefore, we propose a series of reflections on new questions and challenges that arise for teachers and, consequently, for teacher educators as a result of implementing SSA as a means for assessment for learning.

To begin with, we have been able to detect areas in severe need of improvement regarding assessment literacy: exam design (with and without SSA) is far from satisfying and incoherent with the assessment for learning discourse (Hernández Nodarse, 2007; Hernández Nodarse et al., 2018). Crucial questions arise; for example, what is the link between the experience in implementing SSA and the level of cognitive demand of the proposed exam designed by the instructors? Does Instructor #4 pose a more demanding exam due to the fact (or as a consequence) of having more experience in the implementation of SSA? Or are they independent phenomena? Or perhaps linked to other factors?

The emotional impact of assessment on the teacher has not yet been sufficiently studied. Brown et al.' chapter (2018) is one of the few publications, if not the only one by date, on this important issue. Much work has been done on the side of students' emotions related to assessment (Schutz and Pekrun, 2007), but previous works on the teacher's side (e.g., Schutz and Zembylas, 2009; Sutton et al., 2009), do not specifically consider *assessment* amongst the emotion-loaded phenomena in the teaching profession. One of the latest advances in instruments design for the study of teacher emotions, still does not pay attention to this particular chapter of the teaching profession (Hong et al., 2016). We wonder, thus, if the emotional concerns exposed by these three novel teachers are related to a lack of confidence in students' capability of assuming more agency and responsibility in the assessment of their learning. Or else, their emotional concerns could relate to the loss of (teacher's) power provoked by the SSA strategy.

In all three cases of novices in SSA to a greater or lesser extent, we identified notable inconsistencies between their discourse and the designed exam, as well as between the basic agreements of norms and macro-and microstructure, on the one hand, and the final design and development, on the other. Despite all the participating instructors being members of the same teaching team, with a shared pedagogical approach and teaching programme (Fulton and Britton, 2011), the evaluative artefacts (exams in this case) designed by each of these instructors differ notably. The contrasted exams share nothing more than the superficial structure and the conceptual contents object of evaluation, which are part of the course programme. It becomes evident once again that educational assessment needs continuous reflection and constant and real teamwork; otherwise, its validity and even the fairness of treatment received by the students are firmly at stake (Buckley-Walker and Lipscombe, 2021).

The results also highlight the need for training for university instructors, even those with long teaching experience, regarding the ability to identify the cognitive demand of assessment activities and tasks. For many years, the training towards assessment literacy has focused on what we could call 'superior' or 'meso/macro' assessment levels (Popham, 2009), such as using alternative and complementary instruments (for example, the implementation of rubrics). However, at a 'lower' or 'micro' level, deep reflection on the very nature of assessment activities and tasks is still pending (Bonner, 2017). We will only be able to respond to the call of Villarroel et al. (2021) to implement competency and authentic assessment if we are concretely aware

of the shortcomings of our current praxis. We contend that the very participation in this case study led the four instructors, and particularly those novices to SSA, to reflect on their conceptions and practices, however, it was only a first step.

Our results also show that instructors' conceptions about 'ordinary' or 'traditional' students' self-assessment affect the chances of acceptance and effective implementation of SSA. Undoubtedly, much remains to be explored from the view of the most general conceptions about assessment, be they linked to accountability or pedagogical regulation (Remesal, 2011; Estrada, 2021). It might be attractive to link SSA to summative or accrediting purposes since it occurs in a traditional summative practice -such as a written exam. However, we know that both assessment functions, accreditation and pedagogical or formative regulation, are always related and in constant tension in any educational system (Taras, 2009; Remesal, 2011; Black and Wiliam, 2018). Therefore, each assessment activity contains the potential to positively inform both summative and formative decisions that affect students' learning (Lau, 2016). We want to underline that our exam definition does not find a limit in paper-and-pencil format or an individual resolution (Remesal et al., 2022). In our understanding, the key and the specificity of the 'exam' lies in the explicitness of the shared purpose (for both teacher and student) to expose an optimal performance under a time-constrained condition. This broad definition of exam actually accepts a great variety of forms in the classroom. It would be worth exploring alternative conditions for implementing SSA, such as spoken or enactive resolution, or as a group experience. In short, it would be convenient to explore the potential of exam situations from an essentially formative point of view.

In terms of practical implications of SSA, we see two additional advantages in SSA in comparison with more traditional evaluation practices, which we also submit for consideration. In the first place, the correction or marking phase is usually the most tedious for the instructor in traditional conditions since it is repetitive when all the students must solve the same activities in a traditional "one-size-fits-all" exam. Under SSA conditions, however, this marking phase, due to students' particular choices, becomes more diverse and colourful and thus less monotonous, facilitating a higher level of attention and constant reflection on the teachers' side. With SSA, each exam can be unique due to the double decision of each student. The information that the instructor may collect is more complex and qualitative, both about the individual student and the group class as a whole. In turn, the specific students' choice of assessment activities provides rich information about the activities *per se* (their intelligibility, the level of challenge they pose, the interest they raise, and the understanding of the referred contents). Secondly, students' decisions about the weighting of answers can give clues about their self-competence or perceived difficulties during the exam. Every instructor faces the challenge of interpreting this information, so the assessment situation is also an opportunity to make decisions that effectively improve teaching. In an ideal case of an essential trust established between instructor and students, the students themselves could share the reasons for their decisions with the instructor, which further enriches all the formative potential of SSA.

As a matter of fact, our proposal of SSA is closely related to other recent concepts, such as evaluative judgement. (Boud et al., 2018; Tai et al., 2018; Panadero et al., 2019). We contend that SSA is indeed a practical proposal for developing evaluative judgement in the long

run, since it opens up occasions for reflection on personal excellence criteria. When using SSA to actively develop students' evaluative judgement, it can be introduced gradually, for example, one type of choice at a time (activity or weighted grading). This same gradual strategy could be applied in teachers' professional development for enhancing their assessment literacy, offering the chance to reflect upon changes more deeply.

5. Conclusions and open roads

The quadruple case study we have presented has obvious limitations: too many factors remain inevitably unattended, and specific context features might not apply to other situations. However, the results are robust enough to deserve our attention and raise essential questions for future research and praxis revision. From the multidimensional model of classroom assessment practices (MMCAP), that allows us to examine classroom assessment in a more comprehensive and qualitative way, we propose the following list of challenges for future research and practice improvement: How does SSA contribute to the development of evaluative judgement and students' self-efficacy? How can the instructor intervene in this direction? Regarding pedagogical potential, when is it best to inform students of the particular conditions of an SSA experience? How may this affect students' study strategies and learning approaches? How to design challenging exam situations that suppose real new learning opportunities? How can instructors be helped to manage the power rebalance with the student? How to promote a fair evaluation for all students? How can students' agency be raised? How can a teacher take pedagogical advantage of students' choices -both at an individual and group level? How is the instructor affected emotionally in this process? How could an instructor interpret the results of SSA globally as well as individually? What formative decisions can SSA promote? How can this strategy cater for the diversity of students? How can it contribute to educational excellence? Is SSA a suitable strategy in response to the new claim of personalised educational practices?

Finally, this study has been carried out at the university level, but we must also regard implementing this assessment strategy at earlier educational levels. In this case, we could ponder the appropriate adaptations depending on the students' developmental stage, or the curricular area. Many questions remain open; from these pages, we urge the educational community to accept the challenge of SSA. This first small case study is a tiny step into a new long road.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and

institutional requirements. The patients/participants provided their written informed consent to participate in this study.

Author contributions

AR carried out the data collection prior to the incorporation of FE to the research team. Once FE started the research collaboration, her contribution to the analysis and elaboration of the text is remarkable. AR and FE are equally in charge of the final result. All authors contributed to the article and approved the submitted version.

Acknowledgments

The authors want to acknowledge and thank the contribution of C. L. Corrial (Pontificia Universidad Católica de Valparaíso, Chile) to

References

- Anderson, L. W., and Krathwohl, D. R. (Eds.) (2001). *A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives*. New York: NY: White Plains Longman.
- Andrade, H. L. (2019). A critical review of research on student self-assessment. *Front. Educ.* 4:87. doi: 10.3389/feduc.2019.00087
- Beca, J. P., Castillo, E., Cid, A., Darrichon, E., and Lagos, S. (2019). Diseño de un examen de grado por competencias en la carrera de Derecho. *Rev. Pedagog. Univ. y Didáct. del Derecho* 6, 101–130. doi: 10.5354/0719-5885.2019.53747
- Black, P., and Wiliam, D. (2018). Classroom assessment and pedagogy. *Assess. Educ. Princ. Policy Pract.* 25, 551–575. doi: 10.1080/0969594X.20181441807
- Bonner, S. (2017). "Using learning theory and validity theory to improve classroom assessment research, design, and implementation" in *Paper presented at the Annual Meeting of the National Council on Measurement in Education*. New York.
- Boud, D., Ajjawi, R., Dawson, P., and Tai, J. (2018). *Developing Evaluative Judgement in Higher Education: Assessment for Knowing and Producing Quality Work*. Routledge, England.
- Braun, V., and Clarke, V. (2006). Using thematic analysis in psychology. *Qual. Res. Psychol.* 3, 77–101. doi: 10.1191/1478088706qp063oa
- Brown, G. T., Gebriel, A., Michaelides, M. P., and Remesal, A. (2018). "Assessment as an emotional practice: emotional challenges faced by L2 instructors within assessment" in *Emotions in Second Language Teaching*, ed. J. D. Martínez Agudo (Cham: Springer), 205–222.
- Brown, G. T., and Remesal, A. (2017). Teachers' conceptions of assessment: comparing two inventories with Ecuadorian teachers. *Stud. Educ. Eval.* 55, 68–74. doi: 10.1016/j.stueduc.2017.07.003
- Buckley-Walker, K., and Lipscombe, K. (2021). Validity and the design of classroom assessment in instructor teams. *The Aust. Educ. Res.* 20pp 49, 425–444. doi: 10.1007/s13384-021-00437-9
- Coll, C., Mauri, T., and Rochera, M. (2012). La práctica de la evaluación como contexto para aprender a ser un aprendiz competente. *Profesorado, revista de curriculum y formación del profesorado* 16, 49–59. <http://www.ugr.es/~recfpro/rev161ART4.pdf>, <https://revistaseugr.es/index.php/profesorado/articulo/view/19769>
- Crowell, T. L. (2015). Student self grading: perception vs. reality. *Am. J. Educ. Res.* 3, 450–455. doi: 10.12691/education-3-4-10
- Estrada, F.G. (2021). *Revisión y modificación de concepciones del profesorado sobre la evaluación a través de la autoevaluación sincrónica*. Paper presented at XV-Congreso Internacional de Educación e Innovación. Florence, Italy. December, 13–15th, 2021.
- Fulton, K., and Britton, T. (2011). *STEM Instructors in Professional Learning Communities: From Good Instructors to Great Teaching*. Washington, DC: National Commission on Teaching and America's Future.
- Hernández Nodarse, M. (2007). Perfeccionando los exámenes escritos: reflexiones y sugerencias metodológicas. *Rev. Iberoam. de Educ.* 41, 1–25.
- Hernández Nodarse, M., Gregori, M., and Tomalá, D. (2018). ¿Cómo puede transformarse verdaderamente la evaluación del aprendizaje practicada por los profesores? Una experiencia pedagógica por la mejora del profesorado. *Revista Ciencias Pedagógicas e Innovación* 6, 88–106. doi: 10.26423/rcpi.v6i1.221

the analysis of the data implied in this study. This research was funded by the University of Barcelona, ref. REDICE182200.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Herrero, P. G., and Medina, A. R. (2019). La exigencia cognitiva en los exámenes tipo test en contexto universitario y su relación con los enfoques de aprendizaje, la autorregulación, los métodos docentes y el rendimiento académico. *Eur. J. Investig. Health, Psychol. Educ.* 9, 177–187. doi: 10.30552/ejihpe.v9i3.333

Hong, J., Nie, Y., Heddy, B., Monobe, G., Ruan, J., You, S., et al. (2016). Revising and validating achievement emotions questionnaire – teachers (AEQ-T). *Int. J. Educ. Psychol.* 5, 80–107. doi: 10.17583/ijep.2016.1395

Imbulpitiya, A., Whalley, J., and Senapathi, M. (2021). Examining the exams: bloom and database modelling and design. In *Australasian Computing Education Conference* (pp. 21–29). doi: 10.1145/3441636.3442301

Lau, A. M. S. (2016). 'Formative good, summative bad?' – a review of the dichotomy in assessment literature. *J. Furth. High. Educ.* 40, 509–525. doi: 10.1080/0309877X.2014.984600

López Espinosa, G. J., Quintana, R., Rodríguez Cruz, O., Gómez López, L., Pérez de Armas, A., and Aparicio, G. (2014). El profesor principal y su preparación para diseñar instrumentos de evaluación escritos. *Edumecentro* 6, 94–109. Available at: http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S2077-28742014000200007&lng=es

Nicol, D. (2021). The power of internal feedback: exploiting natural comparison processes. *Assessment & Evaluation in Higher Education* 46, 756–778. doi: 10.1080/02602938.2020.1823314

Panadero, E., Broadbent, J., Boud, D., and Lodge, J. M. (2019). Using formative assessment to influence self-and co-regulated learning: the role of evaluative judgement. *Eur. J. Psychol. Educ.* 34, 535–557. doi: 10.1007/s10212-018-0407-8

Popham, W. J. (2009). Assessment literacy for instructors: faddish or fundamental? *Theory Pract.* 48, 4–11. doi: 10.1080/00405840802577536

Remesal, A. (2006). *Los Problemas En La Evaluación Del Area De Matemáticas En La Educación Obligatoria: Perspectiva De Profesores y Alumnos*. Tesis Doctoral presentada en la Universidad de Barcelona. Director: Dr.César Coll Salvador.

Remesal, A. (2011). Primary and secondary instructors' conceptions of assessment: a qualitative study. *Teach. Instructor Educ.* 27, 472–482. doi: 10.1016/j.tate.2010.09.017

Remesal, A. (2020). *Autoevaluación sincrónica. Conferencia presentada en el Primer Encuentro Virtual Internacional: Enseñanza, Aprendizaje y Evaluación en la Educación Superior*. Universidad del Desarrollo, Santiago de Chile.

Remesal, A. (2021). "Synchronous self-assessment: assessment from the other side of the Mirror" in *Assessment as Learning: Maximising Opportunities for Student Learning and Achievement*, eds. Z. Yan and L. Yang (England: Routledge)

Remesal, A., Colomina, R., Nadal, E., Vidosa, H., and Antón, A. M. (2021). "Diseño de exámenes para autoevaluación sincrónica: emociones implicadas" in *Paper Presented at XI CIDUI 2020+1. Más Allá De Las Competencias: Nuevos Retos En La Sociedad Digital* (Barcelona). Universitat de Barcelona.

Remesal, A., Estrada, F. G., and Corrial, C. L. (2022). "Exams for the purpose of meaningful learning: new chances with synchronous self-assessment" in *Design and Measurement Strategies for Meaningful Learning*, ed. J. L. Gómez-Ramos (ISI-Global), 192–211. Pensilvania.

Remesal, A., Khanbeiki, A., Attareivani, S., and Parham, Z. (2019). *In Situ, Synchronous Self-Assessment: A New Research Strategy for Accessing Individual SA Processes*. Paper presented at EARLI 2019. Aachen, Germany, 12th–16th, August, 2019.

- Roméu, M. R., and Díaz Quiñones, J. A. (2015). Valoración metodológica de la confección de temarios de exámenes finales de Medicina y Estomatología. *Educ. Méd. Super.* 29, 522–531. Available at: http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S0864-21412015000300011&lng=es&tlng=es
- Rusak, G., and Yan, L. (2021). Unique exams: designing assessments for integrity and fairness. In Proceedings of the 52nd ACM Technical Symposium on Computer Science Education (pp. 1170–1176).
- Schutz, P. A., and Pekrun, R. E. (Eds.) (2007). *Emotion in Education* Elsevier.
- Schutz, P. A., and Zembylas, M. (Eds.) (2009). *Advances in Teacher Emotion Research: The Impact on Teachers' Lives* Springer NY.
- Sutton, P. (2009). Towards dialogic feedback. *Critical Reflective Pract. Educ.* 1, 1–10. Cambridge.
- Sutton, R. E., Mudrey-Camino, R., and Knight, C. C. (2009). Teachers' emotions and classroom effectiveness: implications from recent research. *Theory Pract.* 48, 130–137. doi: 10.1080/00405840902776418
- Tai, J., Ajjawi, R., Boud, D., Dawson, P., and Panadero, E. (2018). Developing evaluative judgement: enabling students to make decisions about the quality of work. *High. Educ.* 76, 467–481. doi: 10.1007/s10734-017-0220-3
- Taras, M. (2009). Summative assessment: the missing link for formative assessment. *J. Furth. High. Educ.* 33, 57–69. doi: 10.1080/03098770802638671
- Van Den Heuvel-Panhuizen, M. (2005). The role of contexts in assessment problems in mathematics. *For the learning of mathematics* 25, 2–23.
- Villarroel, V., Bruna, D., Brown, G. T., and Bustos, C. (2021). Changing the quality of instructors' written tests by implementing an authentic assessment instructors' training program. *Int. J. Instr.* 14, 987–1000. doi: 10.29333/iji.2021.14256a
- Xu, Y., and Brown, G. T. (2016). Instructor assessment literacy in practice: a reconceptualisation. *Teach. Instructor Educ.* 58, 149–162. doi: 10.1016/j.tate.2016.05.010
- Yan, Z. (2020). Self-assessment in the process of self-regulated learning and its relationship with academic achievement. *Assess. Eval. High. Educ.* 45, 224–238. doi: 10.1080/02602938.2019.1629390
- Yan, Z., and Boud, D. (2021). "Conceptualising assessment as learning" in *Assessment as Learning. Maximising Opportunities for Student Learning and Achievement*. eds. Z. Yan and L. Yang (England: Routledge), 11–24.
- Yan, Z., and Brown, G. T. L. (2017). A cyclical self-assessment process: towards a model of how students engage in self-assessment. *Assess. Eval. High. Educ.* 42, 1247–1262. doi: 10.1080/02602938.2016.1260091
- Yin, R. K. (2014). *Case Study Research: Design and Methods*. 5th SAGE, California
- Yu, L., and Suen, H. K. (2005). Historical and contemporary exam-driven education fever in China. *KEDI J. Educ. Policy* 2, 17–33.