

La evaluación del aprendizaje de estudiantes: validación española del Assessment Experience Questionnaire (AEQ)

The Assessment of Student Learning: Spanish Validation of the Assessment Experience Questionnaire (AEQ)

JUAN L. NÚÑEZ

Universidad de Las Palmas de Gran Canaria
juanluis.nunez@ulpgc.es

CARMEN I. REYES

Universidad de Las Palmas de Gran Canaria
creyes@dedu.ulpgc.es

Resumen: El propósito de esta investigación fue traducir al español y analizar las propiedades psicométricas del Assessment Experiences Questionnaire (AEQ) ya que no se dispone de un cuestionario en español que evalúe la experiencia de evaluación del aprendizaje de los estudiantes. Los participantes fueron 329 estudiantes universitarios. Los resultados del análisis de correlaciones y de la regresión lineal múltiple apoyaron la validez predictiva. El análisis factorial confirmatorio confirmó la estructura de nueve factores. Además, la consistencia interna fue aceptable y la correlación test-retest fue moderada. En conclusión, los resultados apoyan preliminarmente el uso de la versión española del AEQ.

Palabras clave: evaluación orientada al aprendizaje; propiedades psicométricas; educación superior; rendimiento académico.

Abstract: The aim of this study was the translation of the Assessment Experiences Questionnaire (AEQ) into Spanish, and the analysis of its psychometric properties. We carried out this study because there is no questionnaire in Spanish which evaluates the experience of learning assessment. Participants were 329 university students. Results of the correlation analysis and the multiple linear regression supported the predictive validity of the questionnaire. The confirmatory factor analysis confirmed the nine-factor structure. In addition, internal consistency was acceptable, and the test-retest correlation was moderate. In conclusion, the results support the use of the Spanish version of the AEQ as an instrument to assess the students' perception of the experience of learning assessment.

Keywords: learning-oriented assessment; psychometric properties; higher education; academic performance.



JUAN L. NÚÑEZ / CARMEN I. REYES

INTRODUCCIÓN

A partir de los años 90 en determinados ámbitos investigadores se inicia un cambio en la concepción sobre la naturaleza del aprendizaje cuyo foco de atención se centrará en la capacidad de un estudiante de articular y utilizar globalmente los conocimientos, destrezas y actitudes necesarios para solucionar problemas específicos (formación basada en competencias) dejando atrás el modelo tradicional preocupado por lo que sabe un alumno en cuanto contenido.

Esto originará un cambio en el marco conceptual de la evaluación educativa desplazando el interés casi exclusivo de principios psicométricos hacia otros centrados en la evaluación educativa (Mateo, 2006). Este nuevo enfoque alternativo de la evaluación es muy amplio y en él podemos encontrar diferentes modalidades o perspectivas: evaluación auténtica, evaluación orientada al aprendizaje, evaluación del desempeño o ejecución, etc. Es viable descubrir algunas diferencias entre el significado de estos términos pero todas estas perspectivas tienen el siguiente objetivo fundamental: enfatizar que el propósito fundamental de la evaluación debería ser promocionar el aprendizaje.

En adelante nos referimos a este enfoque como evaluación orientada al aprendizaje (EOA) porque estos vocablos destacan una de las características principales: la importancia de la evaluación en la optimización del aprendizaje. Este término es utilizado por primera vez en educación superior por Carless (2003) quien en diferentes trabajos ha señalado las principales condiciones de esta conceptualización de la evaluación en la universidad (Carless, 2007, 2009; Carless, Salter, Yang y Lam, 2011): a) las tareas de evaluación son entendidas como tareas de aprendizaje. Esto quiere decir que las tareas de evaluación deben promover experiencias de aprendizaje profundo que se orienten hacia los resultados de aprendizaje esperados; b) la implicación de los estudiantes en la evaluación de tal forma que desarrollen una mejor comprensión de las metas de aprendizaje y que se ocupen más activamente en los estándares y criterios y c) los estudiantes deben recibir retroalimentación o feedback adecuado que podrían utilizar como prealimentación o feedforward en el trabajo futuro.

Desde la EOA, Biggs (2005) enfatiza la importancia del alineamiento constructivo del currículo es decir, la evaluación debe estar alineada con los resultados de aprendizaje y las actividades a realizar, si realmente queremos desarrollar una educación de calidad. Recientemente McDowell, Wakelin, Montgomery y King (2011) aportan una variación a la conceptualización de la EOA a partir de un modelo comprensivo que no sólo se centra en la calidad y oportunidad de feedback que le aportan los tutores o del diseño de tareas (aunque ambos sean importantes),



LA EVALUACIÓN DEL APRENDIZAJE DE ESTUDIANTES

sino en la necesidad que tienen los estudiantes de un ambiente que les permita valorar cómo lo están realizando y ofrecerles oportunidades para mejorarlos. Han definido la EOA como un ambiente de evaluación que es rico en feedback formal, por ejemplo comentarios del tutor, sistemas de autoevaluación, etc.; es rico en feedback informal a través del dialogo equitativo y equilibrado en el aula y la interacción de pares provee oportunidades para probar y practicar conocimientos, comprensión y destrezas; desarrolla tareas de evaluación que son auténticas o relevantes de cara a la actividad profesional; ayuda a los estudiantes a desarrollar autonomía e independencia y presenta un equilibrio apropiado entre evaluación formativa y sumativa.

Desde este enfoque, se asume que la evaluación afecta a la forma de estudiar de los aprendizajes (Gibbs, 2006). La evaluación enmarca el aprendizaje, genera la actividad de aprendizaje y orienta la mayor parte de los aspectos de la conducta de aprendizaje (Gibbs y Simpson, 2004). Todas las percepciones de la evaluación de los estudiantes influyen considerablemente en las aproximaciones al aprendizaje y a la realización de los estudiantes. Turner y Gibbs (2010) han confirmado que las respuestas de aprendizaje de los estudiantes en ambientes diferentes de evaluación pueden ser notablemente diferentes. Las respuestas positivas de aprendizaje (aproximación al aprendizaje profundo, cobertura total del programa, aprendizaje desde la preparación de un examen) se ha mostrado que están asociadas con mucha evaluación formativa, poca evaluación sumativa y mucho feedback oral. Se ha encontrado que las respuestas negativas de aprendizaje (ser estratégico en cuanto estudiar el programa y en una aproximación superficial) está asociado a mucha y variada evaluación sumativa, detallada especificación de resultados y criterios y poca evaluación formativa. Por lo que podemos decir que hay evidencias que confirman que existe una relación entre el tipo de ambiente de evaluación adoptado por una universidad y la actividad de estudio de sus estudiantes.

En síntesis, podríamos concebir la EOA como un constructo teórico que se centra en la evaluación, como un proceso interrelacionado con el aprendizaje basado en la implicación de los estudiantes en tareas apropiadas, en la provisión de abundante feedback incluyendo recursos de auto y coevaluación y sustentado en un compromiso por mejorar el aprendizaje por todos. Esta modificación en la conceptualización de la evaluación en las investigaciones también está generando numerosas innovaciones educativas, aunque en la práctica cotidiana de las universidades españolas la evaluación sigue teniendo un carácter tradicional centrada en la calificación, dirigida por el profesor y donde la participación de los estudiantes es prácticamente nula (Ibarra y Rodríguez, 2010; Reyes, Sosa, Marchena y Marchena, 2012).



JUAN L. NÚÑEZ / CARMEN I. REYES

En nuestro país, observamos que a pesar que, de alguna manera, se están planteando algunas mejoras, necesitamos instrumentos formales que nos proporcionen una visión y un panorama más amplio sobre las formas en las que las experiencias de evaluación de los estudiantes están influyendo en el aprendizaje. Desde este marco conceptual, las tareas o experiencias de evaluación constituyen el elemento que más influye sobre el aprendizaje. De ahí, la necesidad de analizar cómo están diseñadas: tiempo de esfuerzo del alumno, temas en que se centran, etc. y, de esta manera, detectar, si realmente están apoyando el aprendizaje.

Actualmente, prácticamente todas las universidades están sumergidas en procesos de calidad en donde, a través de cuestionarios, entre otros instrumentos, se analiza la calidad de la enseñanza universitaria: claridad de los objetivos, carga de trabajo apropiado, evaluación adecuada, etc. Pese a esto, no disponemos de un cuestionario que, desde esta perspectiva, nos ayude a diagnosticar el impacto de las experiencias de evaluación a fin de que los profesores puedan utilizarlo para evaluar el diseño de sus sistemas de evaluación con objeto de realizar mejoras orientadas a la optimización del aprendizaje.

En el contexto anglosajón podemos encontrar una larga lista de inventarios relacionados con el aprendizaje de los estudiantes con sus experiencias en el curso y módulos (Richardson, 2005). Entre todos ellos, el más reciente y ampliamente conocido en Educación Superior en el Reino Unido es el Assessment Experience Questionnaire (AEQ) desarrollado por Gibbs y Simpson (2003) y que ha sido traducido a diversos idiomas. Este cuestionario supone una herramienta para que el profesor pueda diagnosticar hasta qué punto la evaluación de su asignatura apoya el aprendizaje de sus estudiantes y así, poder establecer cambios en la evaluación. Los ítems y las subescalas del AEQ se desarrollaron a partir de una revisión teórica de la literatura relativa a la promoción del aprendizaje por medio de la evaluación. Esta revisión permitió identificar una serie de condiciones bajo las cuales la evaluación fomentaba el aprendizaje, que se agruparon en cinco apartados: cantidad y distribución del esfuerzo, calidad y nivel de esfuerzo, cantidad y duración del feedback, calidad del feedback y respuesta del estudiante al feedback. Posteriormente, se realizaron entrevistas abiertas a estudiantes sobre su experiencia de evaluación, que dieron lugar a un conjunto potencial de ítems relacionados con los cinco apartados citados anteriormente. A partir de aquí, los autores desarrollaron un cuestionario piloto de seis subescalas de seis ítems cada una. Este cuestionario se administró a dos muestras de dos instituciones diferentes, 1050 y 529 estudiantes, respectivamente. Los resultados de un análisis factorial exploratorio mostraron una solución de seis factores que explicaban el 50% de la varianza y que no coincidían plenamente con los apartados resultantes de la revisión teórica: demanda del tiempo y



LA EVALUACIÓN DEL APRENDIZAJE DE ESTUDIANTES

distribución del esfuerzo, tareas y aprendizaje, cantidad y duración del feedback, calidad del feedback, qué hacer con el feedback y aprendizaje desde el examen. La consistencia interna medida con el alfa de Cronbach osciló entre .44 y .87.

Con el objetivo de mejorar las propiedades psicométricas del AEQ, Gibbs y Dunbar-Goddet (2007) desarrollaron una versión revisada del cuestionario denominada AEQ (V3.3) que contenía cinco escalas del AEQ (cantidad de esfuerzo, cobertura del programa, cantidad y calidad de feedback, uso del feedback y aprendizaje desde el examen), dos escalas del Course Experience Questionnaire (CEQ; Ramsden, 1991), dos escalas adicionales (aproximación profunda y aproximación superficial al estudio) y un ítem de satisfacción general. Esta versión de 28 ítems fue administrada a 516 estudiantes. El análisis factorial exploratorio mostró una estructura coherente con pesos factoriales por encima de .50 y valores alfa de las subescalas entre .61 y .85. Además, el cuestionario fue capaz de discriminar entre dos planes de estudio diferentes.

Puesto que la percepción que tienen los estudiantes de la experiencia de evaluación del aprendizaje es importante para la mejora de los procesos de calidad en educación superior y no se dispone de un cuestionario en español que evalúe esta experiencia, nuestro objetivo es traducir al español y analizar las propiedades psicométricas del cuestionario AEQ (V3.3). Para ello, se llevará a cabo un análisis descriptivo de los ítems del AEQ (V3.3), un análisis de correlaciones y un análisis de regresión lineal múltiple por pasos entre los factores que configuran la escala y el rendimiento académico como variable criterio para analizar la validez predictiva; además, se analizará la evidencia de validez de constructo a través de un análisis factorial confirmatorio y la evidencia de fiabilidad a través de la consistencia interna y la estabilidad temporal.

MATERIAL Y MÉTODOS

Participantes

Los participantes fueron 329 estudiantes universitarios de distintas titulaciones de la Universidad de Las Palmas de Gran Canaria, 286 mujeres y 43 hombres, con una media de edad de 21.70 años ($DT = 3.80$). El tipo de muestreo fue por conglomerado donde la unidad de análisis fue el aula.

Para el análisis de la estabilidad temporal del cuestionario se utilizó una segunda muestra de 56 estudiantes universitarios, 21 hombres y 35 mujeres, con una media de edad de 19.91 años ($DT = 2.36$) que cumplieron el cuestionario en dos ocasiones tras un intervalo de cuatro semanas.



JUAN L. NÚÑEZ / CARMEN I. REYES

Instrumentos

Para medir la experiencia de evaluación del aprendizaje de los estudiantes se utilizó la versión 3.3 del Assessment Experience Questionnaire (AEQ; Gibbs y Dunbar-Goddet, 2007).

Para traducir este cuestionario al español se adoptó la estrategia de traducción inversa (Hambleton, 1996). Se tradujeron los ítems al español y posteriormente otro grupo de traductores volvió a traducirlos al inglés, comparándolos con los originales. Posteriormente, se sometieron los ítems a una evaluación por parte de tres expertos en la temática (Lynn, 1986) que estimaron que los ítems fueron pertinentes para medir el constructo de interés, además de la correcta redacción de los mismos en cuanto a la utilización de la gramática española. Este cuestionario está compuesto por 27 ítems (e.g. utilicé el feedback que recibí para revisar lo que había hecho en mi trabajo) distribuidos en nueve subescalas que miden la experiencia de evaluación del aprendizaje de los estudiantes: cantidad de esfuerzo (ítems 6 y 13), cobertura del programa (ítems 4, 5, 11 y 17), cantidad y calidad de feedback (ítems 3, 15 y 16), uso del feedback (ítems 1, 2 y 8), evaluación apropiada (ítems 10, 14 y 18), claridad de los objetivos y criterios (ítems 7, 9 y 12), aproximación profunda (ítems 20, 21 y 22), aproximación superficial (ítems 19, 23 y 24) y aprendizaje desde el examen (ítems 25, 26 y 27). Además, se incluye un ítem que evalúa la satisfacción en general del estudiante con el curso. Se mantuvo la distribución original de los ítems del cuestionario y la redacción en sentido negativo de 9 ítems (ítems 3, 5, 10, 11, 12, 14, 15, 16 y 18). Las respuestas están expresadas en una escala tipo Likert de 1 (*totalmente en desacuerdo*) a 5 (*totalmente de acuerdo*).

Además, se solicitó a los participantes la nota media final obtenida en las asignaturas presentadas el curso anterior 2010/2011 como medida de rendimiento académico.

Procedimiento

Se requirió la autorización de los profesores, se contactó con los estudiantes durante la jornada académica y se les explicó los objetivos de la investigación. El instrumento se administró en las aulas. Se comunicó a los estudiantes que la participación era voluntaria y confidencial, enfatizando que no había respuestas correctas o incorrectas e instándoles a que la contestaran con honestidad basándose en su experiencia de evaluación del curso anterior 2010/2011. El tiempo empleado para responder fue de aproximadamente de unos 10 minutos.



LA EVALUACIÓN DEL APRENDIZAJE DE ESTUDIANTES

Análisis de datos

En primer lugar, se llevó a cabo un análisis descriptivo de los ítems del AEQ para determinar el cumplimiento de normalidad univariada en el que se utilizó como criterio que la asimetría se situara por debajo del valor 2 y la curtosis por debajo del valor 7 (Curran, West y Finch, 1996). En segundo lugar, se realizaron un análisis de correlaciones de los nueve factores que configuran el cuestionario y la nota media final obtenida por los participantes el curso anterior 2010/2011 y un análisis de regresión lineal múltiple como evidencias de validez predictiva. En tercer lugar, se llevó a cabo un análisis factorial confirmatorio para determinar la validez de constructo del AEQ. Finalmente, se analizó la evidencia de fiabilidad a través de la consistencia interna y de la estabilidad temporal. Para realizar estos análisis se utilizaron los programas estadísticos LISREL 8.54 y PASW 18.

RESULTADOS

Análisis descriptivos

La Tabla 1 presenta los estadísticos descriptivos (media, desviación típica, asimetría y curtosis) de los ítems que configuran el cuestionario AEQ. Como se observa, todos los valores de asimetría y curtosis cumplen con el criterio de normalidad propuesto por Curran et al. (1996), lo que indica semejanza con la curva normal.

Tabla 1: Estadísticos descriptivos de los ítems de la versión española del cuestionario AEQ

ITEM	M	DT	ASIMETRÍA	CURTOSIS
1. Utilicé el feedback que recibí para revisar lo que había hecho en mi trabajo.	3.24	.66	.14	1.52
2. El feedback que recibí me hizo volver a consultar el material visto en el curso.	3.13	.67	.16	1.08
3. Apenas recibí feedback sobre mi trabajo.	3.67	.84	-.63	.61
4. Tuve que estudiar todos los temas del programa para realizar bien la evaluación.	3.55	.93	-.113	-.64
5. El sistema de evaluación me permitió ser bastante selectivo acerca de qué temas del programa debía estudiar.	3.33	.85	.03	-.42
6. Por la forma en la que se planteó la evaluación, tenías que planificarte unas horas de estudio semanales.	3.70	.83	-.02	-.51

JUAN L. NÚÑEZ / CARMEN I. REYES

ITEM	M	DT	ASIMETRÍA	CURTOSIS
7. Fue fácil saber los criterios de evaluación esperados.	3.18	.84	-.13	.36
8. Presté atención al feedback que me dieron los profesores sobre mi trabajo y traté de comprenderlo.	3.52	.83	.08	-.05
9. Los profesores aclararon desde el principio lo que esperaban de sus estudiantes.	3.19	.86	-.16	.26
10. El profesorado parecía más interesado en valorar lo que había memorizado que lo que había comprendido.	3.36	1.04	-.66	-.10
11. Pude ser bastante estratégico y dejar de estudiar algunos temas.	3.63	.85	-.29	-.35
12. A veces fue difícil saber lo que se esperaba de mí en este curso.	3.25	.86	-.21	-.07
13. En este curso fue necesario trabajar de forma constante para cumplir los requisitos de la evaluación.	3.85	.87	-.12	-.96
14. El profesorado muchas veces me preguntó sólo cuestiones de datos memorísticos.	3.23	.94	-.59	-.05
15. No comprendí algunos aspectos del feedback que me dieron los profesores sobre mi trabajo.	3.41	.77	-.41	.49
16. Todo el feedback que me dieron sobre mi trabajo llegó demasiado tarde para ser útil.	3.67	.79	-1.08	1.92
17. Por la forma de evaluar las asignaturas de este curso tenía que estudiar todos y cada uno de los temas.	3.26	.96	.16	-.47
18. Para tener éxito en este curso todo lo que se necesitaba era tener memoria.	3.44	.99	-.66	.029
19. Cuando leía trataba de memorizar los datos importantes que podrían ser útiles más adelante.	3.30	.93	-.13	-.173
20. Generalmente me proponía comprender a conciencia el significado de lo que me pedían que leyera.	3.62	.78	.10	-.32
21. Normalmente me esforzaba en comprender lo que inicialmente parecía difícil.	3.70	.77	-.05	-.22
22. Durante el curso, a veces me cuestioné cosas que se decían en clase o que leía.	3.47	.83	.08	-.22
23. Tenía que concentrarme en memorizar gran cantidad de información que tenía que aprender.	3.14	.94	.22	-.44

LA EVALUACIÓN DEL APRENDIZAJE DE ESTUDIANTES

ITEM	M	DT	ASIMETRÍA	CURTOSIS
24. A veces tenía que estudiar cosas sin tener oportunidad de comprenderlas.	3.13	.97	.27	-.41
25. Hacer los exámenes me ayudó a relacionar e integrar los temas.	3.04	.88	.07	-.01
26. Aprendí cosas nuevas mientras preparaba los exámenes.	3.46	.89	-.05	-.29
27. Comprendí mejor los temas después de haber realizado los exámenes.	3.07	.95	.18	-.41
28. En general, estoy satisfecho con la enseñanza de este curso.	3.52	.86	-.17	.23

Análisis de correlaciones

Se analizó la relación entre las nueve subescalas del AEQ y el rendimiento académico como variable criterio utilizando el coeficiente de correlación de Pearson. Tal y como se observa en la Tabla II, los coeficientes son bajos, si bien, la nota media final obtenida por los estudiantes en el curso se relaciona de forma positiva y significativa con el uso del feedback recibido ($r = .21$; $p < .01$), la cobertura del programa ($r = .16$; $p < .01$), la claridad de los objetivos ($r = .15$; $p < .01$) y la aproximación profunda ($r = .14$; $p < .05$).

Tabla 2: Correlaciones entre las subescalas del AEQ y la nota media y alfas en la diagonal

	1	2	3	4	5	6	7	8	9	10	
1. Nota media		.10	.16**	.06	.21**	.07	.15**	-.08	.14*	.03	
2. Esfuerzo			.73	.21**	-.05	.26**	-.18**	.05	.25**	.31**	.13*
3. Programa				.70	-.15**	.03	-.15**	-.07	.13*	.15**	.05
4. Cant. feb					.74	.19**	.43**	.17*	-.39**	-.18**	.13*
5. Uso feb						.72	.15**	.28**	-.02	.21**	.30**
6. Evaluación							.73	.24**	-.58**	-.20**	.14*
7. Objetivos								.71	-.11*	.05	.11
8. Aprox. Superf.									.70	.26**	.09
9. Aprox. Prof.										.72	.18**
10. Ap. Examen											.74*

$p < .05$; ** $p < .01$.



JUAN L. NÚÑEZ / CARMEN I. REYES

Análisis de regresión lineal múltiple

El análisis de correlaciones permite una primera aproximación a la validez predictiva que se estudia con más precisión mediante el análisis de regresión múltiple. Llevamos a cabo un análisis de regresión lineal múltiple por pasos entre las nueve subescalas del AEQ (variables independientes) y el rendimiento académico (variable dependiente), tomando la nota media de las calificaciones obtenidas por los estudiantes en las asignaturas presentadas en el curso académico 2010/2011.

Los resultados obtenidos a partir del análisis de regresión realizado muestran que, de las nueve subescalas, tres entran en la ecuación y, por tanto, tienen la capacidad de predecir de forma positiva y significativa ($p < .05$) el rendimiento académico de los estudiantes, en concreto: uso del feedback ($\beta = .18$), cobertura del programa ($\beta = .17$) y claridad de objetivos ($\beta = .12$). Los resultados obtenidos en el análisis de varianza de la regresión muestran un valor de F significativo ($F = 9.50$; $p < .05$). El coeficiente R^2 fue de .12, lo que supone que el conjunto de estas tres subescalas explica el 12% de la varianza de las calificaciones académicas.

Análisis factorial confirmatorio

La estructura factorial de la escala AEQ fue evaluada con un análisis factorial confirmatorio utilizando el procedimiento estándar de mínimos cuadrados ponderados diagonalizados y la matriz de correlaciones policóricas como entrada para el análisis de datos puesto que los ítems son variables categóricas o discretas que se responden en una escala tipo Likert (Flora y Curran, 2004). Se utilizó este método de estimación ya que no tiene limitaciones respecto al tamaño de la muestra y no requiere normalidad multivariada (Olatunji et al., 2007). Los índices de ajuste obtenidos fueron los siguientes: $S - B \chi^2 / gl = 1.93$ ($S - B \chi^2 (288, N = 329) = 555.35, p = .00$); IFI = .90; SRMR = .06 y RMSEA = .05 (LO 90 = .04; HI 90 = .06). Con respecto a los pesos factoriales de los ítems, oscilaron entre .43 del ítem 5 “el sistema de evaluación me permitió ser bastante selectivo acerca de qué temas del programa debía estudiar” y .86 del ítem 21 “normalmente me esforzaba en comprender lo que inicialmente parecía difícil”, siendo todos ellos significativos ($p < .05$).

Fiabilidad

La consistencia interna de las nueve subescalas de la AEQ fue evaluada a través del alfa de Cronbach. Así, podemos observar en la diagonal de la Tabla 2 que los valores obtenidos se situaron entre el .70 de las subescalas cobertura del programa y



LA EVALUACIÓN DEL APRENDIZAJE DE ESTUDIANTES

aproximación superficial y el .74 de las subescalas cantidad y calidad de feedback y aprendizaje desde el examen.

Para evaluar la estabilidad temporal de la AEQ se utilizó una muestra de 56 estudiantes de educación superior que cumplimentaron la escala en dos ocasiones con un intervalo temporal de cuatro semanas. Se determinó la estabilidad temporal mediante una prueba no paramétrica como el coeficiente de correlación de Spearman, ya que el tamaño de la muestra no permite utilizar una prueba más robusta. Los coeficientes de correlación de Spearman entre las puntuaciones del test y del retest oscilaron entre .52 de la subescala evaluación apropiada y .79 de la subescala aproximación profunda.

DISCUSIÓN

La EOA destaca la importancia de la evaluación en la optimización del aprendizaje. El cuestionario AEQ fue desarrollado para examinar el impacto de la evaluación sobre la experiencia de aprendizaje de los estudiantes. Ante la ausencia de una medida de estas características validada al español, el objetivo del presente estudio fue traducir al español y analizar las propiedades psicométricas del cuestionario AEQ en una muestra de estudiantes universitarios.

La traducción se realizó siguiendo la estrategia de traducción inversa en la misma línea que estudios anteriores (Martín-Albo, Núñez y Navarro, 2009; Núñez, Martín-Albo y Navarro, 2005).

El análisis de correlaciones entre las nueve subescalas del AEQ y el rendimiento académico como variable criterio mostró relaciones positivas entre el rendimiento y todas las subescalas, excepto con la subescala aproximación superficial. En este sentido, unos altos niveles de aprendizaje basados en la memorización de contenidos y no tanto en la comprensión de los mismos se relacionan con una nota académica baja. El análisis de regresión lineal múltiple por pasos mostró que las subescalas uso del feedback, cobertura del programa y claridad de objetivos tienen capacidad para predecir el rendimiento académico explicando el 12% de las calificaciones académicas. Estos resultados apoyan parcialmente la validez predictiva del cuestionario AEQ y revelan que deben considerarse otras variables académicas de interés, además de la experiencia de evaluación del aprendizaje de los estudiantes, como por ejemplo el clima que genera un profesor en el aula, el tipo de motivación del estudiante o el uso de estrategias cognitivas y metacognitivas para explicar en mayor medida una variable como el rendimiento académico; asimismo, ponen de manifiesto la importancia que tiene para el estudiante tener claros desde el principio los objetivos del programa, la cantidad de temas que se estudian y las respues-

tas positivas de aprendizaje (Turner y Gibbs, 2010), estando en consonancia con diversas investigaciones que destacan la influencia del feedback sobre el rendimiento de los estudiantes (Hattie, 1987, Hattie y Timperley, 2007). En términos parecidos, Black y Wiliam (1998) subrayan los efectos positivos del feedback sobre el aprendizaje y el rendimiento en los diferentes niveles de educación.

Para determinar la validez de constructo se analizó la estructura factorial del cuestionario AEQ mediante un análisis factorial confirmatorio. Los resultados confirmaron la estructura de nueve factores correlacionados, lo que corrobora la estructura propuesta por Gibbs y Dunbar-Goddet (2007) a través de un análisis factorial exploratorio. Además, los pesos factoriales de los ítems se situaron por encima de .43.

En general, los resultados obtenidos mostraron que las subescalas del AEQ presentan unos valores aceptables de consistencia interna de acuerdo a los criterios de George y Mallery (2003) y, en general, más altos que los obtenidos por Gibbs y Simpson (2003) en la primera versión del cuestionario y por Gibbs y Dunbar-Goddet (2007) en la versión 3.3. Los valores obtenidos en la correlación test-retest del cuestionario muestran una estabilidad temporal que podemos calificar de moderada-alta.

Debemos tener en cuenta algunas limitaciones. Los participantes de este estudio fueron estudiantes universitarios, por lo que se deberían analizar las propiedades psicométricas del cuestionario AEQ en otros niveles académicos. Asimismo, las evidencias de validez y fiabilidad deberían ser consideradas como provisionales dado que el tamaño de la muestra, sobre todo en el caso de los hombres, fue pequeño. Futuros estudios deberían comprobar los efectos de género, así como analizar la invarianza del cuestionario en función del género y el poder discriminante del instrumento en diferentes entornos de evaluación (e.g. evaluación formativa, evaluación sumativa, etc.). En próximas investigaciones se debería considerar el uso de ítems redactados en negativo; la redacción en positivo favorece la comprensión del ítem y podría mejorar el comportamiento psicométrico de algunas subescalas como cantidad y calidad de feedback y evaluación apropiada. Además, sería interesante desarrollar una versión reducida del cuestionario de acuerdo a las recomendaciones de Marsh, Martin y Jackson (2010) para seleccionar ítems (peso factorial, correlación ítem-test, alfa de Cronbach e importancia teórica del ítem) con la finalidad de reducir el tiempo de aplicación con las mismas garantías psicométricas.

En conclusión, los resultados mostraron que la versión española del cuestionario AEQ presenta evidencias de validez y fiabilidad, en la misma línea que los resultados obtenidos en las versiones de Gibbs y Simpson (2003) y de Gibbs y Dunbar-Goddet (2007). Por tanto, la versión española puede ser considerada como una adaptación preliminar y los resultados justifican su utilización en el contexto de

LA EVALUACIÓN DEL APRENDIZAJE DE ESTUDIANTES

la educación superior con distintos fines. En primer lugar, podría ser utilizado como herramienta de evaluación para examinar las experiencias de evaluación de los estudiantes e indicar si éstas están orientadas al aprendizaje. Posteriormente, esto podría conducir a mejoras y cambios en el diseño de las asignaturas. Además, también podría poner de manifiesto las diferencias entre los diferentes contextos de asignaturas. Asimismo, el cuestionario AEQ podría ser utilizado para evaluar cuestiones específicas como, por ejemplo, el feedback o la claridad de los objetivos e identificar áreas donde existan lagunas y, en consecuencia, donde deberían introducirse mejoras. Por último, también podría servir como instrumento de investigación para explorar en mayor profundidad las relaciones entre el aprendizaje de los estudiantes y las características de sus ambientes o climas de aprendizaje o las relaciones entre las experiencias de EOA y las características de los estudiantes.

Fecha de recepción del original: 28 de enero de 2013

Fecha de aceptación de la versión definitiva: 14 de abril de 2013

REFERENCIAS

- Biggs, J. (2005). *Calidad del aprendizaje universitario*. Narcea: Madrid.
- Black, P. y William, D. (1998). Assessment and classroom learning. *Assessment in Education*, 5(1), 7-74.
- Carless, D. (2003, Noviembre). *Learning-oriented assessment*. Comunicación presentada en la Evaluation and Assessment Conference, University of South Australia, Adelaide.
- Carless, D. (2007). Learning-oriented assessment: Conceptual bases and practical implications. *Innovations in Education and Teaching International*, 44(1), 57-66.
- Carless, D. (2009). Learning-oriented assessment: Principles, practice and a project. En L. H. Meyer, S. Davidson, H. Anderson, R. Fletcher, P. M. Johnston y M. Rees (Eds.), *Tertiary Assessment & Higher Education Student Outcomes: Policy, Practice & Research Learning-oriented assessment* (pp.79-90). Wellington, New Zealand: Ako Aotearoa.
- Carless, D., Salter, D., Yang, M. y Lam J. (2011). Developing sustainable feedback practices. *Studies in Higher Education*, 36(4), 395-407.
- Curran, P. J., West, S. G. y Finch, J. F. (1996). The robustness of test statistics to nonnormality and specification error in confirmatory factor analysis. *Psychological Methods*, 1(1), 16-29.
- Flora, D. B. y Curran, P. J. (2004). An empirical evaluation of alternative methods

- of estimation for confirmatory factor analysis with ordinal data. *Psychological Methods*, 9(4), 466-491.
- George, D. y Mallery, P. (2003). *SPSS for Windows step by step: A simple guide and reference. 11.0 update*. Boston: Allyn & Bacon.
- Gibbs, G. (2006). How assessment frames student learning. En C. Bryan y K. Clegg (Eds.), *Innovative assessment in higher education* (pp. 23-36). London: Routledge.
- Gibbs, G. y Dunbar-Goddet, H. (2007). *The effects of programme assessment environments on student learning*. York: Higher Education Academy.
- Gibbs, G. y Simpson, C. (2003, Septiembre). *Measuring the response of students to assessment: The Assessment Experience Questionnaire*. Comunicación presentada en el 11th International Improving Students Learning Symposium. Hinckley.
- Gibbs, G. y Simpson, C. (2004). Conditions under which assessment supports student learning. *Learning and Teaching in Higher Education*, 1, 3-31.
- Hattie, J. (1987). Identifying the salient facets of a model of student learning: A synthesis of meta-analyses. *International Journal of Educational Research* 11, 187-212.
- Hattie, J. y Timperley, H. (2007). The power of feedback. *Review of Educational Research* 77(1), 81-112.
- Hambleton, R. K. (1996). Adaptación de test para su uso en diferentes idiomas y culturas: fuentes de error, posibles soluciones y directrices prácticas. En J. Muñoz (Ed.), *Psicometría* (pp. 67-89). Madrid: Universitas.
- Lynn, M. (1986). Determination and quantification of content validity. *Nursing Research*, 35, 382-85.
- Ibarra, M. S. y Rodríguez, G. (2010). Aproximación al discurso dominante de la evaluación del aprendizaje en la universidad. *Revista de Educación*, 344, 25-375.
- Marsh, H., Martin, A. y Jackson, S. (2010). Introducing a short version of the Physical Self-Description Questionnaire: New strategies, short-form evaluative criteria, and applications of factor analyses. *Journal of Sport and Exercise Psychology*, 32, 438-482.
- Martín-Albo, J., Núñez, J. L. y Navarro, J. G. (2009). Validation of the Spanish version of the Situational Motivation Scale (EMSI) in the educational context. *The Spanish Journal of Psychology*, 12(2), 799-807.
- Mateo, J. (2006). Claves para el diseño de un nuevo marco conceptual para la medición y evaluación educativas. *Revista de Investigación Educativa*, 24(1), 163-186.
- McDowell, L., Wakelin, D., Montgomery, C. y King S. (2011). Does assessment for learning make a difference? The development of a questionnaire to explore the student response? *Assessment & Evaluation in Higher Education*, 36(7), 749-765.

LA EVALUACIÓN DEL APRENDIZAJE DE ESTUDIANTES

- Núñez, J. L., Martín-Albo, J. y Navarro, J. G. (2005). Validación de la versión española de la Échelle de Motivation en Éducation. *Psicothema*, 17(2), 344-349.
- Olatunji, B. O., Williams, N. L., Tolin, D. F., Abramowitz, J. S., Sawchuk, C. N., Lohr, J. M., y Elwood, L. S. (2007). The disgust scale: Item analysis, factor structure, and suggestions for refinement. *Psychological Assessment*, 19(3), 281-297.
- Ramsden, P. (1991). A performance indicator of teaching quality in higher education: The Course Experience Questionnaire. *Studies in Higher Education*, 16, 129-150.
- Reyes, C., Sosa, F., Marchena, R. y Marchena, J. (2013). Sistema de evaluación en las prácticas clínicas de la titulación de Medicina. *Revista de Educación Médica*, 16(1), 7-11.
- Richardson, J. (2005). Instruments for obtaining student feedback: A review of a literature. *Assessment & Evaluation in Higher Education*, 30(4), 387-415.
- Turner, G. y Gibbs, G. (2010). Are assessment environments gendered? An analysis of the learning responses of male and female students to different. *Assessment & Evaluation in Higher Education*, 35(6), 687-698.