

La exigencia cognitiva en los exámenes tipo test en contexto universitario y su relación con los enfoques de aprendizaje, la autorregulación, los métodos docentes y el rendimiento académico

Pilar Gandía Herrero y Agustín Romero Medina
Universidad de Murcia (España)

La calidad del rendimiento académico y los resultados de aprendizaje dependen de diversos factores, tanto de tipo psicológico como de tipo contextual. El contexto académico incluye las actividades formativas y el tipo de evaluación o examen, sobre el cual también influyen factores cognitivos y motivacionales, tales como los enfoques de aprendizaje y estudio y la autorregulación. En nuestro contexto universitario el tipo de examen predominante es el de preguntas de opción múltiple. La exigencia cognitiva de estas preguntas puede variar. Desde la típica taxonomía de Bloom se considera que de menor a mayor exigencia cognitiva tenemos preguntas sobre conocimiento factual, conceptual, de aplicación, etc. Normalmente el profesor no tiene en cuenta estas clasificaciones a la hora de elaborar este tipo de examen. Se propone aquí un modelo de adaptación de la clasificación de preguntas de opción múltiple según exigencia cognitiva (memorización asociativa, comprensión, aplicación), poniéndolo a prueba analizando un examen de una asignatura de Grado en Psicología y relacionando los resultados con medidas de enfoques de aprendizaje (cuestionarios ASSIST y R-SPQ-2F) y autorregulación en una muestra de 87 sujetos. Los resultados muestran rendimiento académico diferencial según tipos "cognitivos" de preguntas y diferencias en enfoques de aprendizaje y autorregulación. Se subraya la conveniencia de tener en cuenta estos factores de exigencia cognitiva a la hora de elaborar preguntas de opción múltiple.

Palabras Clave: Enfoques de aprendizaje, autorregulación, estudiantes universitarios, rendimiento académico, evaluación, preguntas de opción múltiple.

The cognitive requirement in the multiple-choice exams in the university context and its relationship with learning approaches, self-regulation, teaching methods and academic performance. The quality of academic performance and learning outcomes depend on various factors, both psychological and contextual. The academic context includes the training activities and the type of evaluation or examination, which also influences cognitive and motivational factors, such as learning and study approaches and self-regulation. In our university context, the predominant type of exam is that of multiple-choice questions. The cognitive requirement of these questions may vary. From Bloom's typical taxonomy, it is considered that from lower to higher cognitive demand we have questions about factual, conceptual, application knowledge, etc. Normally, the teacher does not take these classifications into account when preparing this type of exam. We propose here an adaptation model of the multiple choice questions classification according to cognitive requirement (associative memorization, comprehension, application), putting it to the test analyzing an examination of a subject in Psychology Degree and relating the results with measures of learning approaches (ASSIST and R-SPQ-2F questionnaires) and self-regulation in a sample of 87 subjects. The results show differential academic performance according to "cognitive" types of questions and differences in approaches to learning and self-regulation. The convenience of taking into account these factors of cognitive requirement when elaborating multiple choice questions is underlined

Keywords: Approaches to learning, self-regulation, university students, academic performance, multiple-choice questions.

Correspondencia: Agustín Romero Medina. Universidad de Murcia. Departamento de Psicología Básica y Metodología. Facultad de Psicología. Campus de Espinardo. C.P.: 30100. Murcia (España). E-mail: agustinr@um.es

El conocimiento sobre un dominio académico universitario se va adquiriendo a lo largo de un periodo de tiempo delimitado, generalmente un cuatrimestre. Se adquiere con actividades formativas diversas, tanto presenciales (clases, prácticas) como no presenciales (en ámbito privado, sobre todo en sesiones de estudio) y el nivel y tipo de conocimiento y competencia adquiridos se comprueba en la evaluación, la cual generalmente toma la forma de examen.

Intervienen muchos factores en este proceso. Así, están presentes variables cognitivo-emocionales tanto en el momento de aprender como en el de evaluar dicho aprendizaje. Durante el aprendizaje tenemos variables tales como: (1) conocimiento previo específico de la propia temática o contenidos objeto de estudio, (2) conocimiento procedimental estratégico sobre cómo adquirir la información, que permite comprender mejor o peor el nuevo material; (3) variables psicológicas tales como los “enfoques de aprendizaje” del alumnado (un punto en un continuo entre superficial y profundo, en la terminología clásica de Marton, Entwistle, Biggs, etc.) o la autorregulación; (4) las variables motivacionales-afectivas, que provocan una mayor o menor dedicación y esfuerzo en el estudio y en el momento del examen (ansiedad, expectativas de fracaso o éxito, etc.); y (5) el sistema de evaluación, que impulsa todo el proceso de aprendizaje y por tanto afecta a la calidad de los resultados de aprendizaje logrado por el alumnado (Hubbard, 1997; Pérez, Carretero, Palma, y Rafel, 2000).

En este propósito de analizar la calidad de la evaluación en cuanto a su eficacia para conocer el aprendizaje, los instrumentos operativos más utilizados han sido los de tipo taxonómico, es decir identificar tipos de exámenes y de preguntas de examen. Una primera aproximación ha sido la de taxonomías de tipos de exámenes según el tipo de respuesta requerida: de ensayo, de respuesta breve, de elección múltiple, etc. (Pérez et al., 2000). En este caso, la clasificación se fija en las ventajas e inconvenientes de cada tipo, en términos logísticos (duración de examen, dificultad de elaboración, tiempo de corrección, objetividad o sesgos en la corrección) y muy poco en términos de efectos en el tipo de aprendizaje al que obliga (más asociativo o memorístico en el tipo de elección múltiple, más constructivo y comprensivo en el de desarrollo escrito). El tipo de examen más habitual para evaluar los contenidos teóricos en muchas titulaciones de la universidad española es el “tipo test”, o sea, examen de preguntas de opción o elección múltiple de respuesta (ya sea verdadero/falso, tres opciones de respuesta o incluso más), en adelante examen PEM (de preguntas de elección múltiple, en inglés, multiple choice question, MCQ). No requiere elaboración escrita, pues ya viene dada en las opciones de respuesta y esto activa en el sujeto básicamente procesos de reconocimiento de la opción correcta, aunque según el tipo de pregunta puede exigir procesos cognitivos más complejos. Tradicionalmente es un tipo de evaluación muy criticado pues sobre todo el requerimiento cognitivo es básicamente de memoria reproductiva y de reconocimiento y por tanto evidencia un aprendizaje asociativo memorístico y poco comprensivo o

constructivo; aun así, en España, con una Universidad masificada en muchas disciplinas, es el más utilizado, por su fácil corrección y nulo sesgo subjetivo en la corrección.

Otras taxonomías se han centrado en el ámbito cognitivo, es decir, la exigencia mayor o menor de complejidad y esfuerzo cognitivo para responder a las preguntas de examen, aunque al menos en el contexto universitario español son muy poco frecuentes los casos de diseño de preguntas según estas taxonomías. De hecho, lo habitual es que los profesores elaboren sus preguntas de examen de modo no planificado y con arreglo a su experiencia cotidiana o a las tradiciones y cultura de evaluación típica del sistema educativo universitario. En este contexto, el nivel de exigencia y dificultad de las preguntas lo decide típicamente el profesor en base a su experiencia como docente y al tipo de contenido en el que es especialista y de lo que es enseñante, o en base a sus creencias acerca del grado de dificultad de las preguntas: unos piensan que serán más difíciles las preguntas de datos, las de apartados poco explicados del temario; otros piensan lo contrario, que lo más difícil radicaría en la complejidad de los conceptos. Desde luego no suele decidirlo en base a estudios empíricos que le permitan determinar la utilidad en función de los objetivos o competencias formativas buscadas en la asignatura, aun cuando las recomendaciones incluidas en el Plan Bolonia, seguidas en nuestro país por ANECA (2013), ante todo insisten en la alineación o coherencia entre las exigencias cognitivas de los objetivos formativos de las asignaturas o la titulación y las de los sistemas de evaluación que se utilicen. Sin duda, esta alineación se perfeccionaría si se tuviese en cuenta esta coherencia de exigencia cognitiva que requiere aprendizaje y evaluación. Es un mundo poco explorado, pero que indudablemente puede tener repercusión importante en la calidad del aprendizaje del alumnado.

El punto de partida teórico de estas taxonomías cognitivas asume que las actividades y contenidos formativos y de estudio, así como el tipo de pregunta o reactivo de examen provocan distintos grados de exigencia o rigor cognitivo en el aprendizaje del alumno. Es decir, obligan al aprendiz a realizar un determinado tipo de operación cognitiva sobre la información o contenidos (sea formativo o de evaluación) para que se conviertan en conocimiento (en la memoria a largo plazo) y en competencia (demostración de que se tiene ese conocimiento). Se han propuesto diversas taxonomías de exigencia cognitiva o niveles de profundidad de conocimiento. La más conocida ha sido la ya clásica de Bloom et al. (1956) revisada en 1982 (taxonomía SOLO -Structure of the Observed Learning Outcome-) por Biggs y Collis, y en 2001 por Anderson et al. Aunque sería discutible cómo se han fundamentado estas taxonomías en la psicología cognitiva (probablemente se han seguido criterios heurísticos o intuitivos), lo cierto es que han supuesto un avance operativo para el diseño de actividades formativas y también, aunque menos estudiado, para refinar los sistemas de evaluación, con el fin de alcanzar ese alineamiento o coherencia entre ambos (ANECA, 2013).

La taxonomía revisada de Bloom de 2001 (Anderson et al.) es de las más extendidas en la actualidad y propone seis tipos de conocimiento ordenados de menor a mayor exigencia cognitiva: recordar, comprender, aplicar, analizar, evaluar y crear. Los dos primeros son los de menos nivel y los típicos en los exámenes PEM, los tres siguientes también pueden aparecer en los exámenes PEM y el último (crear) no es apropiado para ello.

El procedimiento habitual para aplicar las taxonomías de exigencia cognitiva de preguntas de examen requiere diseñarlas para que encajen en los niveles de complejidad acordes con las actividades formativas y objetivos de aprendizaje de las asignaturas. Esto obliga a un cierto entrenamiento previo para el profesor, pero tiene la ventaja que posibilita una coherencia formativa muy necesaria para la calidad del aprendizaje, y con los años se mejora e incluso sube el nivel de exigencia cognitiva (Kattan et al., 2014). De hecho, en el procedimiento estándar (sin ningún criterio taxonómico) de elaboración de exámenes PEM, en general la demanda cognitiva es baja, es decir, niveles 1 y 2 de la taxonomía de Bloom (Cárdenas, Blanco, y Cáceres, 2014).

A partir de este marco teórico y metodológico se han sucedido distintas investigaciones tendentes a proponer exámenes PEM que sean capaces de evaluar conocimiento complejo basado en aprendizaje comprensivo y habilidades de razonamiento (Kominsky, 2012; Scully, 2017; Stanger-Hall, 2012; Su, Osisek, Montgomery, y Pellar, 2009).

Por otra parte, si los exámenes suben el nivel de exigencia cognitiva, sin duda eso favorecería la utilización de estrategias y enfoques de aprendizaje menos superficiales (memorísticos) y más profundos o al menos estratégicos. De hecho, en investigaciones en las que se han analizado los enfoques de aprendizaje y el tipo de evaluación (Scouller, 1998; Yonker, 2011), se ha podido evidenciar que las preguntas de desarrollo fomentan enfoques y estrategias de aprendizaje profundos mientras que las de opción múltiple favorecen enfoques superficiales. Centrándonos en exámenes PEM, Yonker (2011) examinó a 56 sujetos con preguntas de elección múltiple de tipo factual y de aplicación y les pasó un cuestionario de enfoques de aprendizaje (R-SPQ-2F, de Biggs); en ambos tipos de preguntas no hay correlación de rendimiento con enfoque profundo; y con enfoque superficial presenta correlación negativa con rendimiento, aunque es más fuerte con preguntas factuales que aplicadas.

Hay una relación mutua todavía poco explicada entre enfoques de aprendizaje y factores de evaluación. En la literatura sobre enfoques de aprendizaje (profundo, superficial, estratégico), estos se modulan según diversos factores del contexto de aprendizaje. Así, por ejemplo, parece claro que el estilo o enfoque profundo de aprendizaje se verá potenciado con una metodología docente activa que fomente la comprensión, el análisis y debate crítico y la reflexión, la motivación por aprender y el compromiso (engagement) con la institución académica. Por el contrario, una

metodología de lección magistral rutinaria, con exigencia muy memorística por exámenes que sobre todo obliguen a memorización asociativa llevarán a enfoques superficiales dirigidos básicamente a aprobar para obtener el título. Al comenzar las clases en una determinada asignatura o en un curso, el estudiante ya sabe qué tipo de exigencia de aprendizaje y motivación va a requerir y por tanto su enfoque de aprendizaje se adaptará a ello.

Objetivos

El objetivo general de esta investigación consiste en operativizar un procedimiento para poder alinear actividades formativas, objetivos de aprendizaje y evaluación, centrándonos en ésta última, y en concreto en la parte probablemente más difícil como es el diseño de exámenes PEM, siguiendo principios y metodología de adscripción a taxonomías de exigencia cognitiva de las preguntas.

Para ello se buscarían los siguientes objetivos específicos: (1) comprobar el tipo de exigencia cognitiva en exámenes PEM elaborados sin criterio definido para ver si en nuestro contexto también se cumple que evalúan baja exigencia cognitiva (recuerdo y comprensión), aplicando la típica taxonomía de Bloom revisada; (2) analizar la relación entre exigencia cognitiva de las preguntas y rendimiento o éxito alcanzado en las notas de los estudiantes a los que se les aplicó en situación real, teniendo en cuenta también indicadores objetivos de dificultad y discriminación de cada pregunta; y (4) por último, se debería tener en cuenta la relación de la exigencia cognitiva demostrada con el estilo o enfoque de aprendizaje y autorregulación del alumnado; la evaluación es un factor del contexto de aprendizaje que puede influir muy decisivamente en el tipo de enfoque general de aprendizaje y en el tipo de estrategias cognitivas y de autorregulación necesarias para obtener un óptimo resultado de aprendizaje.

MÉTODO

Participantes

Participaron un total de 87 estudiantes de un total de 188 matriculados en una asignatura obligatoria de 6 créditos de primer cuatrimestre de 2º curso del Grado de Psicología en la Universidad de Murcia (curso 2017/18). De los 188 se presentaron 117 al examen final, y de estos sólo 87 pudieron además cumplimentar cuestiones de enfoques de aprendizaje y autorregulación. La media de edad es de 20.4 años ($DT= 3.88$) y de ellos tenemos 66 (75.86%) mujeres y 21 (24.13%) hombres.

Instrumentos

Se disponía de un examen de 50 preguntas de opción múltiple, de cuatro opciones de respuesta, que se aplicó como examen final de los contenidos teóricos de la

asignatura. Para la investigación finalmente se dispuso de los resultados de 47 preguntas de examen, después de anular tres por defectos de forma. El examen se realizó sobre formulario de papel para lector óptico, de modo que fue corregido automáticamente en el Servicio de Informática de la Universidad. Esta corrección proporciona, por cada sujeto, datos de aciertos, errores, respuestas en blanco, nota numérica sobre 10 en base a la fórmula estándar de $\{[\text{aciertos} - (\text{errores} / \text{n}^\circ \text{ de opciones de respuesta} - 1)] / \text{n}^\circ \text{ de preguntas}\} * 10$; por cada pregunta proporciona el índice de dificultad e índice de discriminación, con arreglo a las siguientes fórmulas:

- Índice de dificultad: $[(\text{n}^\circ \text{ errores} + \text{pregunta en blanco}) / \text{n}^\circ \text{ estudiantes presentados}] * 100$. Si la pregunta se ha dejado en blanco o se ha contestado mal muchas veces se tiene un valor alto de dificultad, y bajo en el caso contrario.

- Índice de discriminación: $[(A - B) / C] * 100$, siendo A= los aciertos del 25% de sujetos con mejores notas (grupo “fuerte”); B= los aciertos del 25% de sujetos con peores notas (grupo “débil”); C= 25% del total de alumnos presentados. Por ejemplo, en la pregunta 2 en el grupo A se dan 8 aciertos, 6 aciertos en el B, el 25% de 87 es 21.75; aplicando la fórmula, el índice de discriminación es $[(8 - 6) / 21.75] * 100 = 9.19$. El índice de discriminación obtiene valores entre -100 y 100 y es un parámetro con una larga trayectoria de uso pues fue propuesto por Kelley en 1939. Un valor alto y positivo es lo esperado de una buena pregunta: que la contesten bien los que más han estudiado y tengan más errores los que menos han estudiado y, por lo tanto, discrimina positivamente. Un valor negativo es justo lo contrario y un valor cerca de cero no sirve para diferenciar ambos grupos. En definitiva, como dicen Doval y Renom (2010, p. 76), “las respuestas al ítem nos permiten discriminar entre las personas que tienen un conocimiento alto y bajo”. Un índice de discriminación entre 20-29 es regular, de 30 a 39 aceptable y de 40 a 100 excelente. Otra clasificación distingue entre mayor o igual a 40 es excelente, bueno entre 30 y 39, satisfactorio entre 15 y 29, y bajo si es inferior a 15.

Para clasificar la exigencia cognitiva de las preguntas se utilizó la taxonomía de Bloom revisada por Anderson et al. (2001). Dos profesores entrenados adscribieron cada pregunta a una de las seis categorías de procesos cognitivos implicados en el aprendizaje y en la evaluación, de menor a mayor dificultad: recordar (literalmente), comprender, aplicar, analizar, evaluar y crear/sintetizar.

Para los enfoques de aprendizaje se ha utilizado el Approaches and Study Skills Inventory for Students (ASSIST) en su versión reducida en español. Fue elaborado originariamente en inglés por Tait, Entwistle, y McCune (1998). Es un cuestionario de autoinforme que evalúa la autopercepción del uso de estrategias, habilidades y enfoques de aprendizaje en entorno académico. Consta de 18 ítems con respuesta tipo Likert que oscila desde 1= Totalmente en desacuerdo a 5= Totalmente de acuerdo. Los 18 ítems están clasificados en tres factores: enfoque profundo (Factor I), enfoque superficial

(Factor II), enfoque estratégico (Factor III). Se ha utilizado la versión traducida al español y validada por Hidalgo et al. (2009), encontrando niveles adecuados de fiabilidad mediante el coeficiente alfa de Cronbach para los tres factores (Factor I= 0.74, Factor II= 0.68, Factor III= 0.65).

Para la autorregulación académica se aplicó el Cuestionario de Autorregulación en el estudio. Se trata de un cuestionario elaborado por el segundo de los autores y consta de 22 ítems distribuidos en cuatro bloques siguiendo las cuatro fases de la tarea de estudio (antes de empezar a estudiar, mientras estudia, después de estudiar y tras la nota del examen). Las respuestas pertenecen a diferentes escalas de medida (tipo Likert y dicotómicas). El cuestionario trata de medir la capacidad autorregulación del estudiante en las diferentes fases de estudio y en el total. Este cuestionario presenta en estos participantes una fiabilidad mediante el coeficiente alfa de Cronbach de .607, siendo una fiabilidad aceptable pero baja. También se le ha aplicado el modelo de Rasch (Abellán-Pérez, 2018), pudiendo comprobar que los ítems se ajustan a los principios de dicho modelo.

Procedimiento y análisis de datos

Estos cuestionarios se pasaron a los sujetos un mes antes del día del examen. La explicación de la prueba fue realizada en el horario habitual de clases, contando previamente con la autorización institucional y del profesor a cargo del curso. Los estudiantes cumplieron posteriormente un archivo Word con todos los cuestionarios y preguntas, así como el consentimiento informado donde se detallaba el objetivo de la investigación y se comunicaba el carácter voluntario, anónimo y sin retribución de la participación. Dicho archivo ya cumplimentado permitió el análisis de los datos mediante el paquete estadístico SPSS v. 24. Por la realización de la investigación fueron recompensados con bonos canjeables por décimas que podían usarse para la suma de la nota final de asignaturas del Grado.

El examen de la asignatura siguió el procedimiento estándar en la convocatoria oficial de exámenes de la Universidad de Murcia y los datos de calificaciones han seguido los protocolos típicos en cuanto a materia de protección de datos. Desde el principio de curso los estudiantes tuvieron información detallada sobre el tipo de examen, haciendo énfasis en que bastantes preguntas exigían sobre todo un estudio comprensivo y de aplicación de teorías a situaciones y ejemplos prácticos. Las clases expositivas y los seminarios entrenaban frecuentemente en esas situaciones e incluso disponían en aula virtual de unas 40 preguntas tipo test para entrenarse en el tipo de examen que debían realizar (en ningún caso estas preguntas de entrenamiento coincidían con las de examen). Concluido el examen se obtuvo el índice de dificultad y el de discriminación para todas las preguntas y según la clasificación de preguntas de Bloom revisada.

Respecto al rendimiento académico se cuenta con la nota media de las notas del examen objeto de este estudio, con un valor del 0 al 10.

RESULTADOS

En primer lugar, se aplicó la clasificación según exigencia cognitiva de Bloom revisada por Anderson et al. (2001) y de las seis categorías posibles, en este examen se obtuvieron 17 (36.17%) preguntas en la categoría 1 (recordar), 19 (40.42%) en la 2 (comprender), 8 en la 3 (aplicar), 1 en la 4 (analizar) y 1 en la 5 (evaluar). Dado que hay pocas en las categorías 3, 4 y 5 y dado que según algunos autores (Prieto, 2016), éstas estarían dentro de niveles superiores de dicha taxonomía, se han agrupado en una única categoría “procesos superiores” que tiene 11 (23.4%) preguntas. Lo habitual en los exámenes PEM es incluir preguntas sólo de las categorías 1 y 2. En nuestro caso, de dichas categorías tenemos el 76.59% y sorprendentemente (dado que se diseñaron sin tener en cuenta esta clasificación ni ninguna otra) tenemos un 23.41%. De cada una de estas categorías se han obtenido los índices de dificultad y discriminación que aparecen en la tabla 1.

Tabla 1. Índices de dificultad, de discriminación y promedio de aciertos en las preguntas válidas del examen realizado según los tres tipos de preguntas según taxonomía de exigencia cognitiva de Bloom revisada

	Tipos de pregunta		
	1 (memorística) <i>n</i> = 17, %= 36.17	2 (de comprensión) <i>n</i> = 19, %= 40.42	(de procesos superiores) <i>n</i> = 11, %= 23.41
Índice de dificultad (%)	44.39	38.06	30.92
Índice de discriminación (%)	36.36	28.95	23.55
promedio de aciertos (0-1)	0.42	0.49	0.58

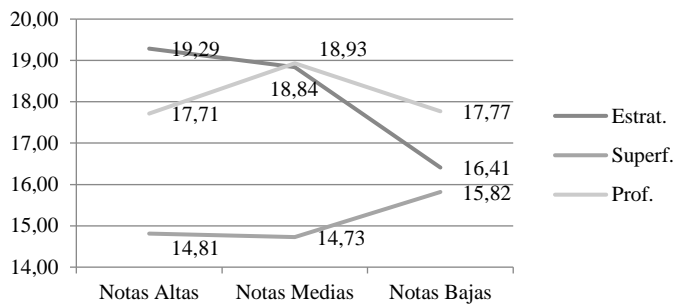
En general vemos una mayor dificultad en las preguntas memorísticas (tipo 1), siendo más fáciles las de comprensión (tipo 2) y por último las de procesos superiores (tipo 3); esto se observa en el índice de dificultad y también en el promedio de aciertos, dándose en esta variable diferencias significativas según tipos de preguntas en el ANOVA ($F(2,258)= 22.980, p< .000$), en concreto entre tipo 1– tipo 2 ($p= .017$), tipo 1– tipo 3 ($p< .000$) y tipo 2– tipo 3 ($p< .000$), según la prueba post hoc de Tukey.

Esto a priori contradice la progresión en dificultad de la clasificación de las preguntas, es decir, debería haber más dificultad en las preguntas tipo 3, luego en las tipo 2 y por último las tipo 1. Pero se podrían explicar estos resultados teniendo en cuenta los enfoques de aprendizaje y la autorregulación en estos sujetos. Así, en la prueba ASSIST de enfoques de aprendizaje, tenemos puntuaciones más altas en enfoque estratégico ($M= 18.33, DT= 4.03$) y profundo ($M= 18.34, DT= 3.73$) que en superficial ($M= 15.02, DT= 3.65$). Estas diferencias son además estadísticamente significativas en el ANOVA ($F(2, 258)= 21.970, p< .000$), en concreto entre tipo 1– tipo 2 ($p< .000$) y tipo 2– tipo 3

($p < .000$) según la prueba post hoc de Tukey. Esto nos da a entender que el estilo de estudio y aprendizaje predominante es estratégico y profundo, permitiendo así asimilar contenidos y responder a preguntas de nivel más complejo (a los cuales ya estaban relativamente entrenados durante el curso).

Sin embargo, cuando subdividimos a los sujetos según nota en la asignatura según primer cuartil de notas (notas 1, $n = 22$), segundo y tercer cuartil (notas 2, $n = 44$) y cuarto cuartil (notas 3, $n = 21$) y estudiamos los enfoques de aprendizaje que corresponden a esos sujetos, los resultados los vemos en la figura 1.

Figura 1. Promedio de puntuación en los tres enfoques de aprendizaje del Assist en los sujetos según tres rangos de rendimiento académico



En general, el conjunto de sujetos puntúa más bajo en enfoque superficial que estratégico y profundo, los cuales puntúan parecido, aunque el que puntúa más alto en estratégico es el subgrupo de notas altas. Analizando posibles diferencias en la puntuación de cada enfoque de aprendizaje según subgrupo de notas, sólo es significativa la diferencia entre los tres subgrupos de nota en puntuación en enfoque estratégico ($F(2.86) = 3.649$, $p = .030$) con un post hoc Tukey significativo entre notas altas y bajas ($p = .047$). Esto nos da a entender que lo que diferencia al grupo de notas altas del de notas bajas es sobre todo la mayor puntuación en enfoque estratégico, en el cual se estudia para aprender, pero de modo autorregulado y atendiendo a las características concretas que se piden en la asignatura.

La nota de examen correlaciona con enfoque estratégico y con autorregulación durante el estudio. Igual ocurre con las notas en las preguntas tipo 3 (procesos superiores). Sin embargo, las preguntas tipo 2 (comprensión) no correlacionan con nada.

En cuanto a la relación entre enfoques de aprendizaje y autorregulación, tenemos que con enfoque estratégico correlaciona autorregulación global; enfoque superficial correlaciona negativamente con autorregulación al terminar de estudiar, tras

la nota de examen y con autorregulación global; enfoque profundo no correlaciona con autorregulación.

DISCUSIÓN Y CONCLUSIONES

En un examen tipo PEM de asignatura universitaria de Psicología, diseñado de modo tradicional, se ha aplicado a posteriori la taxonomía de exigencia cognitiva de Bloom revisada y se ha podido constatar la presencia típica de preguntas (un 76.59%) de bajo nivel de exigencia cognitiva (memorización literal y comprensión) y también de un 23.41% de preguntas de exigencia cognitiva media o superior (aplicar, analizar, evaluar). Ni los índices de dificultad han seguido un orden creciente según la exigencia cognitiva de los tipos preguntas, ni decreciente en las notas promedio, como era de esperar. Ello se explica porque en estos sujetos se ha detectado predominancia de enfoques estratégicos y profundos y de mayor autorregulación que les ha permitido nivelar dificultad y notas.

Este estudio exploratorio abre un camino hacia una mayor racionalización en el diseño de este tipo de exámenes para conseguir una coherencia prefijada con los objetivos formativos y por tanto una mejora en la calidad del aprendizaje del alumnado universitario.

REFERENCIAS

- Abellán-Pérez, C. (2018). *Aplicación del modelo de Rasch en un cuestionario de autorregulación en el estudio* (Trabajo fin de grado). Universidad de Murcia, Murcia.
- Agencia Nacional de Evaluación de la Calidad y Acreditación (2013). *Guía de apoyo para la redacción, puesta en práctica y evaluación de los resultados del aprendizaje*. Madrid: ANECA.
- Anderson, L.W., Krathwohl, D.R., Airasian, P.W., Cruikshank, K.A., Mayer, R.E., Pintrich, P.R.,... Wittrock, M.C. (2001). *A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives*. New York: Longman.
- Biggs, J.B., y Collis, K. F. (1982). *Evaluating de quality of learning. The SOLO taxonomy* (Structure of the Observed Learning Outcome). New York: Academic Press.
- Bloom, B., Englehart, M., Furst, E., Hill, W., y Krathwohl, D. (1956). *A taxonomy of educational objectives, Handbook I: Cognitive domain*. New York: David McKay Company.
- Cárdenas, J.A., Blanco, L.J., y Cáceres, M.J. (2016). La evaluación de las matemáticas: análisis de las pruebas escritas que se realizan en la Secundaria. *Revista Iberoamericana de Educación Matemática*, 48, 59-78.
- Doval, E., y Renom, J. (2010). *Desarrollo y verificación de la calidad de pruebas tipo test*. Barcelona: Institut de Formació Continua (Universitat de Barcelona).
- Hidalgo, M., Romero, A., González, F., García-Sevilla, J., Pérez, M.A., Carrillo,... Peñaranda, M. (2009). Herramientas de evaluación de los estilos de aprendizaje en universitarios: Traducción y adaptación del cuestionario ASSIST. *Comunicación presentada a las IV Jornadas sobre el Espacio Europeo de Educación Superior: "El Espacio Europeo de Educación Superior en el horizonte del 2010"*. Murcia, 22, 23 y 24 de septiembre de 2009.

- Hubbard, R. (1997). Assessment and the Process of Learning Statistics. *Journal of Statistics Education*, 5(1), 1-9.
- Kattan, E., Pérez, G., Le Roy, C., Sirhan, M., González, A., Rybertt, T.,... Riquelme, A. (2014). Proceso de mejora de pruebas de conocimiento con preguntas de selección múltiple en un curso teórico de pregrado de medicina. *Revista de Educación en Ciencias de la Salud*, 11(2), 116-123.
- Kelley, T.L. (1939). The selection of upper and lower groups for the validation of test items. *Journal of Educational Psychology*, 30(1), 17.
- Kominski, C.A. (2012). Designing multiple choice tests to measure higher order thinking. *Test Item Writing*, 10, 1-33.
- Pérez, M.L., Carretero, M.R., Palma, M., y Rafel, E. (2000). La evaluación de la calidad del aprendizaje en la universidad. *Infancia y Aprendizaje*, 91, 5-30.
- Prieto, A. (2016). *Taxonomía de Bloom. Usos para los profesores y para los alumnos*. En <https://es.slideshare.net/alfredo.prietomartin/001-taxonoma-de-bloom>
- Scouller, K. (1998). The influence of assessment method on students' learning approaches: Multiple choice question examination versus assignment essay. *Higher Education*, 35, 453-472.
- Scully, D. (2017). Constructing multiple-choice items to measure higher-order thinking. *Practical Assessment, Research y Evaluation*, 22(4), 1-13.
- Stanger-Hall, K.F. (2012). Multiple-Choice Exams: an obstacle for higher-level thinking in introductory science classes. *CBE—Life Sciences Education*, 11, 294-306.
- Su, W.M., Osisek, P.J., Montgomery, C., y Pellar, S. (2009). Designing multiple-choice test items at higher cognitive levels. *Nurse Education*, 34(5), 223-227. doi:10.1097/NNE.0b013e3181b2b546
- Tait, H., Entwistle, N.J., y McCune, V. (1998). ASSIST. A reconceptualization of the Approaches to Studying Inventory. En C. Rust (Ed.), *Improving students as learners* (pp. 262-271). Oxford: Oxford Bookes University.
- Yonker, J.E. (2011). The relationship of deep and surface study approaches on factual and applied test-bank multiple-choice question performance. *Assessment y Evaluation in Higher Education*, 36, 673-686.

Recibido: 25 de julio de 2019

Recepción modificaciones: 2 de septiembre de 2019

Aceptado: 18 de septiembre de 2019