ISSN: 1139-613X

4

STUDENTS' PERCEPTION OF AUTO-SCORED ONLINE EXAMS IN BLENDED ASSESSMENT: FEEDBACK FOR IMPROVEMENT

(PERCEPCIÓN DE LOS ESTUDIANTES ACERCA DE LOS EXÁMENES ONLINE CON CORRECCIÓN AUTOMÁTICA EN UNA EVALUACIÓN MIXTA: RETROALIMENTACIÓN PARA LA MEJORA)

Jaime Riera
Miguel Ardid
José A. Gómez-Tejedor
Ana Vidaurre
José M. Meseguer-Dueñas
Universitat Politècnica de València

DOI: 10.5944/educXX1.19559

Cómo referenciar este artículo/how to reference this article:

Riera, J.; Ardid, M.; Gómez-Tejedor, J.A.; Vidaurre, A. & Meseguer-Dueñas, J.M. (2018). Students' perception of auto-scored online exams in blended assessment: feedback for improvement. *Educación XX1, 21*(2), 79-103, doi: 10.5944/educXX1.19559

Riera, J.; Ardid, M.; Gómez-Tejedor, J.A.; Vidaurre, A. & Meseguer-Dueñas, J.M. (2018). Percepción de los estudiantes acerca de los exámenes online con corrección automática en una evaluación mixta: retroalimentación para la mejora. *Educación XX1, 21*(2), 79-103, doi: 10.5944/educXX1.19559

ABSTRACT

Development of the information and communication technologies has led to an increase in the use of Computer Based Assessment (CBA) in higher education. In the last decade, there has been a discussion on online versus traditional pen-and-paper exams. The aim of this study was to verify whether students have reserves about auto-scored online exams, and if that is the case, to determine the reasons. The study was performed in the context of a blended assessment in which 1200 students were enrolled on a first-year physics university course. Among them, 463 answered an anonymous survey, supplemented by information obtained from an open-ended question and from interviews with students. Three factors (labelled 'F1-Learning,' 'F2-Use of Tool,' and 'F3-Assessment') emerged from the quantitative analysis of the survey, and an additive scale was established. We found significant differences in the 'F3-Assessment' factor compared to the other two factors, indicating a lower acceptance of the tool for student assessment. It seems that even though students are used to computers, they have a lack of confidence in online exams. We carried out an in-depth survey on this topic in the form of an open-ended question and by interviewing a small group of 11 students to confer strength and nuance to the quantitative results of the survey. Although their comments were positive in general, especially on ease-of-use and on its usefulness in indicating the level achieved during the learning process, there was also some criticism of the clarity of questions and the strictness of the marking system. These two factors, among others, could have been the cause of the worse perception of F3-Assessment and the origin of the students' reluctance towards online exams and automatic scoring.

KEYWORDS

Higher Education; Feedback (Response); Student Surveys; Interviews; Alternative Assessment; Statistical Analysis.

RESUMEN

El desarrollo de las tecnologías de la información y la comunicación ha producido un incremento del uso de la *Computer Based Assessment* (CBA, evaluación basada en ordenadores). en la educación superior. En la última década, ha habido un debate sobre los exámenes online *vs* los escritos tradicionales. El objetivo del presente estudio ha sido verificar si los estudiantes tienen prejuicios sobre los exámenes online con corrección automática, y si ese es el caso, determinar los motivos. El estudio se realizó en el contexto de una evaluación mixta que implicó a 1200 estudiantes matriculados en una asignatura de física de primer curso universitario. De entre ellos, 463 respondieron a una encuesta anónima. Del análisis cuantitativo de la encuesta surgieron tres factores (etiquetados «*F1-Learning*», «*F2-Use of Tool*» y «*F3-Assessment*»), y se estableció una escala aditiva. Hemos encontrado diferencias significativas en el factor «*F3-Assessment*» en comparación con los otros dos factores, lo que indica una menor aceptación de la herramienta para la evaluación del estudiante. Parece ser que, a pesar de que los estudiantes están acostumbrados a los ordenadores, tienen una falta de confianza en los exámenes online. Para reforzar y matizar los resultados cuantitativos de la encuesta, incluimos una pregunta abierta y realizamos una entrevista a un pequeño grupo de 11 estudiantes. Aunque sus comentarios fueron en general positivos, especialmente sobre la facilidad de uso y sobre su utilidad para conocer el nivel alcanzado durante el proceso de aprendizaje, hubo algunas críticas sobre la claridad de las preguntas y el rigor del sistema de puntuación. Estos dos factores, entre otros, podrían ser la causa de la peor percepción del factor «*F3-Assessment*» y el origen de las reticencias de los estudiantes a los exámenes online y a la corrección automática.

PALABRAS CLAVE

Enseñanza superior; Retroalimentación (respuesta); Encuestas a los estudiantes; Entrevistas; Evaluación alternativa; Análisis estadístico.

INTRODUCTION

The aim of this study was to collect the students' opinion as users of auto-scored online exams (ACBA), as part of a blended assessment. Blended assessment combines the benefits of the Information and Communication Technologies with the traditional assessment using paper and pencil (Llamas-Nistal, Fernández-Iglesias, González-Tato, & Mikic-Fonte, 2013). Our intention was to verify whether the increasing digitalization of society (particularly among young people) has improved student perception of ACBA and then compare the results with previous findings.

The role of assessment and feedback in the learning process has been clearly established (Espasa & Meneses, 2010; Gibbs, 1999). In this context, Computer Based Assessment (CBA) strategies can include a wide variety of clearly explained assignments on a regular basis, providing meaningful and timely feedback to students regarding the quality of their work (Lafuente, Remesal, & Álvarez Valdivia, 2014; Lawton et al., 2012; Nicol, Thomson, & Breslin, 2014). Some form of credit is often given for CBA, and electronic feedback is offered to students either in the form of a grade or with more extensive comments (Wilson & Scalise, 2006). By means of automatic correction, immediate scoring, instant feedback and adaptive testing, continuous assessment or training can be applied without overloading the teacher (Bain, 2004; Chao, Hung, & Chen, 2012; Gipps, 2005; Hwang, Hsu, Shadiev, Chang, & Huang, 2015; Jordan & Mitchell, 2009; Kuo & Wu, 2013; Pacheco-Venegas, López, & Andrade-Aréchiga, 2015), although the reluctance of some educator's to use these systems may be an impediment to the full benefit of their advantages (Debuse & Lawley, 2016). Information about students' achievements in ACBA can be exploited for evaluating a blended assessment project (Jassó, Milani, & Pallottelli, 2008). Recent initiatives in education place great emphasis on developing rich Computer-Based Environments of assessment that make student thinking and reasoning visible (Rosen & Tager, 2014). According to Lee's study on effective online learning, student satisfaction level is closely associated with clear guidelines, rubrics and constructive feedback (Carless, 2015; Lee, 2014). Ardid et al. concluded that the use of ACBA was a useful tool for blended assessment, observing a rather good equivalency between ACBA and traditional offprint exams (Ardid, Gómez-Tejedor, Meseguer-Dueñas, Riera, & Vidaurre, 2015).

When student's opinions of CBA in blended assessment is analysed, most authors find positive acceptance, even though many students prefer the traditional pen-and-paper exams. Some general studies concerning e-Learning found that learner computer anxiety has a negative impact on learner satisfaction (Barbeite & Weiss, 2004; Sun, Tsai, Finger, Chen, & Yeh, 2008). Smaill found that whereas students highlighted the fact that online exams helped them to prepare for the assessment, their perception was not so favourable when they were asked whether they preferred offprint or online exams (Smaill, 2005). Noyes & Garland concluded that this situation should be changing and that there should be greater equivalence between web- and paper-based testing, as students are increasingly using computers in their daily tasks (Noyes & Garland, 2008). Later on, Hewson found that performance scores did not differ depending on whether the assessment was completed in the preferred or non-preferred mode (Hewson, 2012). However, Llamas-Nistal et al. designed a semi-automated grading method to correct traditional pen-and-paper exams, as they considered that many teachers and students prefer this assessment method rather than online exams(Llamas-Nistal et al., 2013). On the other hand, Jawaid et al. found that the majority of the students (61.8%) rated CBA as better as paper based assessment despite experiencing it for the first time (Jawaid, Moosa, Jaleel, & Ashraf, 2014). More recently, Yuan & Kim found that both instructors and students perceived the benefits of CBA although they also expressed the pitfalls and needs for better application (Yuan & Kim, 2015). Therefore, there is still some controversy about the students' opinions on CBA.

Thus, this study was focused on the students' opinions on ACBA. For this purpose, we pose three research questions:

- How do students perceive the ACBA tool? In particular, its use as a learning and assessment tool and possible difficulties encountered in its use.
- Do students consider ACBA a good assessment tool?
- Are there significant differences in the perception of ACBA tool depending on the learning methodology used? In other words, does the way in which ACBA is applied determine the students' perception?

To answer them, a survey was carried out to obtain student's perception of ACBA in the proposed blended assessment model. The qualitative data from the open-ended question and the interview responses were analysed for deeper insights.

METHOD

Auto-scored Computer Based Assessment

Blended assessment and the ACBA tool that we use in this paper were fully described by Ardid et al. (Ardid et al., 2015). The evaluation tool is part of the software application developed through the Sakai project (https://sakaiproject.org). Briefly, exams were randomly obtained from question pools. In these questions the students should complete the different equations involved in solving the problem, and give some relevant intermediate results. In the numeric results, intervals were accepted to cover rounding-off differences. With this procedure, the resolution of the problem was also evaluated, and the student's understanding and ability to solve different problems could be better assessed, while errors in calculations were decreased and weighted. The exams were automatically graded by the application. This aspect together with the equivalency observed between this tool and traditional offprint exams ensures the required quality of the tool for conducting the proposed research.

These exams were proposed for two academic years to different groups in the subject of physics for first-year Engineering degrees at the *Escuela Técnica Superior de Ingeniería del Diseño* (www.etsid.upv.es) of *Universitat Politècnica de València* (www.upv.es). ACBA was used in three different situations: in the classroom in an Evaluation Proctored environment (EP), outside the classroom in an Evaluation-UnProctored environment (EUP) and as a Training-Homework (TH) task. The ACBA exams were used as part of the assessment of the subject (Ardid et al., 2015).

Student survey and interview about the ACBA tool

An anonymous survey was conducted to obtain the students' opinion about the ACBA (Sakai/PoliformaT) exams with the purpose of improving the method. A first version of the survey was designed by two professors based on the analysis of the literature and their own experience of over 30 years. This first design was reviewed by a panel of experts (8 teachers with different experience), to obtain the final design of the survey, which consisted of 11 questions (see list in Table 1). The survey, together with a cover letter, was provided online, being accessible for 10 days to all participants. 463 completed questionnaires were obtained, resulting in a response rate of 39%, with the sampling error = 4.6% ($p = q = 0.5$; significance level, $\alpha = 0.05$). All the questions were quantitative, on a Five-point Likert scale (Likert, 1932) which was converted into a numerical scale from 0 to 10 in the following way: strongly agree (10), agree (7.5), neutral (5), do not agree (2.5), and strongly disagree (0).

Table 1
List of questions clustered by factors. Mean value (MV), standard deviation (SD), rotated factor load (RFL), commonality (C) for each of the questions and Alpha Cronbach value (✓) for each factor. The questions are numbered in the order they were asked, and they are clustered in this table by factor, and ordered by the commonality value, according to the quantitative analysis

Factor/Question	MV	SD	RFL	C
F1-Learning (α = 0.86)				
Q8: ACBA tests are good tools that help students understand the course material better.	7.3	2.4	0.890	0.841
Q7: ACBA tests are good training tools that help students to prepare for the assessment.	7.9	2.4	0.842	0.816
Q11: It has been a useful tool in my learning.	7.3	2.3	0.787	0.772
Q10: This tool gives me a motivation to study.	5.7	2.8	0.550	0.549
F2-Use of tool (α = 0.69)				
Q1: The ACBA exam tool is easy to use.	7.7	2.4	0.710	0.556
Q3: The questions are clear, so there is no possibility of misunderstanding.	5.7	2.3	0.691	0.534
Q2: The questions are based on relevant aspects of the topic.	7.7	2.1	0.629	0.508
Q5: The structure and test scheduling (number of questions, duration, time, etc.) are adequate.	6.7	2.6	0.603	0.450
F3-Assessment (α = 0.80)				
Q9: ACBA tests are a good tool to evaluate student's knowledge.	6.3	2.7	0.778	0.779
Q4: The marks I have obtained in ACBA exams are in accordance with the level of knowledge I had at the time.	5.8	2.7	0.768	0.735
Q6: ACBA tests are a good tool to help students evaluate their level of knowledge.	7.2	2.7	0.574	0.703

Open-ended question: Q12a: Briefly discuss the problems you encountered with the use of the ACBA exam tool. Q12b: Briefly comment on any aspect you believe to be of interest in your assessment of the ACBA examination tool that has not been included in previous questions.

The open-ended question was different in the two school years of the present study. In the first year it focused on the difficulties that students found when performing online exams (Q12a, answered by 86 students). In the second year, the question was more nuanced and gave them the opportunity to highlight not only problems but also positive factors (Q12b, answered by 130 students).

Also, a volunteer group of 11 students were interviewed (3 from the TH group, 5 from the EP group, and 3 from the EUP group). In a first round, all of the participants introduced themselves and gave their general opinion on the topics raised by the teacher. The next round involved a general discussion among the students on different topics while the teacher just listened to them. All of the students participated at some point. The interview was 28 minutes long and was recorded on video.

Quantitative analysis of the students' survey

A statistical analysis of the survey was carried out to validate its consistency and synthesize the results. For the first goal, the structure, reliability and validity of the survey was checked by means of an Exploratory Factor Analysis (EFA) of the internal structure of the answers to the survey, followed by a Confirmatory Factor Analysis (CFA) conducted to examine the stability of the exploratory factor structure according to a widespread method (Ballantine, Guo, & Larres, 2015; Hair, Black, Babin, Anderson, & Tatham, 2010; Xiong, So, & Toh, 2015). The sample was divided randomly into two halves for each analysis. The EFA was performed on IBM SPSS software in Windows Version 16 (IBM, Somers, NY, USA). The results were validated by a CFA using the LISREL statistical package (Linear Structural Relation Statistics Package Program), v. 9.1 (Jöreskog & Sörbom, 1999). Once the consistency of the survey was assured, an additive scale based on the survey was established. The resulting factors were studied by two-way analysis of variance (ANOVA) on IBM SPSS. A significance α level of 0.05 was set for all the statistical tests. With this procedure, by using the sample and the methodology described, we intend to give consistency and validity to the survey, so allowing generalizing the results to similar situations of technical studies in higher education.

Qualitative analysis of the open-ended question and interview

The number of answers to the open-ended question (216) is sufficient to generalize the issues remarked in them. The students' answers were examined looking for recurring themes (Ellis, Goodyear, Bliuc, & Ellis, 2011), making a chart to organize the students' responses (Smith & Suzuki, 2015) taking into account the factors that appeared from the quantitative analysis and selecting illustrative sentences to summarize the most recurrent themes. A similar procedure was followed to analyse the interview. The interview technique is useful to obtain the maximum and more accurate information from homogeneous groups on a target subject through a divergence process (Brill & Galloway, 2007). The interview was oriented to complete and clarify some results of the survey and opinions of the open question to better interpret the results. In this paper the student's statements from the open-ended question and interview are identified by '*italics*' and quotation marks.

RESULTS AND DISCUSSION

From a first observation of the survey results (Table 1) it seems that students consider the ACBA exam tool easy to use, and that the questions are based on relevant aspects of the topic. It seems that they also think that the ACBA tests are good for training and prepare them for the assessment. On the other hand, they may think that this tool does not motivate them to study and that the marks obtained in ACBA exams are not in accordance with their level of knowledge with doubts about the ACBA tests being a good tool to evaluate student knowledge.

Analysis of the validity and reliability of the survey

The method followed to obtain the factor model and the values of the main parameters are fully described in the appendix. The initial sample was randomly segmented into two equal subsamples. With one, comprising 231 cases, an EFA was performed. The principal component analysis was applied as the factor extraction technique. The commonality values, i.e. the variance percentage of each variable explaining the extracted factors, were generally above 50% (Table 1). With the three factors extracted, the total explained variance obtained was 66%, which was acceptable in the context of the study (Hair et al., 2010).

The results are shown in the fourth column of Table 1. The survey can thus be structured into three factors, labelled as F1-Learning, F2-Use of tool, and F3-Assessment.

Confirmatory factor analysis was conducted using the second sub-sample, composed of 232 cases: Table 2 presents a summary with the most frequently used fit indexes values for the suggested model showing that they were within the acceptable limit values (Kline, 2005; Schermelleh-Engel, Moosbrugger, & Müller, 2003). The reader can find all the fitting parameters in the appendix.

Table 2
Confirmatory Factor Analysis results

Index	**Valid range**	**Model**
Normed chi square (χ^2/df)	1-5	1.716
Goodness of Fit Index (GFI)	0.90-1.00	0.991
Root Mean Square Error of Approximation (RMSEA)	0-0.10	0.057
Standardized Root Mean Square Residual (SRMR)	0.08-0	0.051
Goodness of Fit Index (CFI)	0.85-1.00	0.984

Once the joint model has been accepted, each construct should be assessed separately. The Alpha Cronbach value for the whole test set was 0.87, reaching values of 0.86, 0.69 and 0.80 for each of the constructs (Table 1). Reliability exceeds the recommended 0.7 or at least a minimum 0.6 values for a good fit (Hair et al., 2010; Zlatović, Balaban, & Kermek, 2015). In conclusion, the initial model was confirmed (Figure 1).

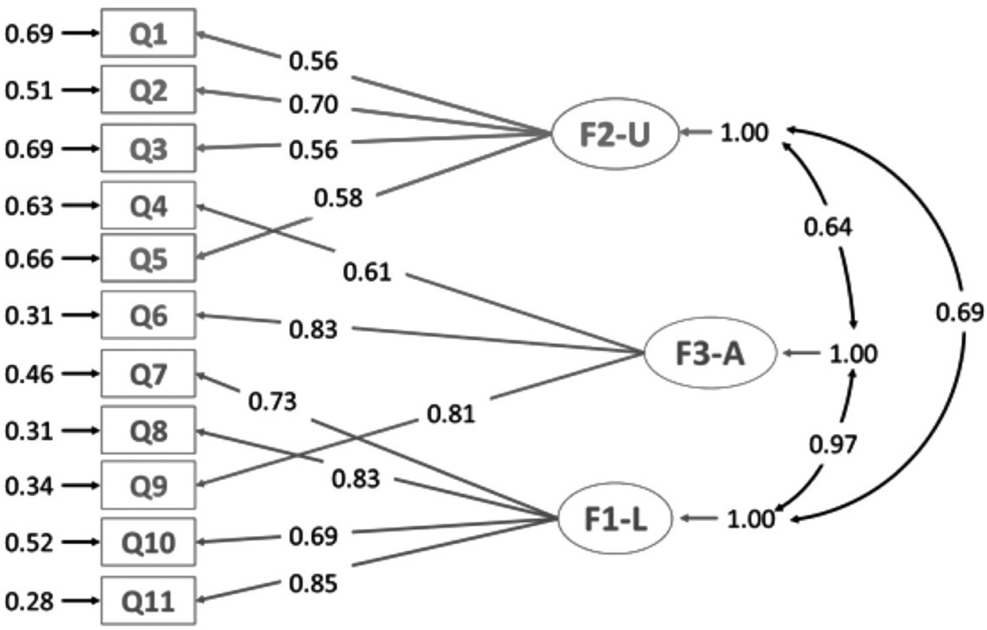

Figure 1. Results of the CFA (standardized solution): a 3-factor model with factor loadings

Students' perception

The results of the survey in global and in terms of the group (methodology) and factor are summarized in Figure 2. All the values are between 6.1 and 7.2 on a scale from 0 to 10 with a global value of 6.8 so that, roughly speaking, we can say that there was a good overall perception of the ACBA tool. However, we were interested in performing an in-depth analysis of the perception of each factor in each group.

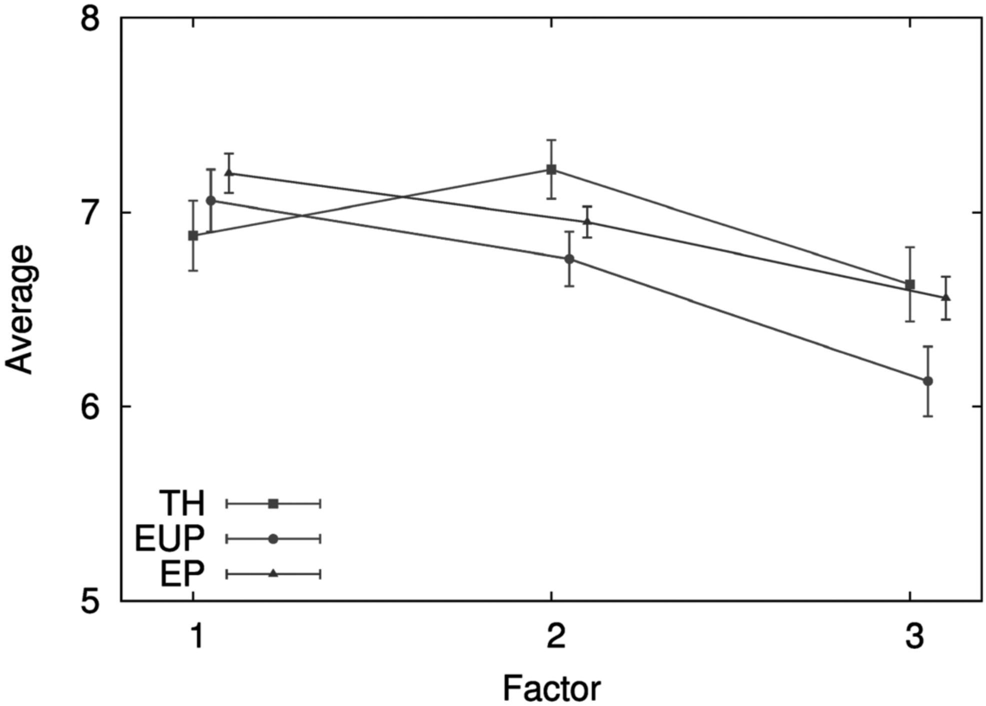

Figure 2. Results of the survey according to CFA as a function of the factor and group. Error bars represent standard error of the mean

With this aim, a two-way analysis of variance ANOVA was carried out with two variables: a) Group, and b) Factors. The null hypotheses that we aimed to study were:

— There are no significant differences between the three factors (H_{01}).
— There are no significant differences between the three groups, i.e., the three methodologies used (H_{02}).
— There is no interaction between variables, i.e. there are no significant differences between groups depending on the factor (H_{03}).

From the statistical analysis, it could be concluded that H_{01} was not fulfilled, showing that there were differences in the satisfaction level in the factors (F_{Factor} = 11.83, df = 2, p < 0.001). On the other hand, we did not find any clear evidence either against H_{02} (F_{Group} = 2.577, p = 0.076) or H_{03} ($F_{Group*Factor}$ = 1.54, df = 4, p = 0.188) (see Table 3).

Table 3
Two-way analysis of the variance. Dependent variable: satisfaction level

Source	F-Ratio	Significant level
Group	2.58	0.076
Factor	11.8	<0.001
Group * Factor	1.54	0.188

Perception for different methodologies (groups). Despite not having any clear evidence that H_{02} was not fulfilled, the fact that the test parameters (F_{Group} = 2.577, p = 0.076) were close to the threshold values might indicate that there could be some particular differences in some aspects between different methodologies. To better address this topic, in Figure 3 we have compared the agreement and disagreement answers depending on the group for the three questions, Q4, Q7 and Q9 that present significant differences between groups. For the rest of the questions the agreement was quite similar for the three groups and no significant differences could be found. The largest difference observed was for Q7 (ACBA tests are good training tools that help students to prepare for the assessment), where 58% of TH students agreed with this and 17% disagreed, whereas for EP and EUP the agreement was much larger (82% and 79%, respectively) and the disagreement much smaller (5%). The results suggest that, although the tool was used mainly for training in TH group, the students appreciated the training use even more if it was linked more directly to the qualification (EUP and EP). There were some differences as well in Q4 (The marks I have obtained in ACBA exams are in accordance with the level of knowledge I had at the time): 50% agreement (15% disagreement) for TH, whereas only 38% and 44% agreement (25% and 23% disagreement) for EP and EUP, respectively. These data can be explained in terms that the mark obtained was less important in the TH group and therefore there was more tolerant to this.

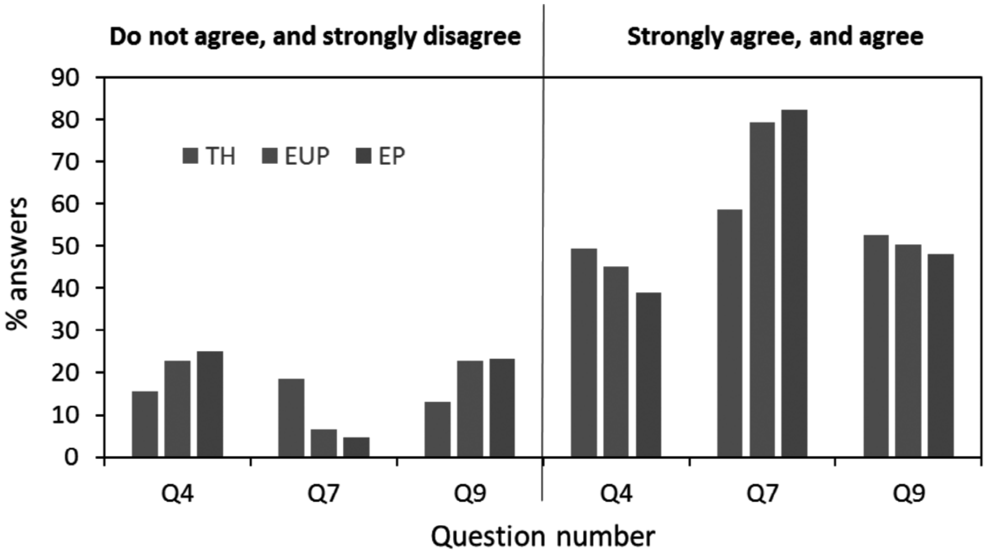

Figure 3. Answer of students on Likert scale by group and question, in %, for the questions that present significant differences between groups. The complete data for all the questions can be seen in the appendix

Q9 (ACBA tests are a good tool to evaluate student's knowledge) presented differences among groups only in disagreement. In our opinion, the reasons for not having found a good acceptance for ACBA exams as assessment tool in EP and EUP cannot be understood in terms of objective reasons, such as these exams being more difficult, since the marks were better in ACBA tests than in traditional offprint exams (Ardid et al., 2015), but it could be more related to students' prejudices that will be further discussed.

Perception for distinct factors. A contrast analysis was carried out to corroborate the differences between the factors. The results show that there were significant differences between factors F1 and F2 with respect to factor F3 ($t = 4.70$, $p < 0.05$), (F1, F2) being larger than F3. On the other hand, there were no significant differences between F1 and F2 ($t = 0.837$, $p = 0.40$).

Since there are no significant differences between the groups, it is better to analyse the perception of the factors globally. Figure 4 shows the histograms of the students' perception of the three factors. We can see that the agreement answer at the F3 is smaller than the other factors.

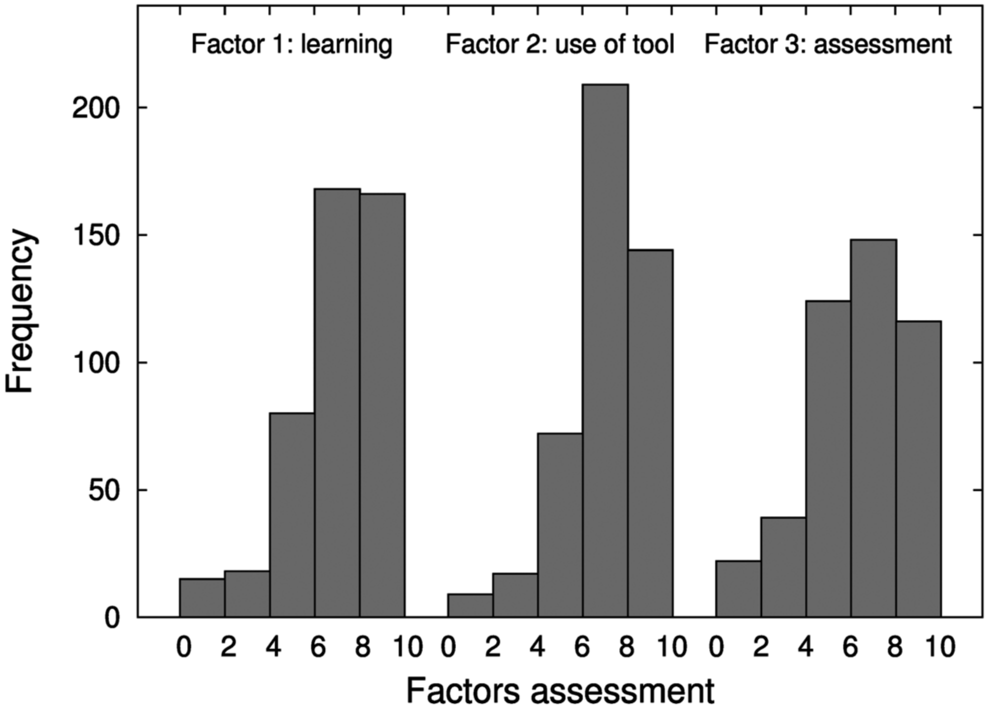

Figure 4. Histograms of students' perception of the different factors

With the aim of obtaining more information about the students' perception on the different factors, we scrutinize the answers to the open-ended question of the survey and the interview.

Factor 1: learning. Some students were satisfied with the method because they thought that ACBA exams were useful for preparing written exams: '*I think it gives you an idea of what the written exam will be like and helps you to focus on studying the fundamental topics*.' They also appreciated that the method '*helps us to make a study plan*.'

Other students, even if they are more critical of the methodology, also consider that it helps to prepare the writing exam because '*it forces us to study for the exam*.'

Factor 2: use of the tool. Some students pointed out the problems they had with the tool. Others proposed improvements to be incorporated into the method, for instance providing additional feedback, besides the total score, after the tests. Some students had difficulties in interpreting some instructions, either because they were confusing, or it was unclear what was required of them: '*some exercises are set out in an ambiguous way and are* *difficult to understand*.' This could be related to the natural insecurity felt by students, mainly when there is an assessment and the teacher is not present and they cannot ask for clarifications. Despite the exhaustive process of revision of the question pools, these comments demonstrate the need for continuous feedback from students to improve the question pools and to mitigate any possible misinterpretation of the questions. Another problem raised involved the platform and the access to Internet: '*a very good internet connection is needed for ACBA exams*.'

We took advantage of the interview to enquire about the clarity of statements of questions in on-line tests. One student said that it is possible that '*some people sometimes look beyond the statement*,' and other student points to the notation that is asked differently in each question: '*in some questions sign and module are asked, and sometimes not*.' A third student said that this issue was not a problem because '*the teacher makes everything clear in class from the beginning, and there is no doubt*.' These opinions suggested that the problem observed in the open-ended question was not as important as might appear at first sight.

Factor 3: Assessment. Even though the Factor 3 mark was lower than the other ones, the students' comments were in general positive: '*They are especially useful to know the level you have in the subject before taking the offprint exam*,' and in addition, '*it helps you to get better scores in the class marks*.' They acknowledged that '*the obtained marks are higher than in the offprint ones and may not actually be representative of the student's knowledge*.'

An important issue they pointed out was the lack of control over the authorship of examinations not taken in the classroom. Students evidenced it very clear: '*certain tests are not performed by a single student but by several*' since '*the exams are online, they can be done by friends, classmates, tutors, etc. Nobody checked (efficiently) if I had really done the exam*.' And also, '*That affects those who do these tests honestly*.' This fact could explain the differences found in the question Q9; the EUP group showing higher disagreement. However, the same opinion was shown by the proctored group EP, which performed the exams in class. Then, this should represent just one of the reasons for disagreement.

In addition, they complained about the objectivity of the machine correction: '*the answer must be very precise to rate the question as correct*,' and consequently ask for '*a larger interval range in the solutions…*'. A related complaint is about the sign of the results, because '*when you're wrong on a sign, the whole score is removed, rather than only a part being taken away*.' From our point of view, this is not a real problem because the accepted interval in the results was calculated to allow for *reasonable* rounding off effects and the *sign* is a fundamental aspect of the solution. We think that this complaint could be due to the students' fear of making mistakes in the calculations. The students also complained that the computer tool was only able to correct numerical answers, but not to assess the method used to solve the problem: '*the whole procedure can be correct, but for one miscalculation you can lose the entire mark*.' In this way, they regard that ACBA exam as a '*good tool for learning, but not a tool for grading knowledge*.' These could be the reasons for obtaining a lower grade in this factor of the survey compared to the others. The length of the test is often a recursive aspect in any kind of exam: some students argued that the exam duration was not correct: '*in some cases 20 minutes are left over and in others more time is needed*.'

In conclusion, there are common problems between offprint and ACBA exams: clarity of statements, test duration, miscalculations, etc. In ACBA the additional difficulty is the inability to ask to the teacher about the doubts that students may arise, and consequently extra efforts should be made to avoid these problems wherever possible. Regarding the doubts about authorship of the tests, it was clearly seen as negative point for assessment but also students interpret it in a positive way for learning.

CONCLUSIONS

This paper presents an analysis of students' opinions on online exams as part of the evaluation process in the context of blended assessment. The factorial analysis of the anonymous survey established that 66% of the total variance was explained by three factors, labelled as 'F1-Learning,' 'F2-Use of tool,' and 'F3-Assessment.' An additive scale was also built and indicated that the students' general perception of the online exam tool was good (6.8 out of 10). The student' opinion was better in reference to the learning aspects or the usability (7.1 for F1 and 7.0 for F2) than what refers to the use as assessment tool (6.4 for F3). No significant differences were observed in the two-way analysis of variance between the groups (methodology used), whereas significant differences were observed in the 'F3-Assessment' factor, compared to the other two factors, indicating a worse perception of the tool for assessment purposes. The sample and methodology used in this analysis gives to the survey a remarkable consistency and validity which allows us to think that these results can be extended to other technical studies in higher education where similar tools are used. In this regard, the results are basically in agreement with previous studies (Noyes & Garland, 2008; Smaill, 2005) but the quantification and causes of students' reticence to ACBA exams were more clearly stated here by combining the statistical analysis with the information obtained from an open-ended question and from the interview with the students. Their comments were in general positive, especially on ease-of-use and its usefulness during the learning process to diagnose the level achieved. On the other hand, there were also some criticisms, especially in terms of clarity of the questions (when the teacher was not present) and of the rigidity of the automatic scoring (Chao et al., 2012). This certainly could be the cause of the poorer perception of F3-Assessment than the other ones. It could also be the reason for the students' reticence towards ACBA. It would therefore be advisable to keep on working on the design of questions, considering not only the final result but also the intermediate results and the procedure. With this study, strong and weak points of ACBA exams have been obtained from the analysis of students' opinions. This can be very useful to set strategies that favour the positive aspects and to mitigate possible problems of limitations towards its use in a better and more efficient blended assessment.

ACKNOWLEDGEMENTS

This work was supported by the Universitat Politècnica de València through the A15/16 Project (Convocatoria de Proyectos de Innovación y Convergencia de la UPV). We would like to thank the ICE in the Universitat Politècnica de València for their help, through the Innovation and Educational Quality Program and for supporting the team *Innovación en Metodologías Activas para el Aprendizaje de la Física (e-MACAFI)*.

APPENDIX: EXPLORATORY AND CONFIRMATORY FACTORIAL ANALYSIS

Input Data

The analysis of the input data showed that missing data (missing values) were randomly distributed between the different variables, so they could be treated as Missing Completely At Random data type (MCAR). As only a small percentage of the total data (approximately 5%) contained missing values, it was decided to use only observations with complete data, following the 'listwise deletion' procedure (Roth, 1994). The sample size is adequate from the point of view of the factorial analysis (Hair et al., 2010).

To test the sample distribution, input data were evaluated using the Shapiro and Kolmorogov tests, which showed statistical significance, indicating that a normal distribution of variables could not be assumed, and therefore that there was no multivariate normality. This made it necessary to use robust estimators for the CFA to be consistent (Coenders, Satorra, & Saris, 1997; Joreskog, 1990).

Exploratory factor analysis

The result of the Bartlett test of sphericity was excellent (0.901) and the Kaiser-Meyer-Olkin test showed statistical significance ($\chi^2 = 1946$, df = 55, $p < 0.001$), which confirmed that the available data was appropriate for a factor analysis.

The initial sample was randomly segmented into two equal subsamples. With one, comprising 231 cases, an EFA was performed. The principal component analysis was applied as the factor extraction technique. Based on the Scree-test, the number of factors to be extracted was established as 'three'. The commonality values, i.e. the variance percentage of each variable explaining the extracted factors, were generally above 50% (see Table I in the paper). The cumulative variance, the total explained variance and the eigenvalues of each of the three factors obtained as a result of the factor analysis are presented in Table I. With the three factors extracted, the total explained variance obtained was 66%, which was acceptable in the context of the study (Hair et al., 2010; Netemeyer, Bearden, & Sharma, 2003).

Varimax vertical axis rotation was used to reset the correlations between factors and to help in their interpretation. The results of the Rotated Factor Load are shown in the fourth column of Table 1 of the paper. The survey can thus be structured into three factors, labelled as F1-Learning, F2-Use of tool, and F3-Assessment.

Table I
Cumulative and total explained variances of the factor analysis. The eigenvalue of each factor is also shown

Initial Eigenvalues			**Rotation Sums of Squared Loadings**		
Total	**% of Variance**	**Cumulative %**	**Total**	**% of Variance**	**Cumulative %**
5.172	47.0	47.0	3.056	27.8	27.8
1.216	11.1	58.1	2.162	19.7	47.4
0.854	7.8	65.8	2.025	18.4	65.8

CONFIRMATORY FACTOR ANALYSIS

Confirmatory factor analysis was conducted using the second sub-sample, composed of 232 cases. Starting with a model in which the 3 factors are exogenous constructs, no correlated measures were proposed within the construct, even though the factors may be correlated. Given that the variables cannot be considered to follow a multi-normal distribution, the diagonally weighted least squares sample analysis was used for the analysis, since it is more appropriate than the traditional maximum likelihood estimation method (Muthen & Kaplan, 1992; Yang-Wallentin, Joreskog, & Luo, 2010). In order to make the scale invariant and each construct comparable, a weight of '1' was fixed for each construct. From the results obtained it can be concluded that there were no infringing estimates, so the quality of the CFA could be evaluated without redesign (Hair et al., 2010).

Table II presents the fit values of the suggested model and the acceptable limit values of the most frequently used fit indexes (Kline, 2005; Schermelleh-Engel et al., 2003). For the sake of clarity, the fitting parameters were classified into three groups: Absolute, Incremental and Parsimony (Schermelleh-Engel et al., 2003).

The absolute adjustment parameters only assess the overall fit of the model, without adjustment for the degree of overfitting. The Incremental Adjustment evaluates the model fit comparatively to a null model, which is a model with a single factor without uncertainty in the measurement. The Parsimony adjustment showed the model fit against the number of estimated coefficients. The review of the different types of adjustment parameters provided additional evidence for acceptance of the proposed model.

Table II
Confirmatory Factor Analysis results

Index	Valid range	Initial Model	Modified Model
ABSOLUTE			
Chi-square (χ^2)	—	75.89	42.97
p-value		0.001	0.093
Degrees of freedom (df)	—	41	32
Goodness of Fit Index (*GFI)*	0.90/1.00	0.991	0.994
Root Mean Square Error of Approximation (RMSEA)	0/0.10	0.057	0.045
Standardized Root Mean Square Residual (*SRMR)*	0.08/0	0.051	0.043
Goodness of Fit Index (CFI)	0.85/1.00	0.984	0.994
INCREMENTAL			
Normed Fit Index (*NFI)*	0.85/1.00	0.966	0.976
Non-Normed Fit Index (NNFI)	0.85/1.00	0.978	0.991
Adjusted Goodness of Fit Index (*AGFI)*	0.85/1.00	0.985	0.989
PARSIMONY			
Normed chi square (χ^2 /df)	1/5	1.716	1.34
Parsimony Normed Fit Index (PNFI)		0.72	0.694
Parsimony Goodness of Fit Index (PGFI)		0.615	0.578

Reliability

Once the joint model has been accepted, each construct should be assessed separately. The Alpha Cronbach value for the whole test set was 0.87, reaching values of 0.86, 0.69 and 0.80 for each of the constructs (see Table 1 of the paper), while the extracted variances were 0.61, 0.36 and 0.57, respectively. Finally, the *t* values associated with the correlation between constructs had values greater than 2.6, so the existence of correlation between the constructs was evident.

Re-specification Model

The goodness results of the fit and the different constructs of the joint model yield evidence to confirm the model with 3 constructs. However, LISREL provided guidance for orienting and performing re-specification of the model to improve the fit. The setting values of the modified model are shown in Table II, showing improvements over the original model in absolute parameters. However, from the Parsimony fit indexes, the initially proposed model had better values. So, there was no justification for re-specifying the model, concluding that the initial model was the most appropriate.

REFERENCES

Ardid, M., Gómez-Tejedor, J.A., Meseguer-Dueñas, J.M., Riera, J., & Vidaurre, A. (2015). Online exams for blended assessment. Study of different application methodologies. *Computers & Education*, *81*, 296-303. https://doi.org/10.1016/j.compedu.2014.10.010

Bain, K. (2004). *What do they know about how we learn? What the Best College Teachers Do*. Cambridge, MA: Harvard University Press.

Ballantine, J., Guo, X., & Larres, P. (2015). Psychometric evaluation of the Student Authorship Questionnaire: a confirmatory factor analysis approach. *Studies in Higher Education*, *40*(4), 596-609. https://doi.org/10.1080/03075079.2013.835910

Barbeite, F.G., & Weiss, E.M. (2004). Computer self-efficacy and anxiety scales for an Internet sample: testing measurement equivalence of existing measures and development of new scales. *Computers in Human Behavior*, *20*(1), 1-15.

Brill, J.M., & Galloway, C. (2007). Perils and promises: University instructors' integration of technology in classroom-based practices. *British Journal of Educational Technology*, *38*(1), 95-105. https://doi.org/10.1111/j.1467-8535.2006.00601.x

Carless, D. (2015). Exploring learning-oriented assessment processes. *Higher Education*, *69*(6), 963-976. https://doi.org/10.1007/s10734-014-9816-z

Chao, K.-J. J., Hung, I.-C. C., & Chen, N.-S. S. (2012). On the design of online synchronous assessments in a synchronous cyber classroom. *Journal of Computer Assisted Learning*, *28*(4), 379-395. https://doi.org/10.1111/j.1365-2729.2011.00463.x

Coenders, G., Satorra, A., & Saris, W. E. (1997). Alternative Approaches to Structural Modeling of Ordinal Data: A Monte Carlo Study. *Structural Equation Modeling-a Multidisciplinary Journal*, *4*(4), 261-282.

Debuse, J.C.W., & Lawley, M. (2016). Benefits and drawbacks of computer-based assessment and feedback systems: Student and educator perspectives. *British Journal of Educational Technology*, *47*(2), 294-301. https://doi.org/10.1111/bjet.12232

Ellis, R.A., Goodyear, P., Bliuc, A.-M., & Ellis, M. (2011). High school students' experiences of learning through research on the internet. *Journal of Computer Assisted Learning*, *27*(6), 503-515. https://doi.org/10.1111/j.1365-2729.2011.00412.x

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

Gibbs, G. (1999). Using assessment strategically to change the way students learn. In *Assessment Matters in Higher Education* (pp. 41-53). https://doi.org/10.1007/s13398-014-0173-7.2

Gipps, C. V. (2005). What is the role for ICT-based assessment in universities? *Studies in Higher Education*, *30*(2), 171-180. https://doi.org/10.1080/03075070500043176

Hair, J.F., Black, W.C., Babin, B.J., Anderson, R.E., & Tatham, R.L. (2010). *Multivariate Data Analysis*. *Prentice Hall*.

Hewson, C. (2012). Can online course-based assessment methods be fair and equitable? Relationships between students' preferences and performance within online and offline assessments. *Journal of Computer Assisted Learning*, *28*(5), 488-498.

Hwang, W.-Y., Hsu, J.-L., Shadiev, R., Chang, C.-L., & Huang, Y.-M. (2015). Employing self-assessment, journaling, and peer sharing to enhance learning from an online course. *Journal of Computing in Higher Education*, *27*(2), 114-133. https://doi.org/10.1007/s12528-015-9096-3

Jassó, J., Milani, A., & Pallottelli, S. (2008). Blended e-Learning: Survey of On-line Student Assessment. In *2008 19th International Conference on Database and Expert Systems Applications* (pp. 626-630). IEEE. https://doi.org/10.1109/DEXA.2008.115

Jawaid, M., Moosa, F.A., Jaleel, F., & Ashraf, J. (2014). Computer Based Assessment (CBA): Perception of residents at Dow University of Health Sciences. *Pakistan Journal of Medical Sciences*, *30*(4), 688-91. https://doi.org/10.12669/pjms.304.5444

Jordan, S., & Mitchell, T. (2009). e-Assessment for learning? The potential of short-answer free-text questions with tailored feedback. *British Journal of Educational Technology*, *40*(2), 371-385. https://doi.org/10.1111/j.1467-8535.2008.00928.x

Joreskog, K.G. (1990). New Developments in Lisrel - Analysis of Ordinal Variables using Polychoric Correlations and Weighted Least-Squares. *Quality & Quantity*, *24*(4), 387-404.

Jöreskog, K.G., & Sörbom, D. (1999). *LISREL 8 user's guide*. *Chicago: Scientific Software International*.

Kline, T.J.B. (2005). *Psychological Testing A Practical Approach to Design and Evaluation*. SAGE Publications, Inc.

Kuo, C.-Y., & Wu, H.-K. (2013). Toward an integrated model for designing assessment systems: An analysis of the current status of computer-based assessments in science. *Computers & Education*, *68*, 388-403. https://doi.org/10.1016/j.compedu.2013.06.002

Lafuente, M., Remesal, A., & Álvarez Valdivia, I. M. (2014). Assisting learning in e-assessment: a closer look at educational supports. *Assessment & Evaluation in Higher Education*, *39*(March 2015), 443-460. https://doi.org/10.1080/02602938.2013.848835

Lawton, D., Vye, N., Bransford, J., Sanders, E., Richey, M., French, D., & Stephens, R. (2012). Online Learning Based on Essential Concepts and Formative Assessment. *Journal of Engineering Education*, *101*(2), 244-287. https://doi.org/10.1002/j.2168-9830.2012.tb00050.x

Lee, J. (2014). An Exploratory Study of Effective Online Learning: Assessing Satisfaction Levels of Graduate Students of Mathematics Education Associated with Human and Design Factors of an Online Course. *International Review of Research in Open and Distance Learning*, *15*(1).

Likert, R. (1932). A technique for the measurement of attitudes. *Archives of Psychology*, *22*, 1-55. Retrieved from http://www.sciepub.com/reference/113453

Llamas-Nistal, M., Fernández-Iglesias, M. J., González-Tato, J., & Mikic-Fonte, F. A. (2013). Blended e-assessment: Migrating classical exams to the digital world. *Computers & Education*, *62*, 72-87. https://doi.org/10.1016/j.compedu.2012.10.021

Muthen, B., & Kaplan, D. (1992). A Comparison of some Methodologies for the Factor-Analysis of Nonnormal Likert Variables - a Note on the Size of the

Model. *British Journal of Mathematical & Statistical Psychology*, *45*, 19-30.

Netemeyer, R. G., Bearden, W. O., & Sharma, S. (2003). Scaling procedures: issues and applications. In *Thousand Oaks* (p. 219-261 Chap. 10). https://doi.org/10.4135/9781412985772

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

Noyes, J.M., & Garland, K.J. (2008). Computer- vs. paper-based tasks: Are they equivalent? *Ergonomics*, *51*(9), 1352-1375.

Pacheco-Venegas, N.D., López, G., & Andrade-Aréchiga, M. (2015). Conceptualization, development and implementation of a web-based system for automatic evaluation of mathematical expressions. *Computers & Education*, *88*, 15-28. https://doi.org/10.1016/j.compedu.2015.03.021

Rosen, Y., & Tager, M. (2014). Making Student Thinking Visible through a Concept Map in Computer-Based Assessment of Critical Thinking. *Journal of Educational Computing Research*, *50*(2), 249-270. https://doi.org/10.2190/EC.50.2.f

Roth, P.L. (1994). Missing Data - A Conceptual Review for Applied Psychologists. *Personnel Psychology*, *47*(3), 537-560. https://doi.org/10.1111/j.1744-6570.1994.tb01736.x

Schermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the Fit of Structural Equation Models: Tests of Significance and Descriptive Goodness-of-Fit Measures. *Methods of Psychological Research Online*, *8*(2), 23-74. https://doi.org/10.1002/0470010940

Smaill, C.R. (2005). The implementation and evaluation of OASIS: A web-based learning and assessment tool for large classes. *Ieee Transactions on Education*, *48*(4), 658-663.

Smith, J. G., & Suzuki, S. (2015). Embedded blended learning within an Algebra classroom : a multimedia capture experiment. *Journal of Computer Assisted Learning*, *31*, 133-147. https://doi.org/10.1111/jcal.12083

Sun, P.-C., Tsai, R.J., Finger, G., Chen, Y.-Y., & Yeh, D. (2008). What drives a successful e-Learning? An empirical investigation of the critical factors influencing learner satisfaction. *Computers & Education*, *50*(4), 1183-1202.

Wilson, M., & Scalise, K. (2006). Assessment to improve learning in higher education: The BEAR assessment system. *Higher Education*, *52*(4), 635-663.

Xiong, Y., So, H.-J., & Toh, Y. (2015). Assessing learners' perceived readiness for computer-supported collaborative learning (CSCL): a study on initial development and validation. *Journal of Computing in Higher Education*, *27*(3), 215-239. https://doi.org/10.1007/s12528-015-9102-9

Yang-Wallentin, F., Joreskog, K.G., & Luo, H. (2010). Confirmatory Factor Analysis of Ordinal Variables With Misspecified Models. *Structural Equation Modeling-a Multidisciplinary Journal*, *17*(3), 392-423.

Yuan, J., & Kim, C. (2015). Effective Feedback Design Using Free Technologies. *Journal of Educational Computing Research*, *52*(3), 408-434.

Zlatović, M., Balaban, I., & Kermek, D. (2015). Using online assessments to stimulate learning strategies and achievement of learning goals. *Computers & Education*, *91*, 32-45. https://doi.org/10.1016/j.compedu.2015.09.012

PERFIL ACADÉMICO Y PROFESIONAL DE LOS AUTORES

Jaime Riera. Received his PhD degree in Physics in 1990 from Universitat de València. He is a full professor in the Department of Applied Physics at the Universitat Politècnica de València (Spain). Dr. Riera has directed projects focused on the transmission of scientific knowledge through Information and Communication Technologies (ICTs). He has produced several documentaries on scientific subjects. He is currently researching at the Multidisciplinary Mathematics Institute, in the area of image processing and digital tracking.

Miguel Ardid. Received his Ph.D. degree in Physics at IFIC-Valencia in 2002. He is a full professor of the Department of Applied Physics in the Universitat Politècnica de València. He is leading the «acoustics for astroparticle detection» research group at UPV, involved in several international Collaborations: KM3NeT, ANTARES, PICO. His research interests include astroparticle physics, particle detectors, acoustics, sensors, underwater systems and education, and he is the author of more than 100 research articles.

José A. Gómez-Tejedor**.** Received his Ph.D. in physical sciences in 1995 from the Universitat de València, Spain. Since 1996 he is a professor of the Department of Applied Physics in the Universitat Politècnica de València. He is the author of seven books and more than 50 articles. His research interests include physical properties of polymer materials for tissue engineering. He is also interested in applying information and communications technology to improve education and training systems.

Ana Vidaurre**.** Received her Ph.D. in 1990 from the Universitat de València, Spain. Since 1980 she has been working in different positions with the Applied Physics Department, Universitat Politècnica de València, Spain. She researches in the Centre for Biomaterials and Tissue Engineering at the Universitat Politècnica de València (Spain); she has written more than 50 articles in the field of polymers, tissue engineering and educational issues related to electrical and electronic engineering education.

José M. Meseguer-Dueñas. Obtained his Ph.D. degree from the Universitat Politècnica de València, Spain in 1988. Since 1985 he is a professor of the Department of Applied Physics in the Universitat Politècnica de València. He currently holds the position of full professor. He is the author of fourteen docent books, more than 70 articles, and more than 100 communications in the field of physical properties of polymer, tissue engineering and the use of ICT in university education.

Dirección de los autores: Jaime Riera
Miguel Ardid
José A. Gómez-Tejedor*
Ana Vidaurre
José M. Meseguer-Dueñas
ETS de Ingeniería del Diseño, Departamento de Física Aplicada
Universitat Politècnica de València
Camino de Vera, s/n
València, Spain
*E-mail: jogomez@fis.upv.es

Fecha Recepción del Artículo: 08. Septiembre. 2017
Fecha Modificación del Artículo: 19. Enero. 2018
Fecha Aceptación del Artículo: 22. Enero. 2018
Fecha Revisión para Publicación: 22. Febrero. 2018