Evaluating Assessment. Validation with PLS-SEM of ATAE Scale for the Analysis of Assessment Tasks

Evaluando la evaluación. Validación mediante PLS-SEM de la escala ATAE para el análisis de las tareas de evaluación

Ibarra-Sáiz, M.S. (iD) **, & Rodríguez-Gómez, G.** (iD)

UNESCO Chair on Evaluation and Assessment, Innovation and Excellence in Education. University of Cadiz (Spain)

Abstract

One of the essential functions of university teachers lies in the decision-making process regarding the various components included in assessment process design, where the quality of assessment tasks is a key aspect. This study presents both validation of an instrument for students to evaluate the assessment tasks and the model that upholds the relationships between constructs that characterise the assessment tasks. Working from a review of the literature, a theoretical model has been devised featuring the characteristics of the assessment tasks and the relationships between them. The *Analysis of the Assessment and Learning Tasks* questionnaire (ATAE) has been designed to check them, based on a formative measurement model. Using a cohort design, a total of 1,166 questionnaires were obtained, completed by students from the Business Administration and Management (BAM) and Finance and Accounting (F&A) degree courses. The measurement model and the structural model were evaluated by means of the *Partial Least Squares Structural Equation Modeling* (PLS-SEM) technique using SmartPLS_3 software. The results show no collinearity problems plus high levels of absolute and relative importance for each questionnaire item. From the students' perception, it should be highlighted that the challenging aspect of an assessment task is related to transfer of learning, and that this is measured by use of communication strategies and demonstration of in-depth understanding.

Keywords: Assessment task; Assessment as learning; Empowerment; PLS-SEM; Partial Least Squares; Structural Equation Modeling, PLSpredict

Resumen

Una de las funciones esenciales del profesorado universitario se concreta en el proceso de toma de decisiones sobre los diferentes componentes que constituyen el diseño de los procesos de evaluación, siendo uno de sus elementos clave la calidad de las tareas de evaluación. En este estudio se presenta tanto la validación de un instrumento para la valoración por el estudiantado de las tareas de evaluación como el modelo que sustenta las relaciones entre los constructos que caracterizan las tareas de evaluación. A partir de una revisión de la literatura se ha elaborado un modelo teórico de las características de las tareas de evaluación y las relaciones existentes entre ellas. Para su comprobación se ha diseñado, sobre la base de un modelo de medida de carácter formativo, el cuestionario *Análisis de las Tareas de Evaluación y Aprendizaje* (ATAE). Mediante un diseño de cohorte se han obtenido un total de 1.166 cuestionarios cumplimentados por estudiantes de los grados de Administración y Dirección de Empresas (ADE) y Finanzas y Contabilidad (FYCO). La evaluación del modelo de medida y del modelo estructural se ha realizado mediante la técnica *Partial Least Squares Structural Equation Modeling* (PLS-SEM) utilizando el software SmartPLS_3. Los resultados muestran la no existencia de problemas de colinealidad y unos niveles elevados de importancia absoluta y relativa de cada uno de los ítems del cuestionario. Es de destacar, desde la percepción de los estudiantes, que el carácter retador de una tarea de evaluación se relaciona con la transferencia del aprendizaje, y cómo el uso de estrategias de comunicación y la demostración de una comprensión profunda son elementos mediadores de esta relación.

Palabras clave: Tarea de evaluación; Evaluación como aprendizaje; Empoderamiento; PLS-SEM; Mínimos cuadrados parciales; Modelo de ecuaciones estructurales; PLS predictivo

Corresponding author / Autor de contacto: **Ibarra-Sáiz, M.S.** University of Cadiz. Republic Saharaui Avenue; Campus Puerto Real, 11519, Puerto Real, Cádiz (Spain). marisol.ibarra@uca.es

Comments are frequently heard from students, often rather negative remarks, regarding the opportuneness, usefulness, or justice of an assessment. However, is this really the case when students tackle challenging assessment tasks? This initial question was used to commence a project that focuses on assessment task quality, university students' perception of these tasks and how university lecturers could improve their design.

One essential function for a lecturer, from an education planning perspective, lies in designing assessment processes, often requiring decision-making on many aspects (Bearman et al., 2014, 2016, 2017). This study focuses on just one of them, regarding the characteristics required for an assessment task to be considered good quality. These aspects have been analysed previously by authors such as Ashford-Rowe et al. (2014), Gore et al. (2009) or Smith & Smith (2014). In short, interest is focussed on the nature of the assessment tasks.

There are many studies regarding students' perception of assessment in a global sense (Wren et al., 2009); or focussing on specific aspects such as how often different means of assessment are used (Pereira et al., 2017). However, barely any research exists on students' experience and perception regarding the specific nature of the assessment tasks.

The study presented here, contextualised within a more global project, specifically proposes to provide an exploratory/predictive model and an instrument that helps to analyse and improve assessment task design and practice in the higher education environment. Specifically, this study aims to:

• Deliver a predictive model for learning transfer from the assessment tasks, by considering the relationships between the challenging nature of these tasks, their depth and the communication.

• Provide an instrument that can help us analyse and understand university students'

perception of the quality of the assessment tasks they perform.

Firstly, the theoretical foundations will be presented, alongside a predictive model that will determine the causal relationships between a set of variables that characterise the assessment tasks and, subsequently, the outcomes will be presented, obtained by corroborating this model from the perception of students from Business Administration and Management (BAM) and Finance and Accounting (F&A) degrees, eventually providing a series of theoretical and practical implications to improve the evaluation processes.

Conceptual framework and hypothesis development

Design and specification of assessment tasks is approached by Sadler (2016) as one of the three reforms required in the context of evaluating learning in higher education. For this author, assessment and grading of student performance imply making a deduction working from the student's products and actions and, logically, the quality of this deduction is determined by the quality of the data (the student's products and actions) and the assessor's skill. This work essentially falls within the second element proposed by Sadler (assessor skill) although considering that the assessor's role can be played both by the lecturers and by the students.

From the perspective of lecturers as assessors we refer to their skill to design and implement assessment tasks and, from the student's point of view, their role in performing these tasks and how they assess each one. Therefore, for the time being, this will not encompass the students' capacity to assess their own work, using self-assessment, or their classmates' work, by means of peer assessment, aspects that are essential in the students' evaluative judgement (Ibarra-Sáiz et al., 2020; Tai et al., 2017) and it will focus on their capability to evaluate not the lecturers' work in general, but the quality of a specific product of the lecturers' design activity such as assessment

tasks. By and large, lecturers design an assessment task and the student completes it, attempting to attain the standard determined by the lecturers and, once the task is finished, he/she feels the relief of having finished it but begins to experience uncertainty concerning the outcome. At this point, we should ask what it meant to the student to tackle this specific task, what assessment the actual task is worth, what value it holds for the student to have completed it, in short, what it represented as a learning experience. This thereby takes us to the challenge set by Dawson et al. (2013) to no longer study assessment practices from an abstract perspective but to attempt to understand how assessment principles can be used to improve learning outcomes and how they are perceived and valued by the students.

Assessment tasks as an essential part of the evaluation process

If we boil evaluation down to activities that must be performed or forced work for lecturers and students, then we are distorting the real meaning of disciplinary learning in higher education. We are failing by not recognising the relevance of curiosity, the importance of asking and answering pertinent and relevant questions. Assessment should be the meeting point where knowledge, ideas, differences of opinion, criticism and comprehension are generated and exchanged and this is the essential purpose of higher education (Sambell et al., 2013).

This importance given to the role played by assessment was highlighted by Biggs & Tang (2011) when, from their conception of constructive alignment, they proposed that their perceptions of the assessment would affect students' implication in the learning process; or by Pereira et al. (2017) who, more specifically, refer to influence from various means of assessment (presentations, reports, portfolio, projects, etc.) on how students learn.

Designing the assessment implies a decision-making process revolving around a series of elements that Bearman et al. (2014) specified in a global framework including the assessment proposals, assessment context,

feedback processes, learning outcomes, interactions and assessment tasks. We must be aware that these elements are not independent of each other but rather that they determine each other, so that certain assessment proposals or contexts will involve the prevalence of one type of assessment task or another, as highlighted by Ibarra-Sáiz & Rodríguez-Gómez (2019, p. 192) "simple, memorising or repetitive tasks cannot capture the complexity of realities and scenarios that require multiple, open solutions."

In short, as Sadler (2016, p. 1083) demonstrates, we should not confuse low quality evidence of the student's performance with evidence of low performance. Along this same line, Boud (2020) warn us about the importance of assessment tasks, insomuch as a poor choice of these tasks will lead to poor learning and will distort students' possible performance. A low-quality assessment task will provide us with weak information that is biased and unfair concerning the student's performance.

Characterisation of the assessment tasks

We can essentially say that an assessment task is an activity designed to gather information on students' capability to apply and use their competences, knowledge, abilities and skills to solve complex problems and be able to check how far the expected learning outcomes have been achieved.

Traditionally, university student learning has been assessed, above all, by the degree of comprehension of a specific field of knowledge focussed on the subject they are taking, thereby focussing on what students knew. Progressively, particularly from the 1990s onwards, the focus has changed and emphasis is being put on the value of transferable, generic skills or essential competences, skills that the student should develop irrespective of the specific discipline around which they wish to develop their future career (Boud, 2014; Strijbos et al., 2015).

This change in direction has implied renewing the means of assessment, changing

from classic tests, quizzes or final exams focused on reproducing knowledge, to a series of new assessment means (portfolio, simulations, case studies, etc.) in an attempt to integrate and give coherence to the learning which we would like to develop by aligning teaching and assessment, which has led certain authors to talk about a new assessment culture (Dochy, 2009). These new means of assessment focus on a student's performance, in terms of what he/she is capable of doing and producing, using critical thinking and creativity to solve complex and current problems.

Assessment task quality is a central axis for this new assessment culture, so that, for example, Sambell et al. (2013) refer to the emphasis on authentic, complex tasks as one of the six central principles of the evaluative approach called assessment for learning. Along this same line, Rodríguez-Gómez and Ibarra-Sáiz (2015) consider assessment tasks as one of the essential challenges that must be tackled from the approach of assessment as learning and empowerment

Assessment task quality must be analysed on the basis of three specific dimensions (Gore et al., 2009): intellectual rigour, meaning and support offered to the student. Intellectual rigour refers to focusing assessment tasks on producing an in-depth understanding of what is important, of the concepts, skills and essential ideas. It requires active construction and implication in high level thinking from students, as well as substantial communication regarding what they have learnt. An assessment task will be relevant to the extent that it helps make the learning more significant and important for the students, connecting it with the intellectual demands of their work. Consequently, assessment tasks require a clear connection with prior knowledge and with academic and extra-academic knowledge. Finally, the assessment task supports the student to the extent that it explicitly sets high expectations on the student's work.

Focussing on the design of assessment tasks for students who are starting their university

course, Thomas et al. (2019) consider that these tasks should facilitate learning for students, encourage students' implication in the learning and provide feedback for future learning, thereby being used not only to analyse the degree of completion, with varying accuracy, but to understand potential future improvement areas.

After analysing different contributions, Dochy (2009) reach the conclusion that new means of assessment maintain five core characteristics. The main characteristic is that a good assessment should require students to construct knowledge. It is not enough for students to faithfully reproduce the knowledge; they must have a good command of the structure and existing interrelations and give coherence to the knowledge. The second characteristic highlights the need to assess the ability to apply knowledge to current cases, which requires analysing how much the students apply the knowledge to problem solving in real life and also make appropriate decisions. The third characteristic is the contextual sensitivity and the multiplicity of perspectives. For Dochy (2009), the student does not just need to know "what" but also "when", "where" and "why". To do this, it is not advisable to use means of assessment based on statements and answers. The student must have a good command and understanding of the underlying causal mechanisms. The student's participation is highlighted by this author as the fourth characteristic of these new means of assessment, where the student plays an active role in debating and participating in the design or drawing up the assessment criteria, the assessment instruments or even acting as an assessor. Finally, the assessment is not something final or separate, but it is built into the learning process and it is consistent with the teaching methods and the learning environment.

As we have seen, there are many different aspects to be considered when designing an assessment task. Based on these prior contributions and other studies, Table 1 presents the four characteristics that we have considered in this research as essential for an

assessment task: represent a challenging stimulus, the need to demonstrate in-depth comprehension, the use of communication strategies and the capability to transfer what has been learnt when performing the task.

Table 1. Definition of constructs

Construct	Definition	References
Challenge	Tackle open, complex problems that require divergent thinking, creativity and forging significant relationships and connections.	Ashford-Rowe et. al., 2014; Dochy & Gijbels, 2006 Gore et al., 2009; Sambell et al., 2013
Depth	Demonstrate in-depth understanding by using investigation methods and reflective and critical thinking	Dochy, 2009; Entwistle & Karagiannopoulou, 2014; Herrington & Herrington, 2006; O'Donovan, 2016
Communication	Use oral, written and symbolic communication strategies by means of presentations, developments or products based on well-founded arguments.	Gore et al., 2009; Gulikers et. al, 2004; Smith & Smith, 2014
Transfer	Relate the knowledge and the experience to other subjects and to the social and professional reality	Ashwin et al. 2015; Glofcheski, 2017; Gulikers et al., 2004, 2006; Ibarra-Sáiz et al., 2020; Strijbos et al. 2015

Investigation model and hypothesis

The model baseline in this paper proposes that the students' perception of the capability to transfer knowledge from the assessment tasks that they tackle is determined by the task's depth and the required communication, and these two aspects are in turn determined by the challenging nature of the assessment task. Figure 1 presents this basic model indicating the relationships that are determined between these different constructs.

Figure 1. *Model to confirm the relationships between elements characterising the assessment tasks*
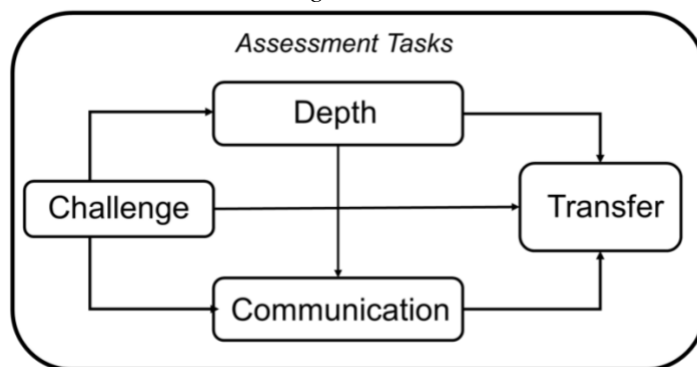
Working from this theoretical model and from the basis of contributions analysed previously in this study, the following hypotheses are proposed:

H1. The transfer is expected to be positively related to the challenge (H1a), the depth (H1b) and the communication (H1c).

H2. The challenge is expected to be positively related to the depth (H2a) and the communication (H2b).

H3. The depth is expected to be directly related to the communication.

H4. The relationship between the challenge and the transfer is expected to be mediated by the depth (H4a) and the communication (H4b).

Methodology

To carry out this study, a survey methodology was followed using a cohort design, since the students' perception was collected during four successive academic years, starting in the 2016/17 academic year and ending in the 2019/20 academic year, so, they were not the same subjects each year. On the contrary, different subjects answered the survey in each academic year.

A set of four assessment tasks were designed, the characteristics of which are described in the work of Ibarra-Sáiz et al. (2020). As students completed each assessment task they answered the *Analysis of Assessment and Learning Tasks* (ATAE) questionnaire expressing their assessment and experience in each case.

Participants

A total of 1,166 ATAE questionnaires were collected, completed by students of the School of Economics and Business Sciences, Cadiz University (Table 2).

Table 2. Demographic characteristics of the sample

		Frequency	**Percentage**
Gender	Male	555	47.6
	Female	611	52.4
Degree	F&A	142	12.2
	BAM	1024	87.8
Cohort	2017	361	31.0
	2018	369	31.6
	2019	240	20.6
	2020	196	16.8

These students were studying the subject Project Management, taught in the last year of the Business Administration and Management (BAM) and Finance and Accounting (F&A) degrees. Table 3 shows the distribution of the questionnaires completed by these students during the four years and for each of the four assessment tasks they addressed.

Table 3. Distribution of questionnaires by year and task

Task	Year				Total
	2017	**2018**	**2019**	**2020**	
1	93	110	75	64	342
2	89	50	67	55	261
3	98	104	51	41	294
4	81	105	47	36	269
Total	361	369	240	196	1,166

Instrument

The constructs and measurement indicators of the ATAE questionnaire were developed on the basis of a literature review and then validated by judges (Figure 2). Different methods used for content validation by expert judges were reviewed (Johnson & Morgan, 2016) and the group consensus method was chosen, as it avoids voting systems. The definition of the constructs was revised at the end of each cycle and the indicators were specified during the discussion process. Finally, in order to analyse the apparent validity, the questionnaire was presented to a group of Master's students, and it was possible to improve the questionnaire in terms of its clarity and ease of understanding.

Figure 2. Process of designing the ATAE questionnaire (Analysis of Assessment Tasks)
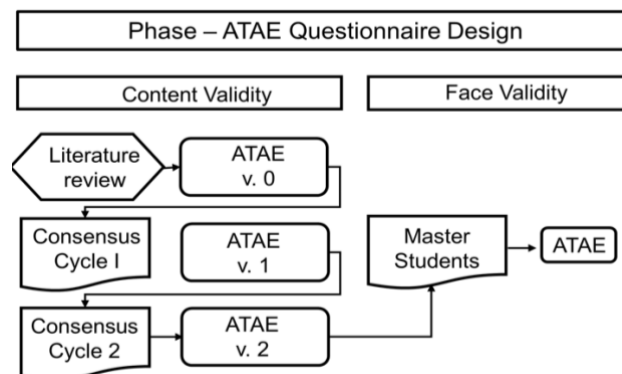
The ATAE questionnaire (Annex I) is structured in four dimensions (Table 4) and consists of 16 items in Likert-type scale format (0-10) distributed in each of the dimensions, and four open questions. Although the debate on the number of options and whether there should be an intermediate option is not closed (Horst & Pyburn, 2018), in this case, we have followed the recommendations of the OECD (2013) to maintain a numerical scale from 0 to 10. The questionnaire took about 10 minutes to complete.

Table 4. Structure of the ATAE questionnaire

Dimensions	# Items
Depth	4
Communication	4
Challenge	3
Transfer	5

Data analysis

The PLS-SEM method (Hair et al., 2017) and PLSpredict (Shmueli et al., 2016) were used to estimate the model. The software SmartPLS 3 was used for carrying out calculations (Ringle et. al., 2015). PLS-SEM is a multivariate analysis approach used to estimate models with latent variables. It is a recommended technique when, as in this study, the aim is the prediction of an objective construct or it is intended to identify relevant constructs, the research model is complex according to the type of relations that are hypothesized (direct and mediation), the constructs that are part of the structural model have been designed following a formative measurement model, the structural model is complex and the data do not follow a normality distribution (Roldán & Sánchez-Franco, 2012; Hair et al., 2016; Jiménez-Cortés, 2019).

To test the adequacy of the measurement model, the Confirmatory Tetrad Analysis (CTA-PLS) has been used. Using this technique, the null hypothesis that the indicators in a model are reflective can be tested (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 its formative nature, for which a multicollinearity analysis and a weighting analysis have been performed. Subsequently, the predictive capacity of the model and the relationships among the constructs have been analysed. The following analyses have been carried out: a) collinearity assessment (VIF); b) structural model Path coefficients; c) determination coefficient (R_2); d) effect size (f_2); e) predictive relevance (Q_2), f) effect size (q_2) and g) predictive power analysis using PLSpredict (Shmueli et al., 2016, 2019).

Results

Evaluation of the measurement model

When analysing reflective indicators it is usual to analyse internal consistency (Cronbach's alpha), convergent validity or discriminant validity (Muñoz-Cantero et al., 2019), but when formative indicators are used, the evaluation of the measurement model is based on the analysis of collinearity, relative importance (external weights) and absolute importance (external loads) (Hair et al., 2019). In our case (Table 5), with co-linearity values (VIF) between 1.18 and 2.11 we can conclude that, taking as reference value 5, collinearity does not reach critical levels in any of the training constructs and, therefore, there is no difficulty in estimating the model. An indicator (COM_3) was found whose weight was not statistically significant but instead had a load close to 0.5, so according to the decision rules expressed by Hair et al. (2016) and the content of the indicator itself, it was decided to maintain all the formative indicators.

In addition, following the orientations of Hair et al. (2019) to check the robustness of the measurement model, a Confirmatory Tetrad Analysis (CTA-PLS) was performed, which has allowed to empirically check the formative character of the RET and TRA constructs, as tetrads significantly different from zero were found.

Table 5. Weights, loads and VIF values of the formative constructs

Constructs	Indicators	Weights	Loads	VIF
CHALLENGE (RET)	RET_1	0.373	0.832	1.700
	RET_2	0.450	0.842	1.590
	RET_3	0.090	0.598	1.447
	RET_4	0.343	0.749	1.518
DEPTH (PRO)	PRO_1	0.269	0.738	1.485
	PRO_2	0.350	0.834	1.793
	PRO_3	0.311	0.814	1.770
	PRO_4	0.323	0.795	1.607
COMMUNICATION (COM)	COM_1	0.416	0.793	1.436
	COM_2	0.669	0.909	1.335
	COM_3	0.133	0.477	1.183
TRANSFER (TRA)	TRA_1	0.312	0.827	2.013
	TRA_2	0.303	0.834	2.110
	TRA_3	0.129	0.666	1.712
	TRA_4	0.239	0.712	1.667
	TRA_5	0.320	0.728	1.363

Evaluation of the structural model

To proceed with the evaluation of the structural model, the following analyses were carried out: a) collinearity; b) significance and relevance of structural relationships; c) predictive power and relevance; d) effect size; and e) predictive power.

With regard to collinearity problems, in the results presented in Table 6 we can see that all VIF values are clearly below the limit of 5, so we can conclude that there are no collinearity problems.

Table 6. VIF values of the structural model

	COM	PRO	RET	TRA
COM				2.610
PRO	2.469			2.900
RET	2.469	1.000		2.947
TRA				

The predictive value of the model has been analysed through the determination coefficient (R_2). Thus, Figure 3 shows how almost 70% of the variance (R_2) of the transfer construct (TRA) is explained by the three other constructs. According to the criteria established by Chin (1998) and Hair et al. (2017) we can consider it substantial. The strongest effect on transfer (TRA) is exerted by the depth construct (PRO, 0.435), followed by the challenge construct (RET, 0.325) and communication (COM, 0.146). Likewise, it is evident that the R_2 values for the PRO (0.595) and COM (0.617) constructs reach levels that can be considered as moderate $(R_2>0.50)$. The model has an SRMR of 0.03, which indicates an adequate level of adjustment taking as a reference the usual criterion of placing it below 0.08.

Figure 3. Results of the structural model
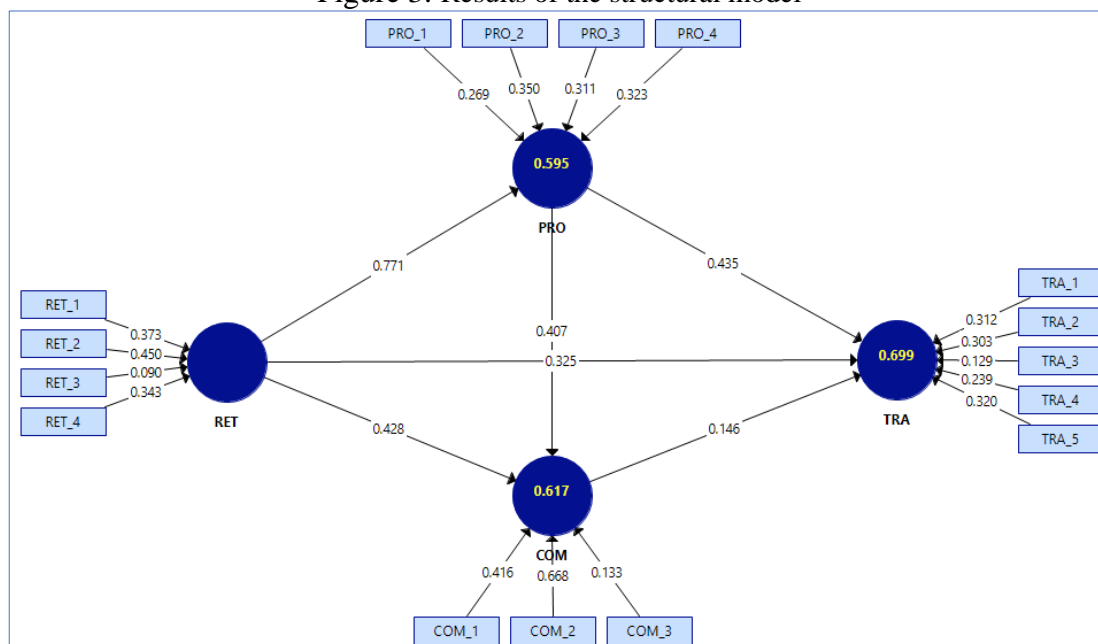
To establish the statistical significance of the path coefficients, according to Hair et al. (2017) a bootstrapping with 5,000 subsamples was performed in order to generate the *t*-statistics and confidence intervals (Table 7).

We observe large effect sizes in the case of the RET->PRO relationship, being medium in the PRO->TRA, RET->COM and PRO->COM relationships and small in the RET->TRA and COM->TRA case.

Table 7. Structural model results using t-values and percentiles with 95% confidence interval (n=5,000 subsamples).

Relationships	Path coefficients				Effect size				Hypo-thesis
	Path(*)	95% CI	*t*	*p*	*f2*(+)	95% CI	*t*	*p*	
RET -> TRA	0.325	[0.252, 0.390]	9.006	0.000	0.119	[0.071, 0.180]	4.347	0.000	H1a
PRO -> TRA	0.435	[0.365, 0.515]	11.660	0.000	0.217	[0.142, 0.312]	4.949	0.000	H1b
COM -> TRA	0.146	[0.077, 0.217]	4.092	0.000	0.027	[0.008, 0.061]	1.977	0.048	H1c
RET -> PRO	0.771	[0.734, 0.808]	41.047	0.000	1.469	[1.179, 1.862]	8.311	0.000	H2a
RET -> COM	0.428	[0.703, 0.779]	37.986	0.000	0.194	[0.130, 0.270]	5.414	0.000	H2b
PRO -> COM	0.407	[0.338, 0.475]	11.763	0.000	0.175	[0.117, 0.247]	5.308	0.000	H3

Notes: (*) 0.75 substantial/0.50 moderate/0.25 weak / (+) 0.35 large/0.15 medium /0.02 small

To check the predictive relevance of the model, the Q_2 values have been calculated by means of the blindfolding procedure (Table 8). We can observe that all values for endogenous constructs are above zero. More specifically, the highest value is presented by TRA (0.396), followed by PRO (0.373) and COM (0.341). These results support the relevance of the predictive model based on endogenous latent variables.

Table 8. Cross-validated construct redundancy (Q_2 values)

	SSO	SSE	Q^2 (=1-SSE/SSO)
COM	3498.000	2306.506	0.341
PRO	4664.000	2922.958	0.373
RET	4664.000	4664.000	
TRA	5830.000	3518.702	0.396

The effect sizes (q_2) allow an evaluation of how an exogenous construct contributes to an endogenous latent construct. In our case we found that small values are achieved in the effect size (Table 9).

Table 9. Effect sizes (q_2)

	COM	PRO	TRA
COM		-0.001	0.002
PRO	0.051		0.069
RET	0.066		0.034
TRA	-0,001	0.000	

Notes: * 0.02 small/0.15 medium/0.35 large

Finally, to test the predictive power of the model, the PLSpredict procedure (Sharma et al., 2018) was used, obtaining the results presented in Table 10. It is evident that, in all cases, the $Q_{2predict}$ values are above zero and in half of the indicators higher RMSE values are obtained using PLS versus LM, which indicates that the model has an medium predictive power (Shmueli et al., 2019; Hair et al., 2019).

Table 10. Summary of prediction of manifest variables (indicators)

Indicators	PLS		LM		PLS-LM (RMSE)
	RMSE	Q^2_predict	RMSE	Q^2_predict	
COM_1	1.275	0.375	1.274	0.376	0.001
COM_2	1.061	0.434	1.057	0.438	0.004
COM_3	1.993	0.109	2.000	0.103	-0.007
PRO_4	1.289	0.369	1.287	0.370	0.002
PRO_1	1.361	0.355	1.363	0.353	-0.002
PRO_2	1.176	0.409	1.177	0.407	-0.001
PRO_3	1.242	0.365	1.243	0.363	-0,002
TRA_1	1.135	0.392	1.137	0.388	-0.002
TRA_2	1.152	0.426	1.155	0.423	-0.003
TRA_3	1.582	0.252	1.567	0.267	0.015
TRA_4	1.496	0.314	1.494	0.317	0.002
TRA_5	1.329	0.295	1.320	0.303	0.009

Mediation analysis

In analysing the mediation, following the guidelines proposed by Zhao et al. (2010) and Nitzl et al. (2016) a two-step process has been used. The first step is to determine the significance of the indirect effects by means of a bootstrapping procedure and secondly to establish the type of mediation following the decision tree proposed by Zhao et al. (2010) and updated by Hair et al. (2017).

a) Depth and communication as mediating variables

In the case of the model we have presented (Fig. 3), depth and communication operate as mediating variables between the challenging nature of assessment tasks and the transfer of learning, so we can say that this is a model of multiple mediation. Table 11 shows the results obtained when checking the effect of this mediation. Thus, we can see that the challenge has a significant *(t=9.006, p<0.05)* direct effect (0.325) and that the total indirect effect of the challenge-transfer relationships (0.444) is also significant (*t*=15.272, *p*<0.05) and, in both cases, the confidence interval does not include zero. We observe how depth (0.336) presents a significant specific indirect effect (*t*=11.378, *p*<.05) and, although to a lesser extent, the effect of communication (0.063) is also significant (*t*=3,914, *p*<.05), as is the multiple indirect effect of depth and communication (0.046).

To analyse the magnitude of mediation, the explained variance index (VAF) has been calculated according to the orientations of

Cepeda-Carrión et al. (2017). The conclusion is that the greatest power of mediation (75.6%) is exercised by depth, followed by communication (14.1%) and the interaction between depth and communication (10.3%).

Since it is always the product of the direct effect and the specific indirect positive one, we can conclude that it is a complementary mediation and thus, as pointed out by Zhao et al. (2010), these mediating variables are consistent with the hypothesized theoretical model, although there may be other mediators not contemplated that could complete this mode.

Table 11. Summary of the checking of the mediation effect of RET on TRA

Total effect	Effect	95% CI	t	p	Type of mediation
RET -> TRA	0.769	[0.733, 0.805]	42.047	0.000	
Direct effect					
RET -> TRA	0.325	[252, 0.390]	9.006	0.000	
Total indirect effect					
RET -> TRA	0.444	[0.389, 0.503]	15.272	0.000	
Specific indirect effects					
RET -> PRO -> COM -> TRA	0.046	[0.023, 0.072]	3.662	0,000	Complementary
RET -> PRO -> TRA	0.336	[0.277, 0.393]	11.378	0.000	Complementary
RET -> COM -> TRA	0.063	[0.033, 0.095]	3.914	0.000	Complementary

Discussion

Initially, this study aims to provide a predictive model for learning transfer, based on the variables that characterise the nature of the assessment tasks and, secondly, offer a useful instrument to analyse and understand university students' perception of the quality of the assessment tasks that they face in their learning process. These results suggest implications both from a theoretical and a practical perspective to understand design and implementation of assessment tasks in university classrooms, whilst making it possible to discern future lines of research.

Theoretical implications

As we have mentioned, the main aim was to provide a predictive model regarding the parts of the nature of assessment tasks, contextualising this study within research on characterisation and nature of assessment tasks.

One of this study's main contributions revolves around confirmation of a model that integrates the relationship between a set of variables which characterise assessment task quality. The results that have been updated can, to a large extent, predict the relationships determined between the outstanding variables and demonstrate, firstly, that the challenging aspect of the assessment tasks is directly related to the transfer of knowledge, the depth and the communication and, on the other hand, the mediation role played by depth and communication.

The hypothesis has been corroborated which affirms the direct relationship of the transfer with the challenging nature of the assessment task (H1a), and with the depth (H1b) and the communication (H1c). In addition, the hypothesis has been corroborated that relates the challenging nature of the tasks with the depth (H2a) and the communication (H2b) and that relates the relationship between depth and communication (H3). Finally, the hypothesis has been corroborated that determines the mediation of the depth (H4a) and the communication (H4b) in the relationship between the challenging nature of the tasks and the learning transfer.

The challenging nature of the assessment tasks, the need to use high level thinking to solve them by means of demonstrating in-depth comprehension or using communication are all characteristics that determine the quality of the assessment tasks and their capacity to transfer the learning to other contexts or situations (Ashford-Rowe et al., 2014; Sambell et al., 2013; Smith & Smith, 2014). This study demonstrates the relevant role of the challenging nature of assessment tasks and the need to get students to tackle complex problems and motivate them to solve these problems and how this affects the learning transfer when solving these tasks.

Practical implications

A second objective driving this study was to provide an instrument to analyse and understand university students' perception of assessment task quality. The measurement model estimation supports the validity of the ATAE questionnaire used to operationalize the latent variables, as the items are relevant and load on the correct construct. Consequently, we have an instrument that is easy to apply after finishing an assessment task through which it is possible to quickly gather the students' assessment of this and, in turn, it is used for critical reflection by the students themselves regarding the usefulness of the actual assessment task.

Limitations and prospective

From a methodological perspective, this research is affected by a series of limitations that, in turn, become chances to make improvements and explore new lines of research. Firstly, this research has taken place in the context of Economic and Business Sciences degree studies, more specifically with final year students. In addition, the assessment tasks that these students had to tackle were designed from an evaluative approach based on assessment as learning and empowerment (Rodríguez-Gómez & Ibarra-Sáiz, 2015), all reasons that lead to weakening the chances of generalising any findings. Therefore, one future line of research revolves around

carrying out other studies that, on the one hand, allow extrapolation to other different contexts both in the field of social sciences and in other areas of knowledge and, on the other hand, compare evaluations made concerning assessment tasks designed by lecturers from different evaluative approaches, which would make it possible to corroborate the assessments made by the students among very different types of assessment tasks.

Secondly, the measurement instrument hereby presented is based on the students' assessment from a completely subjective perspective, so it could be improved by using instruments that facilitate alternative indicators (Panadero et al., 2018), or by incorporating sources of information other than the actual student, combining measuring and intervention (Panadero et al., 2016).

Conclusion

In this research we have presented outcomes that demonstrate how the challenging nature of the assessment tasks, or their requirement to implement high level knowledge using communication strategies are characteristics that encourage learning transfer in assessment processes. An instrument has been provided that can be adapted or replicated in other contexts and new lines of work have been proposed that will improve comprehension of the nature of assessment tasks.

This study asserts the importance of the university lecturer's role as a designer of learning experiences and, specifically, as the designer of high-quality assessment tasks that require putting all the student's potential into play. Nevertheless, this is not an easy role, as demonstrated by research from Bearman et al. (2016, 2017), and requires thinking about the rationality and justification of the assessment, the activities that are going to be used to be reviewed or scored, the criteria that are going to be used to assess whether the chosen learning outcomes have been achieved, what the actual student can bring to the assessment process or the possible time frame for the different assessment tasks over the academic

year. It is a decision-making process for the lecturers that, to a large extent, is made more difficult due to contextual limitations that require the lecturers under evaluation to take more, specific training and for university policies to be developed that encourage new assessment means.

Designing challenging assessment tasks that require the use of in-depth knowledge of the subject matter, command of communication strategies and, consequently, that the learning that takes place when solving the task can be transferred to other contexts, requires lecturers to have a good command of new means of assessment and their contextualised use. Such means of assessment would no longer have to be evaluated using former reliability and validity criteria, but new alternative criteria (Dochy, 2009) such as transparency, justice, direct (not inferred) appreciation or its authenticity and cognitive complexity that, obviously, require lecturers and students to embrace a new vision and perspective regarding the assessment processes.

Acknowledgement

References

Ashford-Rowe, K., Herrington, J., & Brown, C. (2014). Establishing the critical elements that determine authentic assessment. *Assessment & Evaluation in Higher Education*, *39*(2), 205–222. https://doi.org/10.1080/02602938.2013.819566

Ashwin, P., Boud, D., Coate, K., Hallet, F., Keane, E., Krause, K.-L., … Tooher, M. (2015). *Reflective teaching in higher education*. Bloomsbury. https://doi.org/10.1142/S0129183114500405

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

Bearman, M., Dawson, P., Boud, D., Bennett, S., Hall, M., & Molloy, E. (2016). Support for assessment practice: developing the Assessment Design Decisions Framework. *Teaching in Higher Education*, *21*(5), 545–556. https://doi.org/10.1080/13562517.2016.1160217

Bearman, M., Dawson, P., Boud, D., Hall, M., Bennett, S., Molloy, E., & Joughin, G. (2014). *Guide to the Assessment Design Decisions Framework*. Retrieved from http://www.assessmentdecisions.org/guide/

Biggs, J., & Tang, C. (2011). *Teaching for quality learning at university. What the students does* (4th ed.). McGraw-Hill-SRHE & Open University Press.

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 University Press Ltd. https://doi.org/10.3366/edinburgh/9780748694549.003.0002

Boud, D. (2020, en prensa). Challenges in reforming higher education assessment: a perspective from afar. *RELIEVE - Electronic Journal of Educational Research, Assessment and Evaluation*.

Cepeda-Carrión, G., Nitzl, C., & Roldán, J. L. (2017). Mediation Analyses in Partial Least Squares Structural Equation Modeling: Guidelines and Empirical Examples. In H. Latan & R. Noonan (Eds.), *Partial Least Squares Path Modeling: Basic concepts, methodological issues and applications* (pp. 173–195). Springer. https://doi.org/10.1007/978-3-319-64069-3

Chin, W. W. (1998). The partial least squares approach to structural equation modeling. In

G. A. Marcoulides (Ed.), *Modern methods for business research* (pp. 295–398). Lawrence Erlbaum.

Cubero-Ibáñez, J., & Ponce-González, N. (2020). Aprendiendo a través de tareas de evaluación auténticas: Percepción de estudiantes de Grado en Educación Infantil. *Revista Iberoamericana de Evaluación Educativa*, *13*(1), 41–69. https://doi.org/10.15366/riee2020.13.1.002

Dawson, P., Bearman, M., Boud, D. J., Hall, M., Molloy, E. K., Bennett, S., & Joughin, G. (2013). Assessment Might Dictate the Curriculum, But What Dictates Assessment? *Teaching & Learning Inquiry: The ISSOTL Journal*, *1*(1), 107–111. https://doi.org/10.2979/teachlearninqu.1.1.107

Dochy, F. (2009). The Edumetric Quality of New Modes of Assessment: Some Issues and Prospect. In G. Joughin (Ed.), *Assessment, Learning and Judgement in Higher Education* (pp. 85–114). Springer. https://doi.org/10.1007/978-1-4020-8905-3_6

Dochy, F., & Gijbels, D. (2006). New learning, assessment engineering and edumetrics. In L. Verschaffel, F. Dochy, M. Boekaerts, & S. Vosniadou (Eds.), *Instructional Psychology: Past, present and future trends. Sixteen essays in honour of Erik De Corte*. Elsevier.

Entwistle, N., & Karagiannopoulou, E. (2014). Perceptions of Assessment and their Influences on Learning. In C. Kreber, C. Anderson, N. Entwistle, & J. McArthut (Eds.), *Advances and innovations in university assessment and feedback* (pp. 75–98). Edinburgh University Press Ltd. https://doi.org/10.3366/edinburgh/9780748694549.003.0005

Garson, G. D. (2016). *Partial Least Squares: Regression & Structural Equation Models*. Statistical Publishing Associates.

Glofcheski, R. (2017). Making Assessment for Learning Happen Through Assessment Task Design in the Law Curriculum. In D. Carless, S. M. Bridges, C. K. Y. Chan, & R. Glofcheski (Eds.), *Scaling up Assessment for Learning in Higher Education* (pp. 67–80). Singapore. https://doi.org/10.1007/978-981-10-3045-1_5

Gore, J., Ladwig, J., Eslworth, W., & Ellis, H. (2009). *Quality assessment framework: A guide for assessment practice in higher education*. Callaghan, NSW Australia: The Australian Learning and Teaching Council. The University of Newcastle. Retrieved from http://www.olt.gov.au/system/files/resources/QAF FINAL doc for print.pdf

Gulikers, J. T. M., Bastiaens, T. J., Kischner, P. A., & Kester, L. (2006). Relations between student perception of assessment authenticity, study approaches and learning outcomes. *Studies in Educational Evaluation*, *32*, 381–400. https://doi.org/10.1016/j.stueduc.2006.10.003

Gulikers, J. T. M., Bastiaens, T., & Kirschner, P. A. (2004). A five-dimensional framework for authentic assessment. *Educational Technology Research and Development*, *52*(3), 67–85. https://doi.org/10.1007/BF02504676

Hair, J. F., Hult, G. T. M., Ringle, C. M., & Sarstedt, M. (2017). *A Primer on Partial Least Squares Structural Equation Modeling (PLS-SEM)* (2nd ed.). Sage.

Hair, J. F., Risher, J. J., Sarstedt, M., & Ringle, C. M. (2019). When to use and how to report the results of PLS-SEM. *European Business Review*, *31*(1), 2–24. https://doi.org/10.1108/EBR-11-2018-0203

Hair, J. F., Sarstedt, M., Ringle, C. M., & Gudergan, S. P. (2018). *Advanced Issues in Partial Least Squares Structural Equation Modeling*. Sage.

Herrington, J., & Herrington, A. (2006). Authentic conditions for authentic assessment: Aligning task and assessment. *Proceedings of the 29th HERDSA Annual Conference*, 146–151.

Horst, S. J., & Pyburn, E. M. (2018). Likert scaling. In B. B. Frey (Ed.), *The SAGE encyclopedia of educational research, measurement, and evaluation* (pp. 974–976). Sage.

Ibarra-Sáiz, M. S., & Rodríguez-Gómez, G. (2019). Una evaluación como aprendizaje [An Assessment as Learning]. In J. Paricio, A. Fernández, & I. Fernández (Eds.), *Cartografía de la buena docencia. Un marco para el desarrollo del profesorado basado en la investigación* (pp. 175–196). Narcea.

Ibarra-Sáiz, M. S., Rodríguez-Gómez, G., & Boud, D. (2020). Developing student competence through peer assessment: the role of feedback, self-regulation and evaluative judgement. *Higher Education*. https://doi.org/10.1007/s10734-019-00469-2

Jiménez-Cortés, R. (2019). Aprendizaje de las mujeres en las redes sociales: Validación de la escala MAIA con PLS. *Revista de Investigación Educativa*, *37*(2), 431–449. https://doi.org/http://dx.doi.org/10.6018/rie.37.2.325721

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

Muñoz-Cantero, J.-M., Rebollo-Quintela, N., Mosteiro-García, J., & Ocampo-Gómez, C.-I. (2019). Validación del cuestionario de atribuciones para la detección de coincidencias en trabajos académicos. *RELIEVE - Revista Electrónica de Investigación y Evaluación Educativa*, *25*(1), 1–16. https://doi.org/10.7203/relieve.25.1.13599

Nitzl, C., Roldán, J. L., & Cepeda, G. (2016). Mediation Analysis in Partial Least Squares Path Modeling: Helping researchers discuss more sophisticated models. *Industrial Management & Data Systems*, *116*(9), 1849–1864. https://doi.org/10.1108/IMDS-07-2015-0302

O'Donovan, B. (2016). How student beliefs about knowledge and knowing influence their satisfaction with assessment and feedback. *Higher Education*, *74*(4), 617–633. https://doi.org/10.1007/s10734-016-0068-y

OECD. (2013). Methodological considerations in the measurement of subjective well-being. In *OECD Guidelines on measuring subjective well-being* (pp. 61–138). Paris: OECD Publishing. https://doi.org/10.1787/9789264191655-en

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

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

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

Ringle, C. M., Wende, S., & Becker, J.-M. (2015). *SmartPLS 3*. Bönningstedt: SmartPLS. Retrieved from 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. https://doi.org/10.1007/978-3-319-10804-9_1

Roldán, J. L., & Sánchez-Franco, M. J. (2012). Variance-Based Structural Equation Modeling: Guidelines for Using Partial Least Squares in Information Systems Research. In *Research Methodologies, Innovations and Philosophies in Software Systems Engineering and Information Systems* (pp. 193–221). IGI Global. https://doi.org/10.4018/978-1-4666-0179-6.ch010

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*. Routledge.

Sharma, P. N., Shmueli, G., Sarstedt, M., Danks, N., & Ray, S. (2018). Prediction-Oriented Model Selection in Partial Least Squares Path Modeling. *Decision Sciences*. https://doi.org/10.1111/deci.12329

Shmueli, G., Ray, S., Velasquez Estrada, J. M., & Chatla, S. B. (2016). The elephant in the room: Predictive performance of PLS models. *Journal of Business Research*, *69*(10), 4552–4564. https://doi.org/10.1016/j.jbusres.2016.03.049

Shmueli, G., Sarstedt, M., Hair, J. F., Cheah, J. H., Ting, H., Vaithilingam, S., & Ringle, C. M. (2019). Predictive model assessment in PLS-SEM: guidelines for using PLSpredict. *European Journal of Marketing*, *53*(11), 2322–2347. https://doi.org/10.1108/EJM-02-2019-0189

Smith, J. K., & Smith, L. F. (2014). Developing Assessment Tasks. In C. Wyatt-Smith, V. Klenowski, & P. Colbert (Eds.), *Designing Assessment for Quality Learning* (pp. 123–136). Springer. https://doi.org/10.1007/978-94-007-5902-2

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

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, T., Jacobs, D., Hurley, L., Martin, J., Maslyuk, S., Lyall, M., & Ryan, M. (2019). Students' perspectives of early assessment tasks in their first-year at university. *Assessment & Evaluation in Higher Education*, *44*(3), 398–414. https://doi.org/10.1080/02602938.2018.1513992

Wren, J., Sparrow, H., Northcote, M., & Sharp, S. (2009). Higher Education Students' Perceptions of Effective Assessment. *International Journal of Learning*, *15*(12), 11–23. https://doi.org/10.18848/1447-9494/CGP/v15i12/46037

Zhao, X., Lynch, J. G., & Chen, Q. (2010). Reconsidering Baron and Kenny: Myths and truths about mediation analysis. *Journal of Consumer Research*, *37*(2), 197–206. https://doi.org/10.1086/651257

Appendix I: Dimensional structure of the ATAE
(Analysis of Assessment and Learning Tasks) questionnaire

Weigh up the assessment task and score on a scale from 0 (nothing) to 10 (total), to what degree or extent you have implemented or developed the aspects below.

Dimension	LABEL	ITEM No.	Items
DEPTH (PRO)	PRO_1	I1	Use investigation and research methods
	PRO_2	I6	Demonstrate in-depth comprehension of fundamental concepts and ideas.
	PRO_3	I11	Identify, articulate and relate the subject's fundamental concepts and subjects.
	PRO_4	I12	Develop reflexive and critical thinking
COMMUNICATION (COM)	COM_1	I2	Make use of oral or written communication strategies
	COM_2	I8	Provide reasoned and well-founded arguments
	COM_3	I16	Present products to internal or external audiences
CHALLENGING (RET)	RET_1	I3	Establish significant relationships and connections
	RET_2	I4	Coordinate the process and the action to respond to what is required in the task
	RET_3	I5	Assume risks by choosing solutions that imply creativity, greater difficulty or uncertainty
	RET_4	I7	Seek alternative solutions or perspectives
TRANSFER (TRA)	TRA_1	I9	Integrate and relate the prior knowledge, skills and experiences to other new ones, establishing significant and relative connections.
	TRA_2	I10	Relate prior knowledge, skills and experiences to other new ones.
	TRA_3	I13	Relate knowledge and experiences to other matters.
	TRA_4	I14	Relate knowledge and experiences to social reality.
	TRA_5	I15	Make specific products (projects, trials, presentations, debates, performances, etc.)
OPEN QUESTIONS		P1	Do you consider this assessment task to have been challenging? Why? Explain your answer with reasoning
		P2	Do you consider this assessment task to have required intellectual rigour? Why? Explain your answer with reasoning
		P3	Do you consider that this assessment task is related to the professional field? Why? Explain your answer with reasoning
		P4	Overall, what did this task require from you, what skills do you think you have implemented? How do you consider your performance in this task?

Authors / Autores
Ibarra-Sáiz, M.S. (marisol.ibarra@uca.es) 🆔 0000-0003-4513-702X
Senior Lecturer in Educational Assessment and Evaluation at University of Cadiz. Director of UNESCO Chair on Evaluation and Assessment, Innovation and Excellence in Education. Director of EVALfor Research Group-SEJ509 *Assessment & Evaluation in Training Contexts* from the Andalusian Programme of Research, Development and Innovation (PAIDI). She develops her research mainly in the field of assessment and evaluation in higher education. She has been principal researcher of more than 10 European, international and national projects, whose results have been published in various articles, book chapters and contributions to international conferences. She is currently the main co-researcher of the FLOASS Project - Learning outcomes and learning analytics in higher education: An action framework from sustainable assessment (RTI2018-093630-B-I00) in which 6 Spanish universities participate.
Rodríguez-Gómez, G. (gregorio.rodriguez@uca.es) 🆔 0000-0001-9337-1270
Professor of Educational Research Methods at the University of Cadiz. He is the coordinator of the strategic area "Studies and research in assessment and evaluation" of the UNESCO Chair on Evaluation and Assessment, Innovation and Excellence in Education. Founding member of the EVALfor Research Group "SEJ509" Assessment and evaluation in training contexts. His research interest is focused on research methods and assessment and evaluation in higher education. He is currently the main co-researcher of the FLOASS Project - Learning outcomes and learning analytics in higher education: A framework for action from sustainable assessment (RTI2018-093630-B-I00). Author of articles, book chapters and contributions to international conferences. He has been President of the *Interuniversity Association for Pedagogical Research* (AIDIPE). He is currently President of the RED-U *Spanish University Teaching Network*.

RELIEVE

Revista **EL**ectrónica de **I**nvestigación y **EV**aluación **E**ducativa
E-Journal of Educational Research, Assessment and Evaluation
[ISSN: 1134-4032]