

Developing student competence through peer assessment: the role of feedback, self-regulation and evaluative judgement

AUTHOR(S)

M S Ibarra-Sáiz, G Rodríguez-Gómez, David Boud

PUBLICATION DATE

14-01-2020

HANDLE

10536/DRO/DU:30134272

Downloaded from Deakin University's Figshare repository

Deakin University CRICOS Provider Code: 00113B

Higher Education
https://doi.org/10.1007/s10734-019-00469-2

Developing student competence through peer assessment: the role of feedback, self-regulation and evaluative judgement

María Soledad Ibarra-Sáiz[1] · Gregorio Rodríguez-Gómez[1] · David Boud[2]

Published online: 14 January 2020
© The Author(s) 2019

Abstract

How can students' competence be developed through peer assessment? This paper focuses on how relevant variables such as participation, evaluative judgement and the quality of the assessment interact and influence peer assessment. From an analysis of 4 years of data from undergraduate classes in project management, it develops a model of causal relationships validated using the PLS-SEM method. It demonstrates relationships between these variables and considerers the influence of students' competence and the mediating nature of feedback and self-regulation on the process. It points to how peer assessment practices can be improved whilst highlighting how evaluative judgement and feedback are two key elements that can be addressed to deliver the effective development of students' competence.

Keywords Peer assessment · Feedback · Evaluative judgement · Self-regulation · Structural equation modelling · Importance-performance map analysis

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s10734-019-00469-2) contains supplementary material, which is available to authorized users.

✉ María Soledad Ibarra-Sáiz
marisol.ibarra@uca.es

Gregorio Rodríguez-Gómez
gregorio.rodriguez@uca.es

David Boud
david.boud@deakin.edu.au

[1] EVALfor Research Group, Faculty of Educational Sciences, University of Cadiz, Puerto Real, Spain

[2] Centre for Research in Assessment and Digital Learning (CRADLE), Deakin University, Geelong, Australia

Introduction

In recent years, we have witnessed a progressive evolution of assessment processes that has changed the focus of attention towards students' strategic and lifelong learning. From the recall of knowledge, emphasis has moved onto students' ability to respond to assessment tasks that are divergent rather than convergent and complex rather than simple (Sadler 2012). Among the approaches accompanying this refocusing are assessment for learning (Carless et al. 2017; Lai 2006; Sambell et al. 2013), learning-oriented assessment (Carless et al. 2006; Carless 2015), assessment as learning (Dann 2014; Earl 2013), sustainable assessment (Boud and Soler 2016; Nguyen and Walker 2016) and assessment as learning and empowerment (Rodríguez-Gómez and Ibarra-Sáiz 2015).

In these different approaches, particular significance is given to participatory modalities of assessment, such as self-assessment and peer assessment. The reviews carried out by Dochy et al. (1999), Gielen et al. (2011) and Panadero (2016) illustrate the variety of ways in which peer assessment can be implemented in practice.

Given the centrality of peer assessment in reforming assessment practices, the purpose of this paper is to examine how peer assessment practices can be analysed and thereby improved. It develops an exploratory and predictive model that considers the key variables involved in peer assessment. To do this, it discusses theoretical foundations that suggest possible causal relationships between relevant variables. The model developed is validated using data from peer assessment activities with undergraduates in Spain taking an Economics and Business Degree over four academic years. The students experienced peer assessment as part of the course, using the EvalCOMIX® web service, which was expressly developed to promote participatory modes of assessment (Ibarra-Sáiz and Rodríguez-Gómez 2017).

Specifically, this paper aims to:

- Provide a predictive model of the competence development of students, based on the practice of peer assessment and which illustrates the relationships between variables such as evaluative judgement, participation, feedback, self-regulation and the quality of the assessment.
- Offer an instrument that facilitates analysis and understanding of the perception of university students about peer assessment practices using technological resources.
- Orientate the practice of peer assessment towards those aspects with the greatest potential for improving students' competence development.

Framework and development of hypotheses

Practicing peer assessment

Until the 1990s, assessment processes in universities tended to focus on what students knew. Students were assessed, above all, on their understanding of some domain of specific knowledge within the subject area they studied. Progressively, the emphasis has been refocused onto what students can do and the value of transferable, generic or essential skills, that is, the skills and competences that all students should develop (Boud 2014; Strijbos et al. 2015).

As other authors have argued (Nicol et al. 2014; Thomas et al. 2011), learning with peers is assessed because it is a key skill required for lifelong learning, which involves critical thinking and reflection and being able to evaluate one's own work and that of others. Universities increasingly focus their efforts on these skills so that they represent an essential part of what students learn throughout their university studies.

At the beginning of any peer assessment practice we are faced with a situation in which the role of students needs to substantially change. Giving the student a voice implies modifying the traditional relationship of power in assessment processes, from one in which the lecturer holds a dominant position to a more equitable and democratic relationship in which students assume responsibility themselves as an assessor. Assessment therefore transforms from a unidirectional process, dominated by lecturers, to a socio-constructive and dynamic process in which lecturers and students interact (Rust et al. 2005).

This new situation requires numerous variables to be taken into consideration. Aspects such as participation (Falchikov 2005), students' evaluative judgement (Boud et al. 2018a, b), self-regulation (Hawe and Dixon 2017; Panadero et al. 2017), feedback (Boud and Molloy 2013; Nicol et al. 2014), a climate of trust (Carless 2009, 2013) and the quality of the assessment (Sadler 2016) are all elements that play a vital role in assessment practice. In this study, we focus primarily on the two elements that can be considered as basic to peer assessment: student participation and their evaluative judgement. Secondly, we analyse the role played by feedback and self-regulation, as well as the value that students attribute to peer assessment in contributing to their competence development.

Participation and evaluative judgement

Contemporary theories acknowledge the central role of the student in the construction of their own learning (Penuel and Shepard 2016). In the case of peer assessment, the importance of student participation has been highlighted by Falchikov (2005), Thomas et al. (2011), Reinholz (2016) and López-Pastor and Sicilia-Camacho (2017). Participation means encouraging dialogue with students and enabling them to collaborate in the process of assessing their learning in ways that are transparent. This participation can be established during all three phases of the assessment process: planning, development and judgement.

During the planning, phase tutors can decide or agree with their students the selection or definition of criteria, the means of assessment, the design of assessment instruments or the grading system. When the assessment is carried out, students can participate by assessing their own work and actions and those of their peers, through assessment modes such as self-assessment or peer assessment. Finally, students can participate fully in the grading process through dialogue and consensus around the grades allocated.

For the students' judgement on assessed work to be fair and equitable, it is vital that they have significant competence in the practice of assessment. The importance of students' ability to make judgements has been evidenced by Boud and Falchikov (2007), Cowan (2010), Yucel et al. (2014), Nguyen and Walker (2016) or Murillo and Hidalgo (2017). In fact, Boud (2014, p. 27) highlights the importance of the development of informed judgement as one of the strategic axes in the assessment change agenda, because it is the "'sine qua non' of assessment".

Tai et al. (2018, p. 471) define evaluative judgement as "the ability to make decisions about the quality of work of self and others". In short, this implies the identification or discernment of standards, the application of them to a given piece of work, techniques for calibrating judgement and mechanisms to avoid being fooled (Boud 2016).

Under this conception of evaluative judgement lies a double dimensionality, in so far as it supposes an assessment of one's own work and that of others. Furthermore, if we add another determining factor in the assessment process, such as trust or a lack of trust (Carless 2009, 2013), we can consider that, in the context of peer assessment, evaluative judgement is underpinned by trust in one's own judgement and trust in the judgement of others.

Feedback and self-regulation

In peer assessment, the role of feedback is crucial. In general, research shows that feedback is associated with learning and performance because, as Hounsell (2007) states, feedback can improve learning in three different ways: by accelerating learning, by optimizing the quality of what is learned and by raising the level of achievement of individuals and of groups.

There are numerous contributions to the characteristics of high-quality feedback, as well as suggestions on how further improvements can be made (Ajjawi and Boud 2017, 2018; Boud and Molloy 2013; Espasa and Meneses 2010; Evans 2013; Gielen et al. 2011; Pardo 2018). However, in recent years, we have seen a change in the meaning and purpose of feedback. Previously, special attention was paid to features such as speed, level of detail, clarity, structure or relevance in the delivery of information to students; but nowadays, attention has shifted towards the meaning of feedback for the student and the interaction between the student and the giver of the feedback (Rowe 2017). Feedback has evolved from being perceived as a one-way process of transmitting information from lecturer to student, to being considered as a process using multiple communication channels, through which lecturers and students interact with each other in order to lead to improved outcomes. This highlights the importance of facilitating the participation of students as a source of feedback and learning (Moore and Teather 2013; Nicol et al. 2014)

In the context of peer assessment, it is essential for students to understand what quality feedback involves. They need to learn to evaluate and make judgements about the quality of the work of others whilst maintaining a dialogue with their lecturers and colleagues about the quality of their assessment (Sadler 2012). Consequently, it is important for students to develop the ability to make judgements and evaluate constructively. However, that alone is not sufficient. In addition, students must be able to use the feedback offered in such a way as to reduce the gap between the feedback given and the feedback used (Cartney 2012).

This change of the focus and the actors in feedback implies that lecturers should pay less attention to delivering unidirectional, focused and direct feedback and pay more attention to how students understand and interpret multidirectional feedback from their self-regulatory and self-productive identities (Dann 2014).

The self-regulation of learning is an essential feature of effective learning processes, as supported by multiple existing models (Panadero 2017). One of the assumptions underlying self-regulated learning is the importance of the nature of mediation that takes place between personal and contextual characteristics and the level or degree of achievement or execution (Pintrich 2000; Järvelä et al. 2016). The work of Panadero et al. (2018) shows that many studies have explored the relationships between self-regulated learning, the use of learning strategies and academic performance. However, few studies have focused their attention on the role of assessment as an element conducive to self-regulated learning.

Assessment quality

The achievement of high and consistent quality in assessment practices is required by students (Smith and Coombe 2006). With regard to assessment standards, it has been proposed that "Classroom assessment practices meet the standards of quality when teachers can be confident that their assessment practices provide accurate and dependable information about student learning" (Klinger et al. 2015). In fact, the generating of quality evidence is one basic principle in the Berkeley Evaluation and Assessment Research Assessment (BEAR) System, based on four principles: (1) developmental perspective, (2) a match between instruction and assessment, (3) the generating of quality evidence, and (4) management by instructors to allow appropriate feedback and follow-up (Wilson and Scalise 2006, p. 646). However, this is a partial view since it only takes into consideration the perspective of teachers. Quality assessment should be perceived by students as a rigorous evaluation, that is both valuable and interesting. Assessment quality also relates to other aspects mentioned previously, such as trust in the judgement of others and the usefulness of their judgements.

During peer assessment, if students do not receive pertinent, constructive feedback from their peers, there is a risk that they will perceive their judgements as unfair and discouraging. It is vitally important, therefore, that quality feedback information is given during peer assessment to reduce any perception of injustice and increase students' motivation and commitment (Moore and Teather 2013).

Research model and hypotheses

The model adopted here proposes that, within the context of an assessment process based on peer assessment, competency development is determined by self-regulation and feedback which, in turn, are conditioned by the quality of assessment, being itself dependent on participation and evaluative judgement. Figure 1 illustrates the base model indicating the relationships between the constructs. Table 1 summarises the definition of the constructs in the model.

Based on this theoretical model and within the scope of the contributions presented in this study, we propose the following hypotheses:

H1. Evaluative judgement consists of two components; trust in one's own judgement and confidence in the judgement of others

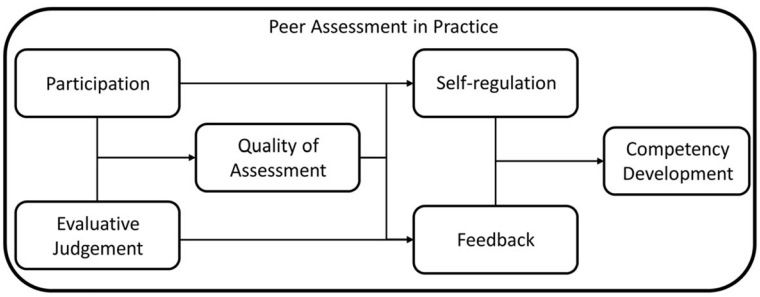


Fig. 1 Model for testing drivers of competency development in the context of peer assessment

Table 1 Constructs definition

Construct	Definition	Key references
Competency	Competences are an integration of knowledge, skills and attitudes and are situated in or aligned to a professional context	Brabrand and Dahl 2009; Strijbos et al. 2015; Zlatkin-Troitschanskaia et al. 2015
Feedback input	Specific and detailed information on the quality of work and improvement in commitment aimed at enabling improvement	Hattie and Timperley 2007; Boud and Molloy 2013; Nicol et al. 2014; Pardo 2018
Self-regulation	Self-generated thoughts, feelings and actions planned and cyclically adapted to the attainment of personal goals	Hawe and Dixon 2017; Panadero et al. 2017, 2018
Quality of assessment	Assessment that is rigorous, credible, interesting and useful for learning	Klinger et al. 2015; Sadler 2016; Carless et al. 2017
Evaluative judgement	The ability to make decisions about the quality of work of self and others	Tai et al. 2016; Boud et al. 2018a, b;Tai et al. 2018
Participation	Design and wording of assessment criteria and instruments to assess their own work and that of their peers	Falchikov 2005; Rodríguez-Gómez and Ibarra-Sáiz 2015; López-Pastor and Sicilia-Camacho 2017

H2a. Evaluative judgement is expected to be positively related to the quality of the assessment
H2b. Participation is expected to be positively related to the quality of the assessment
H3a. Participation is expected to be positively related to the development of competence
H3b. Evaluative judgement is expected to be directly related to the development of competence
H3c. Self-regulation is expected to be positively related to the development of competence
H3d. Feedback is expected to be positively related to the development of competence
H4a. The relationship between evaluative judgement and the development of competence is expected to be mediated by feedback.
H4b. The relationship between participation and the development of competence is expected to be mediated by feedback
H5a. The relationship between evaluative judgement and the development of competence is expected to be mediated by self-regulation
H5b. The relationship between participation and the development of competence is expected to be mediated by self-regulation
H6a. The relationship between participation and the development of competence is expected to be mediated by the quality of the assessment
H6b. The relationship between evaluative judgement and the development of competence is expected to be mediated by the quality of the assessment

Methodology

This study was carried out during four academic years (2012/2013, 2013/2014, 2016/2017 and 2017/2018) using peer assessment as an integral part of the assessment process in the subject Project Management. A cohort design for the research was used, since the self-perception of the students was collected at the end of each semester. Consequently, different students

responded in each academic year. A set of three assessment tasks was designed in which the students had to perform both self-assessment and peer assessment of the products or actions that were being assessed.

Peer assessment in practice

The way in which peer assessment is implemented in practice is of fundamental importance if the intention is to enable the study to be replicated and to compare and synthesise results (Topping 2010). Consequently, with the intention of avoiding what Makel and Plucker (2014) refer to as "replication crisis", the characteristics of each of these three assessment tasks are described in Appendix A (Online Resource 1). In addition, based on the nineteen elements that should be considered in the description of peer assessment proposed by Adachi et al. (2018), Appendix B (Online Resource 2) describes each of these nineteen elements. The assessment tools used in each of the tasks are summarised in Table 2 and can be found in Appendix C (Online Resource 3).

The EvalCOMIX® web service (Ibarra-Sáiz and Rodríguez-Gómez 2017), integrated into the Moodle server of the university's virtual campus, was used to design, manage and apply all assessment instruments. Using EvalCOMIX® facilitated the peer assessment process, the delivery of feedback and the final calculation of grades, based on the criteria and weighting of each of the elements within the assessment task.

Participants

A total of 301 students from the Faculty of Economic and Business Sciences of the University of Cádiz (Spain) participated in the study (Table 3). These students were taking the Project Management module, taught in the final year of the Business Administration and Management Degree (BAM) and the Finance and Accounting Degree (FINA).

Instrument

In order to obtain the students' views on participating in assessment processes through peer assessment, we designed an ad hoc questionnaire, the "Student Perception of Peer Assessment in Practice" questionnaire (Appendix D) (Online Resource 4).

Figure 2 illustrates the process followed in the design and validation of the questionnaire. The constructs and their items of measurement were developed first, following an extensive literature review. The content validity was then determined using the process of validation by

Table 2 Assessment tools used for every assessment task

	Assessment task	Assessment tool
Task 1	Design and completion of a comparative report on documents relating to Business and Economics	• Rubric for comparative reports on professional documents • Rating scale for oral presentations
Task 2	Oral presentation on predictive and agile methodologies for Project Management	• Rating scale for oral presentations on methodologies
Task 3	Design and planning of a project in the area of Business and Economic Science	• Rating scale for project assessment

Table 3 Demographic characteristics

		Female		Male		Total	
		n	%	*n*	%	*n*	%
Degree	FINA	22	62.9	13	37.1	35	11.6
	BAM	152	57.1	114	42.9	266	88.4
Cohort	2013	27	61.4	17	38.6	44	14.6
	2014	36	56.3	28	43.7	64	21.3
	2017	43	62.3	26	37.7	69	22.9
	2018	68	54.8	56	45.2	124	41.2
Total		174	57.8	127	42.2	301	100

experts. Of the various methods available (Johnson and Morgan 2016; Litwin 2003), we chose to use the group consensus option in which the experts consulted arrive at a final product on which they all agree, following an incremental and iterative process in several cycles. Five experts in assessment in higher education were consulted, chosen using the criteria proposed by Skjong and Wentworth (2000) regarding making judgements, decision making, availability, motivation and impartiality. Starting with an initial proposition, two iterative cycles were completed before the final version was agreed. The definition of the constructs was revised during each cycle and the appropriateness of each of the indicators was considered during discussions of approximately an hour and a half. Finally, to achieve face validity, the final version of the questionnaire was shown to a group of 8 Masters students who amended it from the perspective of the clarity of the language and ease of understanding.

The final version comprised 40 items in a Likert scale format (1–6) structured around seven dimensions (Table 4).

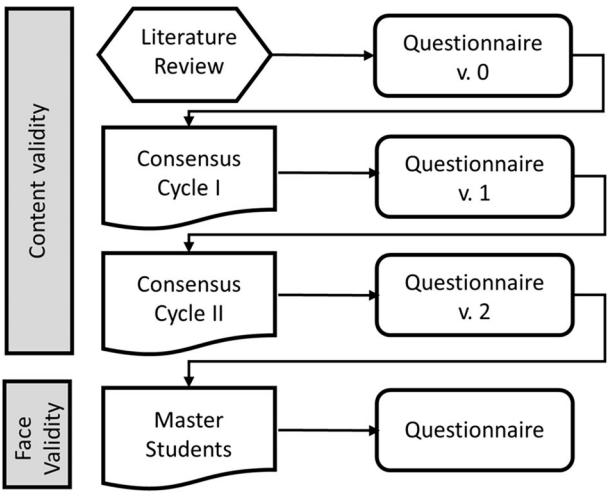


Fig. 2 Questionnaire design process

Table 4 Participation satisfaction questionnaire structure

Dimension		No. of items
CODEVP	Competency development	11
SELFRP	Self-regulation	5
PARTIP	Participation	6
FEEDFP	Feedback	4
QUASSP	Quality of the assessment	6
OWNJUP	Confidence in their own judgement	4
OTHJUP	Confidence in the judgement of others	4

Data analysis

Partial least squares structural equation modelling (PLS-SEM) method and the statistical software SmartPLS 3 (Ringle et al. 2015) were used. This is a second-generation technique designed to overcome the weakness in more traditional exploratory type first-generation methods such as cluster analysis, exploratory factor analysis or multidimensional scaling (Hair et al. 2017). PLS-SEM is used specifically to develop theories in exploratory research by focusing on the explanation of variances in dependent variables when analysing a model.

PLS-SEM is recommended when, as here, the objective is the prediction of an objective construct or the intention is to identify "driver constructs"; the research model is complex according to the type of relationships hypothesised (direct and mediation) and the levels of dimensionality (first-order and second-order constructs); formatively measured constructs are part of the structural model; the structural model is complex and the data is non-normal (Hair et al. 2017).

Confirmatory tetrad analysis (CTA-PLS) was employed to confirm the formative or reflective nature of the constructs. This is used to check the adequacy of the specification of the measurement model and test the null hypothesis that the indicators for a model are reflective (Garson 2016), so that the reflective or formative nature of the latent variables can be confirmed (Hair et al. 2018).

The evaluation of the model has been carried out according to the reflective (Mode A) or formative (Mode B) character of the model. Once the measurements of the constructs were confirmed as reliable and valid, we proceeded to analyse the predictive capacity of the model and the relationships between the constructs.

Finally, the importance-performance map analysis (IPMA) technique was used to identify predecessor constructs that have a relatively high importance for predicting the target construct, but "also have a relatively low performance so that improvements can be implemented" (Hair et al. 2018, p. 105). This technique allows constructs and indicators to be easily identified so they can be modified to improve results in an effective way and enable improvements or changes to be prioritised.

Results

Before proceeding to the results of our evaluation of the measurements model and the structural model, Table 5 offers descriptive results for each of the variables measured and an

Table 5 Descriptive statistics and contrast tests (Mann-Whitney *U* and Kruskal-Wallis)

Constructs	Descriptives			Gender		Degree		Cohort	
	Mdn	*M*	*SD*	*U*	Sig.	*U*	Sig.	*H*	Sig.
Competency development (CODEVP)	4.18	4.09	.84	10742	.680	3287.5	.005**	21.2	.000***
Self-regulation (SELFRP)	4.20	4.22	.99	9772	.086+	2995.5	.001***	30.5	.000***
Participation (PARTIP)	4.17	4.02	.97	10514	.472	3899.5	.118	26.6	.000***
Feedback (FEEDFP)	4.00	3.78	1.01	10671	.611	3436.0	.012*	23.5	.000***
Quality of the assessment (QUASSP)	3.83	3.75	.98	10970	.916	3339.5	.006**	11.4	.010**
Confidence in their own judgement (OWNJUP)	4.00	4.01	.97	10936	.880	3999.5	.174	4.58	.205
Confidence in the judgement of others (OTHJUP)	3.50	3.53	.52	10668	.605	3686.5	.043*	6.60	.086+

+ $p < .10$; * $p < .05$; ** $p < .01$; *** $p < .001$

analysis of the possible differences between groups by gender, degree studied and cohort. Table 5 shows that in terms of gender, the only statistically significant difference is in self-regulation whilst there are differences in practically all of the variables when considered by degree studied and cohort.

We initially considered all the constructs to be formative but, after carrying out confirmatory tetrad analysis (CTA-PLS), no empirical evidence of this could be seen for the constructs feedback (FEEDFP), confidence in their own judgement (OWNJUP) and confidence in the judgement of others (OTHJP) so the decision was taken to consider them as reflective.

Evaluation of the measurement model

(a) Reflective model

The evaluation of the measurement model for reflective indicators in PLS-SEM is based on internal consistency reliability, convergent validity and discriminant validity (Hair et al. 2017). As the values of internal consistency reliability and Cronbach's Alpha values are above the 0.70 threshold, we can conclude that the four constructs are reliable. Average variance extracted (AVE) values for latent variables are greater than 0.61. Thus, the measure of the four reflective constructs has high levels of convergent validity. The Heterotrait-Monotrait Ratio (HTMT) uses 0.85 as the relevant threshold level, a criterion which is also met in this study. This means that all the constructs are empirically distinct (Online Resource 5).

(b) Formative model

With values of variance inflation factor (VIF) between 1.43 and 4.06, we can conclude that collinearity does not reach critical levels in any of the formative constructs and is not an issue for the estimation of the PLS path model (threshold value of 5). Some indicators were found whose values were not statistically significant but instead had loads greater than 0.5, so according to the rules of thumb expressed by Hair et al. (2017, p. 151), all the formative indicators were maintained (Online Resource 6).

Evaluation of structural model

All variance inflation factor (VIF) values are clearly below the threshold of 5. Therefore, collinearity among the predictor constructs is not a critical issue in the structural model. In order to assess the statistical significance of the path coefficients, consistent with Hair et al. (2017), bootstrapping (5000 resamples) was used to generate *t*-statistics and confidence intervals (Table 6). Figure 3 shows the results obtained in the evaluation of the model.

We can confirm the predictive value of the model through the analysis of the coefficient of determination (R^2). Thus, it is evident how 65.9% of the variance (R^2) of the competence development construct (CODEVP) is explained by four essential constructs. The strongest effect is exercised by the feedback construct (FEEDFP, 0.429), followed by self-regulation (SELFRP, 0.261), evaluative judgement (JUDGEP, 0.140) and participation (PARTI, 0.128).

Evaluative judgement (JUDGEP) is a hierarchical component model (HCM) constructed under a repeated indicators approach (Hair et al. 2017, 2018). That is to say, it is a multidimensional construct, formed by the confidence in one's own judgement (OWNJUP) and trust in the judgement of others (OTHJUP). Our research model achieves SRMR of 0.072 (Fig. 2), which means an appropriate fit, taking into account the usual cut-off of 0.08.

All Stone-Geisser's Q^2 values (predictive relevance) for endogenous constructs are considerably above zero (Online Resource 7). More precisely, evaluative judgement (JUDGEP) has the highest Q^2 value (0.530), followed by feedback (FEEDF, 0.397), quality of the assessment (QUASSP, 0.404), competency development (CODEVP, 0.350) and, finally, self-regulation (SELFRP, 0.347). These results provide clear support for the model's predictive relevance regarding the endogenous latent variables. A medium value is reached by effect sizes (q^2) in the cases of evaluative judgement (JUDGED) on quality of the assessment (QUASSP) and participation (PARTI) on feedback (FEEDFP), with smaller values in the other cases.

Table 6 Structural model results using *t* values and percentile bootstrap 95% confidence interval ($n = 5000$ subsamples)

Relationship	Path coefficients (*)				Effect size (+)				Hypothesis
	Path	95% CI	*t*	*p*	f^2	95% CI	*t*	*p*	
JUDGEP → OTHJUP	0.901	[0.864, 0.924]	63.457	0.000	4.303	[0.833, 0.833]	5.859	0.000	H1
JUDGEP → OWNJUP	0.872	[0.824, 0.905]	42.247	0.000	3.187	[0.753, 0.753]	5.022	0.000	
JUDGEP → QUASSP	0.629	[0.547, 0.705]	15.585	0.000	0.858	[0.475, 0.475]	4.745	0.000	H2a
PARTI → QUASSP	0.278	[0.184, 0.357]	6.267	0.000	0.168	[0.138, 0.138]	2.761	0.006	H2b
PARTI → CODEVP	0.128	[0.032, 0.246]	2.521	0.007	0.033	[− 0.065, 0.031]	1.182	0.237	H3a
JUDGEP → CODEVP	0.140	[0.032, 0.246]	2.519	0.012	0.034	[− 0.063, 0.039]	1.125	0.261	H3b
SELFRP → CODEVP	0.261	[0.099, 0.412]	3.276	0.000	0.099	[− 0.026, 0.089]	1.473	0.141	H3c
FEEDF → CODEVP	0.429	[0.286, 0.594]	5.435	0.000	0.215	[0.170, 0.170]	2.504	0.012	H3d

*0.75 substantial/0.50 moderate/0.25 weak // + 0.35 large/0.15 medium/0.02 small

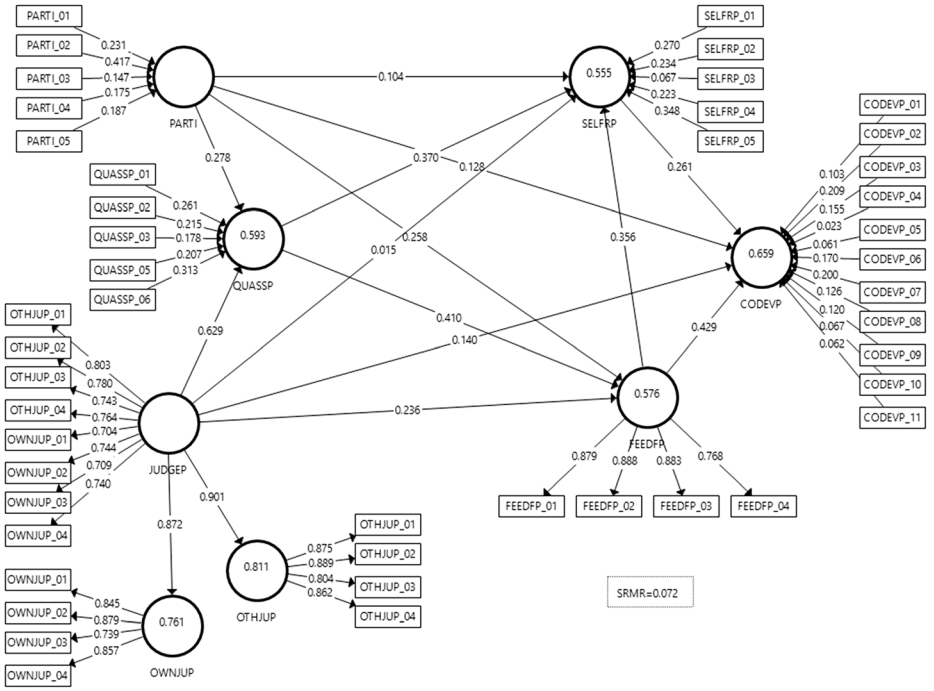


Fig. 3 Structural model results

Mediation analysis

(a) Feedback as mediator

In our model (Fig. 2), feedback operates as a mediating variable between participation and competence development (Table 7). We found a significant specific indirect effect of the relationship from participation to competence development (0.111). Participation has a significant direct effect (0.128) on competence development. Both the direct and the specific indirect effects from participation to competence development are significant, which indicates that feedback partially mediates the relationship between them (Hair et al. 2017). Moreover, the product of the direct effect and specific indirect effects are positive. Hence, the result reveals that feedback represents a complementary partial mediation for the path from participation to competence development. This complementary mediation suggests that feedback explains the relationship between participation and competence development. When students participate by being involved in producing assessment criteria or in selecting assessment instruments, it contributes to their competence development due to the learning produced by assessing their own learning and that of their peers.

Feedback also operates as a mediating variable between evaluative judgement and competence development (Table 8). We found a significant specific indirect effect of the relationship from evaluative judgement to competence development (0.101). Evaluative judgement has a significant direct effect (0.140). Both the direct and the indirect effects from evaluative judgement to competence development are significant, which indicates that feedback partially

Table 7 Summary of mediating effect test of PARTI on CODEVP

	Effect	95% CI	*t*	*p*	Mediation type
Total effects					
PARTI → CODEVP	0.376	[0.247, 0.476]	6.504	0.000	
Direct effects					
PARTI → CODEVP	0.128	[0.026, 0.224]	2.521	0.012	
Total indirect effect					
PARTI → CODEVP	0.248	[0.175, 0.321]	6.670	0.000	
Specific indirect effect					
PARTI → FEEDF → CODEVP	0.111	[0.062, 0.176]	3.782	0.000	Complementary
PARTI → QUASSP → FEEDF → CODEVP	0.049	[0.025, 0.085]	3.189	0.001	Complementary
PARTI → FEEDF → SELFRP → CODEVP	0.024	[0.007, 0.053]	1.995	0.046	Complementary
PARTI → QUASSP → FEEDF → SELFRP → CODEVP	0.011	[0.003, 0.024]	1.994	0.046	Complementary
PARTI → SELFRP → CODEVP	0.027	[−0.003, 0.072]	1.443	0.149	No mediation
PARTI → QUASSP → SELFRP → CODEVP	0.027	[0.009, 0.052]	2.363	0.018	Complementary

mediates the relationship between them. Moreover, the product of the direct effect and indirect effect are positive. Hence, the result reveals that feedback represents a complementary partial mediation for the path from evaluative judgement to competence development. This complementary mediation suggests that feedback explains the relationship between evaluative judgement and competence development. Ultimately, the information gained by students through peer assessment aimed at improving their work serves to measure the relationship established between their confidence in their own judgements and those made by their peers about the development of their generic competences.

So far, we have presented the results of the analysis of simple mediation, but as we can see in Fig. 2 and Tables 7 and 8, feedback also operates in the context of multiple mediation, that is, in the mediation that occurs when an exogenous construct exerts its influence through more than one mediating variable. This multiple analysis allows us to consider all the mediators

Table 8 Summary of mediating effect test of JUDGED on CODEVP

	Effect	95% CI	*t*	*p*	Mediation type
Total effects					
JUDGEP → CODEVP	0.462	[0.335, 0.583]	7.315	0.000	
Direct effects					
JUDGEP → CODEVP	0.140	[0.032, 0.246]	2.519	0.012	
Total indirect effect					
JUDGEP → CODEVP	0.322	[0.240, 0.416]	7.229	0.000	
Specific indirect effect					
JUDGEP → FEEDF → CODEVP	0.101	[0.043,0.188]	2.809	0.005	Complementary
JUDGEP → QUASSP → FEEDF → CODEVP	0.111	[0.061,0.187]	3.483	0.001	Complementary
JUDGEP → FEEDF → SELFRP → CODEVP	0.022	[0.006, 0.056]	1.769	0.077	Complementary
JUDGEP → QUASSP → FEEDF → SELFRP → CODEVP	0.024	[0.008, 0.055]	2.038	0.042	Complementary
JUDGEP → SELFRP → CODEVP	0.004	[−0.037, 0.050]	0.182	0.855	No mediation
JUDGEP → QUASSP → SELFRP → CODEVP	0.061	[0.023, 0.111]	2.680	0.007	Complementary

simultaneously in one model, enabling us to obtain a better representation of the mechanisms through which an exogenous construct affects an endogenous construct (Hair et al. 2017).

In the case of our model, we can see in Table 7 how feedback also intervenes in this multiple mediation, between participation and competence development, together with the quality of the assessment (0.049) and self-regulation (0.024). This multiple mediation of feedback between evaluative judgement and competence development is also evident (Table 7), together with the quality of the assessment (0.111) and self-regulation (0.022 and 0.024). In all cases, it is a partial (complementary) mediation.

(b) Self-regulation as mediator

Regarding the mediation of self-regulation between participation and competence development (Table 7), we find that the value of the specific indirect effect is not significant (0.027). Since the direct effect is significant and this indirect effect is not significant, we conclude that there is no mediation, but when analysing multiple mediation, the results indicate that it occurs in conjunction with the quality of the assessment (0.027), with feedback (0.024) and with the quality of the assessment and the feedback together (0.011). In this case, it would be a partial (complementary) mediation.

In the case of the mediation of self-regulation between evaluative judgement and competence development (Table 8) something similar happens. The results indicate that there is no simple mediation (0.004). On the other hand, it is evident that it mediates in conjunction with feedback (0.022), the quality of the assessment (0.061) and with both together (0.024). In this case, there is also a partial (complementary) mediation. Therefore, within the context of collaboration, the process by which students critically analyse their own work and that of their peers and identify omissions or errors that help them improve their own and their peers' work is seen to be a valuable element in developing their competences.

Importance-performance map analysis

In addition to the evaluation of the measurement model, the structural model and the analysis of simple and multiple mediation, an importance-performance map analysis (IPMA) was carried out (Online Resource 8). The rationale of IMPA is "to identify predecessor constructs that have a relatively high importance for predicting the target construct, but also have a relatively low performance so that improvements can be implemented" (Hair et al. 2018, p. 105). In our case, the constructs on which action could be taken to improve competence development are, firstly, feedback and evaluative judgement, followed by participation. If we make self-regulation the objective, we could act on the quality of the assessment. And, finally, if we take as an objective the improvement of feedback, the variable which we should act on is evaluative judgement.

Discussion

In this paper, we intended, firstly, to provide a predictive model of students' competence development based on the practice of peer assessment. Secondly, we sought to propose an instrument through which to analyse and understand the perception of university students about peer assessment practices using technological resources. Finally, we wanted to guide the

practice of peer assessment towards formats with the greatest potential for change and improvement. The results achieved in this study suggest there are important implications, both from a theoretical and practical perspective, to understanding peer assessment processes. At the same time, they also provide insight into future lines of research.

Theoretical implications

The primary objective of this study was to provide a predictive model of students' competence development based on the practice of peer assessment. In this regard, our study reflects the proposals made by Panadero et al. (2018) who suggest that an analysis is needed of the influence that different models of formative assessment and self-regulation have on each other, what practices considered as formative can promote self-regulated learning and under what conditions.

One of the main contributions of this work is the construction of a model that integrates the relationships between significant variables of peer assessment in a university context. The results obtained demonstrate that the hypothesised model can, indeed, predict a large part of the relationships between the variables and show, on the one hand, that participation and evaluative judgement are directly related to competence development and, on the other, the mediating role of feedback and self-regulation in the context of peer assessment.

Hypothesis H1, in which the construction of evaluative judgement is achieved through two constructs: trust in one's own judgement and trust in the judgement of others, has been tested and verified. Likewise, hypothesis H2, that directly relates evaluative judgement and participation with the quality of the assessment, as well as hypothesis H3 regarding the positive relationship between participation, evaluative judgement, self-regulation and feedback on the one hand with competence development on the other, have also been confirmed. Finally, the remaining hypotheses concerning the mediating character of feedback (H4), self-regulation (H5) and quality of evaluation (H6) have also been established.

One of the essential purposes of assessment as learning is that students should become the protagonists of their learning (Coombs et al. 2018; DeLuca et al. 2016). This means tutors must assume more of a role as facilitator. Our study confirms student participation in assessment as a variable that is directly related to their competence development and which exerts a direct influence on other aspects such as feedback or self-regulation.

It has been shown in this study how evaluative judgement, in terms of trust in one's own judgement and in the judgement of others, is directly related to students' competence development, as well as to feedback and self-regulation. The systematic development of evaluative judgement is currently an important challenge for the university curriculum, since it places it at the very centre of education (Boud et al. 2018b).

Practical implications

A second objective that has guided our research has been to develop a technology-enabled instrument to analyse and understand the perception of university students about peer assessment practices. In relation to this, our evaluation of the measurement model employed supports the validity of the questionnaire used to operationalise the latent variables, since the items are relevant and all the items load on the right construct. As a result, tutors now have access to an easy-to-use instrument through which they can collect students' perceptions on the implementation of peer assessment.

In this research, the peer assessment process has been carried out using the EvalCOMIX® web service, which allows for greater speed and efficiency in the assessment process. It requires tutors to design and manage the assessment instruments used and monitor the process to address any problems students have with it. This web service can be an excellent technological tool to facilitate the *impetus for change* identified by Bearman et al. (2017)

The last objective that guided this study was to influence the practice of peer assessment to focus on areas of greatest potential for change and improvement. The results obtained in the IPMA analysis confirm the importance of evaluative judgement and feedback as the primary elements on which to act in order to significantly improve competence development. These results are consistent with the contributions of authors such as Boud et al. (2018a), Dawson et al. (2018), Hernández (2012), Nicol et al. (2014), Rodríguez-Gómez and Ibarra-Sáiz (2015) and Sadler (2016). Consequently, the importance of the mediating role of tutors is crucial. They must foster a climate of trust among students that allows them to carry out a rigorous, credible and objective assessment, whilst providing useful and relevant information for the improvement of future activity. This requires lecturers to educate students about assessment so that they, in turn, can participate and deliver judgements that can be reviewed and contrasted, allowing them to progressively acquire greater confidence in their own judgements and those of their peers.

Limitations and future research

From a methodological perspective, the research described in this paper suffers from three specific limitations. In the first place, the research was carried out in a specific context, with final year project management students in Spain. Research needs to be undertaken in other subject areas and with students at other stages of their studies. Secondly, it is an investigation carried out on the basis of a design with a post-test measurement, meaning the degree of control over the intervening variables is reduced and, in line with the caution advised by Stone-Romero and Rosopa (2008), the inferences that can be made about the mediation model are limited. Finally, the measurement instrument is based on the perception of the students themselves, which, as indicated by Panadero et al. (2018), could be improved by the use of alternative measuring instruments.

To be able to generalise our results more widely, further studies need to be carried out using experimental designs in which both the independent variable (the practice of peer assessment) and the mediating variables (essentially feedback and self-regulation) can be manipulated. Despite these limitations, though, we have been able to verify the great diversity and variability of current assessment practices. This diversity and variability make comparison and generalisation difficult, especially in the context of formal education where experimental studies are often difficult to carry out, but the modes of self-assessment and peer assessment in particular seem to be an issue in the future (Pereira et al. 2017).

Struyven et al. (2005) highlight that students' perceptions serve to guide us in our reflective attempts to improve our educational practices and achieve a higher quality of learning and education for our students. However, a second line of research would be to improve the instrumentation used in the data collection process, incorporating other ways of collecting the students' perceptions and collecting data from sources other than students, whilst incorporating measurements that combine both measurement and intervention (Panadero et al. 2016).

The variables involved in the process of peer assessment are highly complex and interact in so many different ways that it is important to try and achieve a greater level of detail, precision

and understanding of them. It would therefore be valuable if further research was undertaken using mixed methodologies which, on the basis of explanatory sequential design, could explain the quantitative results in terms of qualitative data obtained. As Creswell and Clark (2010, p. 82) argue, this type of design "is most useful when the researcher wants to assess trends and relationships with quantitative data but also be able to explain the mechanism or reasons behind the resultant trends".

Conclusion

In this paper, we have demonstrated how the practice of peer assessment is perceived by students as an element that promotes their competence development. We have devised tools that can facilitate adaptation or replication in other different contexts and have suggested future lines of research that will lead to further improvements in assessment. Likewise, we have shown how the implementation of participatory assessment involves a series of interrelationships between different aspects, highlighting the need to address the improvement of feedback processes and the development of evaluative judgement. In fact, providing a context where assessment processes are rigorous, credible, objective and participative, as well as delivering useful and relevant information for the ongoing development of peer assessment practices.

It is vital that policies are developed in higher education that encourage the creation of contexts in which peer assessment can be incorporated, both from a pedagogical and conceptual perspective and also from a technological perspective. Lecturer education on these practices should be promoted and technological resources provided so that the implementation of participatory assessment methods does not become a continuous struggle to overcome bureaucratic difficulties or technological limitations, which can often frustrate and limit educational improvement and change.

Acknowledgements This work was supported by the DevalS Project funded by the Ministry of Economy and Competitiveness (Ref: EDU2012-31804) and the TransEval Project (Ref: R + D + i 2017/01) funded by University of Cadiz.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

Adachi, C., Tai, J., & Dawson, P. (2018). A framework for designing, implementing, communicating and researching peer assessment. *Higher Education Research and Development, 37*(3), 453–467. https://doi.org/10.1080/07294360.2017.1405913.

Ajjawi, R., & Boud, D. (2017). Researching feedback dialogue: an interactional analysis approach. *Assessment & Evaluation in Higher Education, 42*(2), 252–265. https://doi.org/10.1080/02602938.2015.1102863.

Ajjawi, R., & Boud, D. (2018). Examining the nature and effects of feedback dialogue. *Assessment & Evaluation in Higher Education, 43*(7), 1106–1119. https://doi.org/10.1080/02602938.2018.1434128.

Bearman, M., Dawson, P., Bennett, S., Hall, M., Molloy, E., Boud, D., & Joughin, G. (2017). How university teachers design assessments: a cross-disciplinary study. *Higher Education, 74*(1), 49–64. https://doi.org/10.1007/s10734-016-0027-7.

Boud, D. (2014). Shifting views of assessment: from secret teachers' business to sustaining learning. In C. Kreber, C. Anderson, N. Entwistle, & J. McArthut (Eds.), *Advances and inovations in university assessment*

and feedback (pp. 13–31). Edinburgh: Edinburgh University Press Ltd.. https://doi.org/10.3366/edinburgh/9780748694549.003.0002.
Boud, D. (2016). Current influences on changing assessment: implications for research to make a difference. In EARLI SIG1 Conference. Munchen.
Boud, D., & Falchikov, N. (2007). Developing assessment for informing judgement. In D. Boud & N. Falchikov (Eds.), *Rethinking assessment in higher education* (pp. 181–197). London: Routledge.
Boud, D., & Molloy, E. (Eds.). (2013). *Feedback in higher and professional education*. London: Routledge.
Boud, D., & Soler, R. (2016). Sustainable assessment revisited. *Assessment & Evaluation in Higher Education, 41*(3), 400–413. https://doi.org/10.1080/02602938.2015.1018133.
Boud, D., Ajjawi, R., Dawson, P., & Tai, J. (Eds.). (2018a). *Developing evaluative judgement in higher education. Assessment for knowing and producing quality work*. London: Routledge.
Boud, D., Dawson, P., Tai, J., & Ajjawi, R. (2018b). Creating an agenda for developing students' evaluative judgement. In D. Boud, R. Ajjawi, P. Dawson, & J. Tai (Eds.), *Developing evaluative judgement in higher education: assessment for knowing and producing quality work* (pp. 186–195). London: Routledge.
Brabrand, C., & Dahl, B. (2009). Using the SOLO taxonomy to analyze competence progression of university science curricula. *Higher Education, 58*(4), 531–549. https://doi.org/10.1007/s10734-009-9210-4.
Carless, D. (2009). Trust, distrust and their impact on assessment reform. *Assessment & Evaluation in Higher Education, 34*(1), 79–89. https://doi.org/10.1080/02602930801895786.
Carless, D. (2013). Trust and its role in facilitating dialogic feedback. In D. Boud & E. Molloy (Eds.), *Feedback in higher and professional education* (pp. 90–103). London: Routledge.
Carless, D. (2015). Exploring learning-oriented assessment processes. *Higher Education, 69*(6), 963–976. https://doi.org/10.1007/s10734-014-9816-z.
Carless, D., Joughin, G., & Liu, N.-F. (2006). *How assessment supports learning: learning-oriented assessment in action*. Hong Kong: Hong Kong University Press.
Carless, D., Bridges, S. M., Chan, C. K. Y., & Glofcheski, R. (Eds.). (2017). *Scaling up assessment for learning in higher education*. Singapore: Springer.
Cartney, P. (2012). Exploring the use of peer assessment as a vehicle for closing the gap between feedback and feedback used. In S. Hatzipanagos & R. Rocho (Eds.), *Approaches to assessment that enhance learning in higher education* (pp. 61–72). London: Routledge.
Coombs, A., DeLuca, C., LaPointe-McEwan, D., & Chalas, A. (2018). Changing approaches to classroom assessment: an empirical study across teacher career stages. *Teaching and Teacher Education, 71*, 134–144. https://doi.org/10.1016/j.tate.2017.12.010.
Cowan, J. (2010). Developing the ability for making evaluative judgements. *Teaching in Higher Education, 15*(3), 323–334.
Creswell, J. W., & Clark, V. P. (2010). *Designing and conducting mixed methods research* (2nd ed.). Thousands Oaks: SAGE Publications.
Dann, R. (2014). Assessment as learning: blurring the boundaries of assessment and learning for theory, policy and practice. *Assessment in Education: Principles, Policy & Practice, 21*(2), 149–166. https://doi.org/10.1080/0969594X.2014.898128.
Dawson, P., Henderson, M., Mahoney, P., Phillips, M., Ryan, T., Boud, D., & Molloy, E. (2018). What makes for effective feedback: staff and student perspectives. *Assessment & Evaluation in Higher Education, 2938*, 1–12. https://doi.org/10.1080/02602938.2018.1467877.
DeLuca, C., Valiquette, A., Coombs, A., LaPointe-McEwan, D., & Luhanga, U. (2016). Teachers' approaches to classroom assessment: a large-scale survey. *Assessment in Education: Principles, Policy & Practice, 25*(4), 355–375. https://doi.org/10.1080/0969594X.2016.1244514.
Dochy, F., Segers, M., & Sluijsmans, D. (1999). The use of self-, peer and co-assessment in higher education: a review. *Studies in Higher Education, 24*, 331–350.
Earl, L. (2013). *Assessment as learning. Using classroom assessment to maximize student learning*. London: Sage.
Espasa, A., & Meneses, J. (2010). Analysing feedback processes in an online teaching and learning environment: an exploratory study. *Higher Education, 59*(3), 277–292. https://doi.org/10.1007/s10734-009-9247-4.
Evans, C. (2013). Making sense of assessment feedback in higher education. *Review of Educational Research, 83*(1), 70–120. https://doi.org/10.3102/0034654312474350.
Falchikov, N. (2005). *Improving assessment through student involvement. Practical solutions for aiding learning in higher education and further education*. London: RoutledgeFalmer.
Garson, G. D. (2016). *Partial least squares: regression & structural equation models*. Asheboro: Statistical Publishing Associates.
Gielen, S., Dochy, F., & Onghena, P. (2011). An inventory of peer assessment diversity. *Assessment & Evaluation in Higher Education, 36*(2), 137–155.
Hair, J. F., Hult, G. T. M., Ringle, C. M., & Sarstedt, M. (2017). *A primer on partial least squares structural equation modeling (PLS_SEM)*. London: Sage.

Hair, J. F., Sarstedt, M., Ringle, C. M., & Gudergan, S. P. (2018). *Advanced issues in partial least squares structural equation modeling*. London: Sage.

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77*(1), 81–112.

Hawe, E., & Dixon, H. (2017). Assessment for learning: a catalyst for student self-regulation. *Assessment & Evaluation in Higher Education, 42*(8), 1181–1192. https://doi.org/10.1080/02602938.2016.1236360.

Hernández, R. (2012). Does continuous assessment in higher education support student learning? *Higher Education, 64*(4), 489–502. https://doi.org/10.1007/s10734-012-9506-7.

Hounsell, D. (2007). Towards more sustainable feedback to students. In D. Boud & N. Falchikov (Eds.), *Rethinking assessment in higher education: Learning for the longer term* (pp. 101–113). London: Routledge.

Ibarra-Sáiz, M. S., & Rodríguez-Gómez, G. (2017). EvalCOMIX®: a web-based programme to support collaboration in assessment. In T. Issa, P. Kommers, T. Issa, P. Isaías, & T. B. Issa (Eds.), *Smart technology applications in business environments* (pp. 249–275). Hershey: IGI Global. https://doi.org/10.4018/978-1-5225-2492-2.ch012.

Järvelä, S., Järvenoja, H., Malmberg, J., Isohätälä, J., & Sobocinski, M. (2016). How do types of interaction and phases of self-regulated learning set a stage for collaborative engagement? *Learning and Instruction, 43*, 39–51. https://doi.org/10.1016/j.learninstruc.2016.01.005.

Johnson, R. L., & Morgan, G. B. (2016). *Survey scales. A guide to development, analysis, and reporting*. London: The Guilford Press.

Klinger, D., McDivitt, P., Howard, B., Rogers, T., Muñoz, M., & Wylie, C. (2015). *Classroom assessment standards: practices for PK-12 teachers*. Seattle: Kindle Direct Press.

Lai, P. (2006). Review of assessing for learning: building a sustainable commitment across the institution. *Assessment & Evaluation in Higher Education, 31*(4), 479–481. https://doi.org/10.1080/02602930600679258.

Litwin, M. S. (2003). *How to assess and interpret survey psychometrics*. Thousands Oaks: SAGE Publications.

López-Pastor, V., & Sicilia-Camacho, A. (2017). Formative and shared assessment in higher education. Lessons learned and challenges for the future. *Assessment & Evaluation in Higher Education, 42*(1), 77–97. https://doi.org/10.1080/02602938.2015.1083535.

Makel, M. C., & Plucker, J. A. (2014). Facts are more important than novelty: replication in the education sciences. *Educational Researcher, 43*(6), 304–316. https://doi.org/10.3102/0013189X14545513.

Moore, C., & Teather, S. (2013). Engaging students in peer review : feedback as learning. *Issues in Educational Research, 23*(2), 196–212.

Murillo, F. J., & Hidalgo, N. (2017). Students' conceptions about a fair assessment of their learning. *Studies in Educational Evaluation, 53*, 10–16. https://doi.org/10.1016/j.stueduc.2017.01.001.

Nguyen, T. T. H., & Walker, M. (2016). Sustainable assessment for lifelong learning. *Assessment & Evaluation in Higher Education, 41*(1), 97–111. https://doi.org/10.1080/02602938.2014.985632.

Nicol, D., Thomson, A., & Breslin, C. (2014). Rethinking feedback practices in higher education: a peer review perspective. *Assessment & Evaluation in Higher Education, 39*(1), 102–122. https://doi.org/10.1080/02602938.2013.795518.

Panadero, E. (2016). Is it safe? Social, interpersonal, and human effects of peer assessment: a review and future directions. In G. T. L. Brown & L. R. Harris (Eds.), *Handbook of social and human conditions in assessment* (pp. 247–266). New York: Routledge.

Panadero, E. (2017). A review of self-regulated learning: six models and four directions for research. *Frontiers in Psychology, 8*(APR), 1–28. https://doi.org/10.3389/fpsyg.2017.00422.

Panadero, E., Klug, J., & Järvelä, S. (2016). Third wave of measurement in the self-regulated learning field: when measurement and intervention come hand in hand. *Scandinavian Journal of Educational Research, 60*(6), 723–735. https://doi.org/10.1080/00313831.2015.1066436.

Panadero, E., Jonsson, A., & Botella, J. (2017). Effects of self-assessment on self-regulated learning and self-efficacy: four meta-analyses. *Educational Research Review, 22*, 74–98. https://doi.org/10.1016/j.edurev.2017.08.004.

Panadero, E., Andrade, H., & Brookhart, S. (2018). Fusing self-regulated learning and formative assessment: a roadmap of where we are, how we got here, and where we are going. *The Australian Educational Researcher, 45*(1), 13–31. https://doi.org/10.1007/s13384-018-0258-y.

Pardo, A. (2018). A feedback model for data-rich learning experiences. *Assessment & Evaluation in Higher Education, 43*(3), 428–438. https://doi.org/10.1080/02602938.2017.1356905.

Penuel, W. R., & Shepard, L. A. (2016). Social models of learning and assessment. In A. A. Rupp & J. P. Leighton (Eds.), *Handbook of cognition and assessment: frameworks, methodologies, ans applications*. Hoboke: Wiley.

Pereira, D., Niklasson, L., & Flores, M. A. (2017). Students' perceptions of assessment: a comparative analysis between Portugal and Sweden. *Higher Education, 73*(1), 153–173. https://doi.org/10.1007/s10734-016-0005-0.

Pintrich, P. R. (2000). The role of goal orientation in self-regulated learning. In M. Boekhaerts, P. R. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 451–502). San Diego: Academic Press.

Reinholz, D. (2016). The assessment cycle: a model for learning through peer assessment. *Assessment & Evaluation in Higher Education, 41*(2), 301–315. https://doi.org/10.1080/02602938.2015.1008982.

Ringle, C. M., Wende, S., & Becker, J.-M. (2015). *SmartPLS 3*. Bönningstedt: SmartPLS. http://www.smartpls.com.

Rodríguez-Gómez, G., & Ibarra-Sáiz, M. S. (2015). Assessment as learning and empowerment: towards sustainable learning in higher education. In M. Peris-Ortiz & J. M. Merigó Lindahl (Eds.), *Sustainable learning in higher education. Developing competencies for the global marketplace (pp. 1–20)*. Springer International Publishing. https://doi.org/10.1007/978-3-319-10804-9_1.

Rowe, A. D. (2017). Feelings about feedback: the role of emotions in assessment for learning. In D. Carless, S. M. Bridges, C. K. Y. Chan, & R. Glofcheski (Eds.), Scaling up assessment for learning in Higher Education (pp. 159–172). Singapore: Springer. https://doi.org/10.1007/978-981-10-3045-1_11

Rust, C., O'Donovan, B., & Price, M. (2005). A social constructivist assessment process model: how the research literature shows us this could be best practice. *Assessment & Evaluation in Higher Education, 30*(3), 231–240. https://doi.org/10.1080/02602930500063819.

Sadler, D. R. (2012). Beyond feedback: developing student capability in complex appraisal. In S. Hatzipanagos & R. Rochon (Eds.), *Approaches to assessment that enhance learning in higher education* (pp. 45–60). London: Routledge.

Sadler, D. R. (2016). Three in-course assessment reforms to improve higher education learning outcomes. *Assessment & Evaluation in Higher Education, 41*(7), 1081–1099. https://doi.org/10.1080/02602938.2015.1064858.

Sambell, K., McDowell, L., & Montgomery, C. (2013). *Assessment for learning in higher education*. London: Routledge.

Skjong, R., & Wentworth, B. (2000). *Expert judgement and risk perception*. Hovik: Det Norsrke Veritas. http://research.dnv.com/skj/Papers/SkjWen.pdf . Accessed on 02/09/2018.

Smith, E., & Coombe, K. (2006). Quality and qualms in the marking of university assignments by sessional staff: an exploratory study. *Higher Education, 51*(1), 45–69. https://doi.org/10.1007/s10734-004-6376-7.

Stone-Romero, E. F., & Rosopa, P. J. (2008). The relative validity of inferences about mediation as a function of research design characteristics. *Organizational Research Methods, 11*(2), 326–352. https://doi.org/10.1177/1094428107300342.

Strijbos, J., Engels, N., & Struyven, K. (2015). Criteria and standards of generic competences at bachelor degree level: a review study. *Educational Research Review, 14*, 18–32. https://doi.org/10.1016/j.edurev.2015.01.001.

Struyven, K., Dochy, F., & Janssens, S. (2005). Students' perceptions about evaluation and assessment in higher education: a review1. *Assessment & Evaluation in Higher Education, 30*(4), 325–341. https://doi.org/10.1080/02602930500099102.

Tai, J. H. M., Canny, B. J., Haines, T. P., & Molloy, E. K. (2016). The role of peer-assisted learning in building evaluative judgement: opportunities in clinical medical education. *Advances in Health Sciences Education, 21*(3), 659–676. https://doi.org/10.1007/s10459-015-9659-0.

Tai, J., Ajjawi, R., Boud, D., Dawson, P., & Panadero, E. (2018). Developing evaluative judgement: enabling students to make decisions about the quality of work. *Higher Education, 76*(3), 467–481. https://doi.org/10.1007/s10734-017-0220-3.

Thomas, G., Martin, D., & Pleasants, K. (2011). Using self-and peer-assessment to enhance students' future-learning in higher education. *Journal of University Teaching & Learning Practice, 8*(81).

Topping, K. J. (2010). Methodological quandaries in studying process and outcomes in peer assessment. *Learning and Instruction, 20*, 339–343.

Wilson, M., & Scalise, K. (2006). Assessment to improve learning in higher education: the BEAR Assessment System. *Higher Education, 52*(4), 635–663.

Yucel, R., Bird, F. L., Young, J., & Blanksby, T. (2014). The road to self-assessment: exemplar marking before peer review develops first-year students' capacity to judge the quality of a scientific report. *Assessment & Evaluation in Higher Education, 39*(8), 971–986. https://doi.org/10.1080/02602938.2014.880400.

Zlatkin-Troitschanskaia, O., Shavelson, R. J., & Kuhn, C. (2015). The international state of research on measurement of competency in higher education. *Studies in Higher Education, 40*(3), 393–411. https://doi.org/10.1080/03075079.2015.1004241.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.