

Análisis psicométrico de un cuestionario Moodle propuesto como examen online en la asignatura de Química Física General

Isabel López-Tocón

Departamento de Química Física, Facultad de Ciencias, Campus de Teatinos s/n, Universidad de Málaga. E-29071 Málaga. E-mail:tocon@uma.es

Cduatcev'

C'xkt wcn's wgnkqppcct g'ku'rt qraquf. 'y kj kp'vj g'O qaf ng'ngctpkpi 'rncvqto. 'cu'cp'qprkpg'gzco 'hqt 'yj g' uwdlgev'ōI gpgtcn'Rj {ukecn'Ej go kat {ō "ql'vj g'lt u'eqwt ug"kp'vj g'f gi tgg"ql'Ej go kat {O'K'ku" c' s wgnkqppcct g'y kj "42"kgō u'qhl'o wnkrg'ej qkeg'v'rg'vj cv'j cxg"o qt g'vj cp'vj tgg'f kmt cevqt u.'y kj "c" ukpi ng'qt "o wnkrg'eqt gev't gur apugu. 'c'f wt cvkqp'qhl; 2'o kpwgu. 'c' ugs wgpvkncpckxi cvkqp'c'pf 'c'f grc'f gf" lggf dcen0'k'q'qt f gt "v'ej gen'vj g'ghgevkxgpgui"cpf "xcn'f kf "ql'vj g's wgnkqppcct g.'yj g'wcvkatecn'cpf" ru'ej qo gvt ke't gumnu. "ecrewn'v'f "qp'vj g'dcuku'ql'c'ercu'kecn'vj gqt {"ql'vguu. "ct g'cpcn'f | gf "hqt 'yj g' s wgnkqppcct g'c'u'c' 'y j qng'c'pf 'hqt'gcej 'kgo 0Vj g'r u'ej qo gvt ke'r ct co vgt u'qt 'kpf kecvt u'imej 'cu 'yj g' l'ektk'f 'kpf gz. 'yj g'f kuet ko kpcv'k'p'k'f gz "c'pf 'yj g'kpvgt p'c'ileqpu'k'w'p'ef "eqgh'ek'g'p'v'c't g'c'p'c'n'f | gf 0Vj g'ug' kpf kecvt u'y kn'o cng'r qu'k'd'ng'v'q'f gv'ev'vj g'y g'c'np'g'u'g'u'q'hl'vj ku'v'rg'q'hl'vgu'cu" c'xkt wcn'cu'gu'g'ō gpv' v'q'q'ō'

Mg'f y q'f u'c'<O qaf ng's wgnkqppcct g.'Qprkpg'cu'gu'g'ō gpv.'Ru'ej qo gvt ke'c'p'c'n'f u'ku'ō'

"

Tguwo gp''

Ug'rt qraq'p'w'ewgnkqppcct k'xkt wcn'f gpvt q'f g'rc' r'ncv'q'q' c'f g'crt g'p'f k'clg"O qaf ng.'eqo q'gzco gp" qprkpg.'c'f k'w'p'ekc.'r'ctc'rc'c'uki pcw'c'S w'p'kec'H'k'ec'I gpgtcn'f g'rt ko gt q'f g'I tcf q'gp'S w'p'kec'ō' Ug'w'c'w'f g'w'p'ewgnkqppcct k'eqp'42'kgō u'f gn'v'k'q'qr ek>p'ō Ankr ng'eqp'w'p'Ap'k'ec'q'xct ku't gur w'g'u'c' eqtt g'ew'u'eqp'w'p'c'f wt ek>p'f g'j'qt c'f "f'gf k'c.'w'p'c'xgi cek>p'ugew'p'ek'c'n'f 'w'p'c'w'q'c'rkō gp'v'ek>p' f'k'g'f'k'f c'ō'R'ctc'eqp'q'eg'it'rc'gh'ev'k'f'cf "f'hc'd'k'f'cf "f'gr'ewgnkqppcct k'ug" c'p'c'n'f'cp'rq'u't gum'w'c'f'qu' g'w'c'f'f'k'ek'qu'f' r'ukeqō 2'v'kequ.'ecrewn'f'qu'gp'd'c'ug'c'rc'v'g'qt 'f'c'ef' u'kec'f'g'v'qu'v'g'u'u.'f'gr'ewgnkqppcct k'gp" eqp'w'p'v'q'f'f'g'ec'f'c'w'p'q'f'g'v'qu'f'g'ō u's w'g'v'q'k'p'v'gi t'c'p'ō'Eqo q'r'ct'ā o gvt qu'q'k'p'f'kec'f'qt g'u'r'ukeqō 2'v'kequ' ug'c'p'c'n'f'cp'gr'v'p'f'keg'f'g'v'ek'k'f'cf.'gn'p'f'keg'f'g'f'k'uet ko k'p'c'ek>p'f'gr'v'ek'g'k'eg'p'v'g'f'g'eqp'uk'w'p'ek'k'p'v'gt'p'c." gpvt g'q'vt qu'ō'Gu'qu'k'p'f'kec'f'qt g'u'x'p'c'r'gt o k'kt'f'g'v'g'v'ct'rc'u'f'g'd'k'k'f'cf'g'u'f'g'g'w'g'v'k'q'f'g'r't'w'g'dc'eqo q" j'gtt co k'p'w'f'g'g'x'c'w'c'ek>p'xkt w'c'it'ō"

Rc'ndt'cu'v'nc'xg'<Ewgnkqppcct k'Q qaf ng.'G'x'c'w'c'ek>p'q'p'rk'p'g.'C'p'ā'rk'ku'r'ukeqō 2'v'keq'ō'

1. Introducción

La etapa final en el proceso de enseñanza-aprendizaje de cualquier asignatura es la evaluación, y posiblemente sea, una de las etapas más importantes por lo que ello implica tanto para el docente como para el estudiante. Para el docente supone medir los logros en los resultados de aprendizaje que se pretenden alcanzar, mientras que para el alumno supone una prueba donde demostrar los conocimientos adquiridos, que servirán para superar la asignatura, y pasar de curso académico.

En estos últimos años, con la implantación del Espacio Europeo de Enseñanza Superior que propone nuevos métodos de evaluar los conocimientos y capacidades de los estudiantes (ENQA, European Association for Quality Assurance in Higher Education, 2015), el procedimiento de evaluación en la mayoría de las asignaturas de cualquier Grado universitario, consiste en lo que se conoce como evaluación continua, donde el mayor peso de la calificación recae sobre un examen escrito, que se realiza al final del periodo docente lectivo, y otra contribución, de menor proporción, corresponde a varias actividades docentes que han ido realizando los estudiantes a lo largo del curso, a modo de evaluación continua de la asignatura. Sin embargo, los nuevos escenarios de docencia, bimodal y online, que han sido adoptados en el ámbito universitario como una nueva metodología didáctica debido a la actual situación socio-sanitaria, han hecho que los docentes programen nuevos métodos de evaluación a distancia, haciendo uso de los medios digitales, en sustitución al tradicional examen escrito presencial.

Hoy en día, las tecnologías de información y comunicación se encuentran prácticamente integradas en el ámbito docente, como son las diferentes plataformas de aprendizaje, entre ellas la plataforma Moodle (Rice, 2010) implementada en el Campus Virtual de la UMA. Dentro de este espacio virtual se pueden realizar diferentes actividades online, algunas de las cuales pueden ser evaluadas o corregidas por la propia plataforma, facilitando la labor del docente en aquellos grupos con elevado número de estudiantes. Son, por ejemplo, los cuestionarios o pruebas de nivel, los cuales pueden incluso utilizarse como herramienta evaluadora a modo de examen virtual, a distancia, en aquellas asignaturas que han sido impartidas en modalidades de docencia telemática. En este sentido, la última versión de la plataforma Moodle permite no solo diseñar cuestionarios con una gran variedad de tipos de preguntas, verdadero/falso, opción múltiple, numérica, emparejamiento, anidada, etc., (López-Tocón, 2016) donde se pueden insertar otros elementos en las preguntas como figuras, tablas, videos y audios, pasando a ser preguntas tipo multimedia, sino además, permite diseñar diferentes modalidades de actuación con los cuestionarios (diferido, adaptativo, interactivo con múltiples intentos, etc.) dependiendo de la estrategia didáctica que se quiera dar a la prueba en sí, bien como actividad formativa o bien como actividad evaluadora.

La evaluación virtual no es un elemento nuevo en el ámbito educativo. Diferentes autores ya han realizado estudios sobre la efectividad de la evaluación incorporando las tecnologías de la información y comunicación (Delgado y Oliver, 2006; Blanco y Ginovart, 2012; Crews y Curtis, 2011; Ferrao, 2010; López-Tocón, 2016) y haciendo uso de diferentes estrategias didácticas dentro del entorno virtual utilizado. Los resultados obtenidos han sido siempre aceptables, aunque con propuestas de mejoras en la elaboración del cuestionario o prueba virtual. Por tanto, dada la importancia del proceso de evaluación y del propio medio electrónico que se utiliza para llevar a cabo la evaluación a distancia, es necesario elaborar herramientas virtuales, o cuestionarios Moodle en este caso, que tengan calidad suficiente como para garantizar el carácter de una evaluación fiable y precisa, adaptados en todo momento, a las necesidades del proceso de enseñanza-aprendizaje de la asignatura, y en concreto a los conocimientos y competencias que se exigen para superar la materia en estudio.

Una forma de conocer la calidad de los cuestionarios Moodle es haciendo uso de las teorías estadísticas, en particular la Teoría Clásica de los Tests (Muñiz, 2010), que va a permitir estimar la fiabilidad de un cuestionario a través de diferentes parámetros estadísticos y psicométricos que se derivan del propio cuestionario, como un conjunto, y de cada uno de los ítems que lo forman. La interpretación y el análisis de estos parámetros van a permitir, también, dos acciones concretas. En primer lugar, conocer el efecto que produce los diferentes tipos de preguntas en la propia evaluación y, por tanto, en la propia calificación de la prueba. En segundo lugar, va a permitir detectar aquellas preguntas/ítems que no son idóneas para discriminar las diferentes capacidades cognitivas de los estudiantes, por lo que habría que eliminarlas del cuestionario, o bien, plantear ajustes y mejoras concretas. Todo este proceso a modo de retroalimentación

o feedback permite la optimización en el diseño del cuestionario para que la evaluación resulte más fiable, con calidad evaluadora.

2. Objetivos

Con este trabajo de investigación docente se pretende alcanzar varios objetivos, entre ellos:

- 1.- Diseñar y optimizar un cuestionario Moodle como herramienta evaluadora, a modo de examen online, en la asignatura Química Física General que se imparte en primero del Grado en Química.
- 2.- Conocer el funcionamiento de las preguntas del tipo opción múltiple con más de tres distractores, bien con una única respuesta correcta o bien con varias respuestas correctas, en la calificación del cuestionario. Este tipo de preguntas son las más utilizadas en los cuestionarios virtuales y presentan como característica el puntuar negativamente aquellas respuestas incorrectas, en el caso de múltiples respuestas correctas, viéndose afectada la calificación final del cuestionario.
- 3.- Analizar e interpretar los resultados estadísticos y psicométricos que se derivan del cuestionario Moodle y que se encuentran disponibles automáticamente en la plataforma. Comparando los resultados obtenidos con los valores de referencia, que propone la propia plataforma para el caso de cuestionarios virtuales de tipo sumativo, se podrá identificar aquellas preguntas que no son adecuadas para discriminar las diferentes habilidades de los estudiantes. También permitirá conocer el comportamiento de estos parámetros en este tipo de pruebas virtuales con unas características específicas de diseño.

"

3. Desarrollo de la innovación

3.1. Entorno docente.

La asignatura de Química Física General se imparte en el primer semestre del primer curso del Grado en Química a un grupo de 80 estudiantes, aproximadamente. Es una asignatura íntegramente teórica de 6 créditos (60 horas en el aula). Se trabaja con una metodología docente basada en un método pedagógico mixto, tipo blended-learning (Baelo, 2009), donde se combinan las clases presenciales desarrolladas de forma tradicional en el aula y las actividades electrónicas virtuales desarrolladas en el Campus Virtual como son cuestionarios virtuales, tutoriales y tareas a entregar. El programa docente consta de 11 temas que se pueden agrupar en cinco bloques temáticos que son: Materia, Disoluciones, Termodinámica, Electroquímica y Cinética Química. Los conocimientos teóricos de esta asignatura se ponen a prueba en las prácticas de laboratorio que se desarrollan dentro de la asignatura de Operaciones Básicas de Laboratorio durante el segundo semestre, donde también participan otras áreas de la química.

3.2. Cuestionario Moodle

El diseño de cualquier tipo de cuestionario no es tarea fácil (Muñiz y Fonseca, 2019), ya que intervienen diferentes elementos a considerar. Primero, la definición de la propia variable que se quiere medir, bien puede tratarse de conocimientos relacionados con definiciones, características de los procesos químicos, etc, o bien competencias y habilidad en resolver problemas, en interpretar datos y gráficas, en deducir consecuencias de un determinado experimento químico, etc., segundo, la construcción de los propios ítems que integran el cuestionario, que son de una gran variedad al hacer uso de las tecnologías de información (Parshall, 2010) como por ejemplo, tipo verdadero/falso, numérica, emparejamiento, etc, y por último, las

propias especificaciones relacionadas con el cuestionario (duración, número de ítems, formato de respuesta, tipo de navegación, etc).

Los cuestionarios desarrollados en el ámbito educativo tratan de evaluar en cierta medida, en qué grado los estudiantes conocen un campo de conocimiento, y esto engloba no solo conocimientos, sino además habilidad en resolver problemas numéricos, y en realizar razonamientos deductivos, sobretodo en asignaturas de carácter científico, por lo que la elaboración de los mismos debe ser muy meticulosa.

El cuestionario Moodle que se ha diseñado como examen online para la asignatura de Química Física General (Tabla 1) trata de recoger los aspectos esenciales y más importantes de la asignatura, tanto en conocimiento como en competencias. Se ha dispuesto, 3 ítems para los bloques de Termodinámica y Cinética, 4 ítems para el bloque de Materia, y 5 ítems para los bloques de Disoluciones y Electroquímica. En total son 20 ítems, únicamente del tipo opción múltiple, con más de tres respuestas o distractores que se ordenan aleatoriamente para cada estudiante en el momento de abrir la prueba. Este tipo de ítem es muy utilizado en los cuestionarios docentes ya que permite la evaluación de niveles cognitivos complejos (Jurado-Núñez y col., 2013) y además, disminuye la probabilidad de respuesta debida al azar frente, por ejemplo, a los ítems del tipo verdadero/falso. Además, en este caso, se propone una serie de ítems que llevan insertados gráficos y tablas numéricas para la interpretación de los mismos, combinándose el conocimiento y el razonamiento abstracto de la asignatura. Cada ítem puntúa en la calificación final con un máximo de 0.4 puntos, ya que el cuestionario, a modo de examen online, supone el 80% (8 puntos) de la calificación final de la asignatura y el 20% restante (2 puntos) corresponde a la evaluación continua. Todos los ítems, excepto tres de ellos, presentan más de una respuesta correcta (Tabla 1). Se ha dispuesto que las respuestas correctas puntúan de forma proporcional positivamente, y las respuestas incorrectas lo hacen de forma negativamente. La proporción en la puntuación depende del número de opciones o distractores que tenga la pregunta y se recoge en la Tabla 1.

Vcdix'30F'kg° q'f'gr'ewgunkpctkq'Oqqf'ng'gp'ix'cuki pcwtc'f'g'S w'p'lec'H'f'kec'I gpgtci0'

Ítem	Tema	Nº Distractores (nº opciones)	Nº Respuestas correctas	Puntuación/Respuesta (sobre la base de 1 punto)		Puntuación/Ítem en el cuestionario (puntos)
				Correcta	Incorrecta	
1	Materia	6	3	0.3333	0.0000	0.4
2	Disoluciones	6	3	0.3333	-0.3333	0.4
3	Materia.	4	2	0.5000	-0.5000	0.4
4	Materia.	4	2	0.5000	-0.5000	0.4
5	Materia.	6	3	0.3333	-0.3333	0.4
6	Disoluciones	4	1	1.0000	-0.3333	0.4
7	Disoluciones	13	7	0.1429	-0.1667	0.4
8	Disoluciones	8	4	0.2500	-0.2500	0.4
9	Disoluciones	6	3	0.3333	-0.3333	0.4
10	Electroquímica	4	1	1.0000	-0.3333	0.4
11	Termodinámica	7	3	0.3333	-0.2500	0.4
12	Termodinámica	4	4	0.2500	0.0000	0.4
13	Termodinámica	6	3	0.3333	-0.3333	0.4
14	Electroquímica	7	3	0.3333	-0.2500	0.4
15	Electroquímica	4	2	0.5000	-0.5000	0.4
16	Electroquímica	4	1	1.0000	-0.3333	0.4
17	Electroquímica	5	3	0.3333	-0.5000	0.4
18	Cinética Química	7	4	0.2500	-0.3333	0.4
19	Cinética Química	4	3	0.3333	-1.0000	0.4
20	Cinética Química	3	1	1.0000	-0.5000	0.4

De esta forma, el cuestionario que se propone no es completamente de carácter sumativo, en el sentido que no siempre suma cantidades positivas, ya que pueden aparecer valores negativos en algunas respuestas de los estudiantes, que restarán a la calificación final del cuestionario.

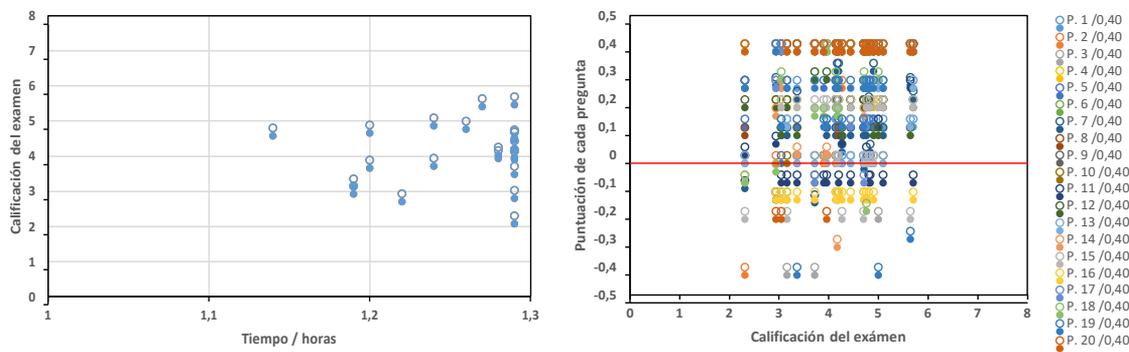
Otras características dentro del marco general de la prueba son, el tiempo, con una duración de hora y media, la disposición de los ítems de forma aleatoria para cada estudiante, la navegación secuencial, es decir, tienen que contestar a las preguntas en el orden que aparecen sin posibilidad de volver atrás, sólo hay un intento para realizar la prueba y no existe retroalimentación en las respuestas. Al acabar el cuestionario se indica únicamente la nota obtenida. Los estudiantes son informados de las características del cuestionario y de los ítems, cuando se realiza el anuncio oficial de la convocatoria del examen, dos semanas antes de realizar la prueba. No obstante, esta dinámica de trabajo en los cuestionarios es previamente conocida por los estudiantes, al realizar actividades docentes formativas online durante la evaluación continua, con otros cuestionarios de características similares.

4. Resultados y discusión

Se analizan los parámetros estadísticos y psicométricos correspondientes al cuestionario Moodle, realizado como examen online y a distancia, para la asignatura de Química Física General en la Convocatoria Ordinaria de Septiembre del año 2020. Estos parámetros son proporcionados por la propia plataforma Moodle en el apartado de Estadísticas (General Public License GNU, 2010) y son calculados en base a la teoría clásica de los tests. El número de estudiantes que realizaron el examen fue de 23, un número inferior al de la clase, al tratarse de la segunda convocatoria.

4.1. Calificaciones obtenidas en el cuestionario

Aunque esté fuera del alcance del análisis estadístico del cuestionario, se va a comprobar en primer lugar el efecto del tiempo empleado en la resolución de la prueba con las calificaciones obtenidas (Fig. 1, Izqda), ya que permitiría en cierto modo detectar conceptos utilizados en el aprendizaje, como estudiantes brillantes o efectivos, así como estudiantes que con puntuaciones bajas, no utilizan todo el tiempo disponible para resolver el cuestionario. En general, todos emplean más de una hora en contestar y a partir de esa hora cierran el cuestionario estudiantes que tienen el examen aprobado con solo 5 puntos y otros estudiantes con el examen suspenso, alrededor de 3 puntos, no emplean todo el tiempo. En general, la mayoría agota todo el tiempo de hora y media y las calificaciones están comprendidas en el rango de 2-6 puntos, no alcanzando la puntuación máxima (8 puntos) del cuestionario ninguno de los estudiantes.



Hli 03"K.sfc +Ecr!kcece»p"j gi'ewgukpctkq"O qqf rg"ht gpyg"cr!wgo r q"go r rncf q0"F tej c +Rwpwce»p"j g'ecf c"jgo lrt gi wpyc"R +"
ht gpyg"cr"ecr!kcece»p"j gi'ewgukpctkq"O qqf rg0"

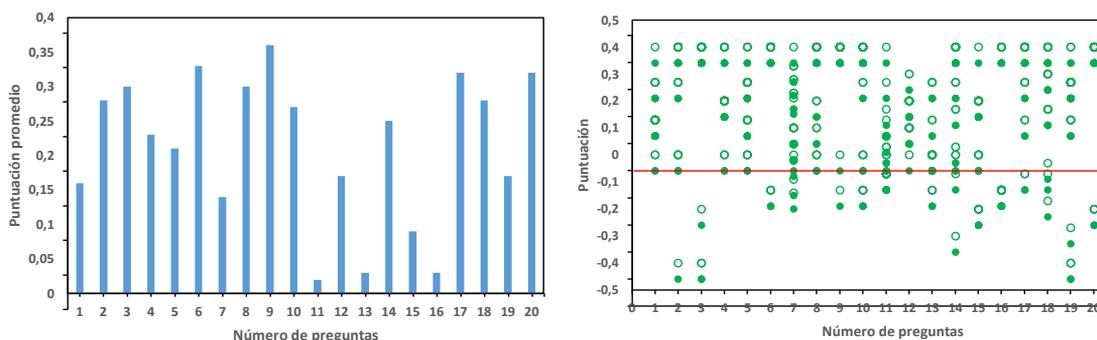


Representando las calificaciones obtenidas en la prueba frente a las puntuaciones obtenidas en cada pregunta (Fig. 1, Drcha) se comprueba que en todo el rango de calificaciones existen respuestas con puntuación negativa, incluso para los estudiantes que obtienen una mayor calificación y que por tanto, no han conseguido llegar a la calificación máxima. Además se observa que, hay preguntas con puntuaciones variadas y otras preguntas como la nº 16 y 20 que tienen una puntuación negativa o máxima, respectivamente, en todo el rango de calificación.

4.2. Análisis estadístico y psicométrico de los ítems/preguntas

El análisis psicométrico basado en principios estadísticos va a permitir determinar la idoneidad de los ítems/preguntas en base a las respuestas y su relación individual con el resto de ítems, detectando así si las preguntas propuestas son adecuadas para evaluar el nivel de conocimiento, grado de dificultad y grado de discriminación entre habilidades conceptuales altas y bajas (Heck y Van Gastel, 2006; Revuelta, Ximénez, y Olea, 2003).

Se inicia el estudio con el análisis de la puntuación promedio y la dispersión en cada ítem/pregunta, representado en la Fig. 2. Todas los ítems tienen una nota promedio positiva (Fig. 2, Izqda), pero se detectan varias preguntas, las nº 11, 13, 15 y 16, con una calificación baja, inferior a 0.1 punto. No obstante, la mayoría de ítems presentan respuestas con puntuaciones negativas (Fig. 2, Drcha) excepto las preguntas 1, 4 y 5 correspondientes al módulo de materia que tienen puntuaciones positivas, al igual que el ítem 12 que estaba diseñado de forma que cualquier respuesta puntuaba positivamente. Por tanto, se puede decir que hay aspectos de la asignatura, excepto las del bloque de materia, que no han sido bien asimilados por los estudiantes llegando a responder de forma inadecuada sobre todo en las nº 11, 13 y 16 que corresponden al bloque de termodinámica y electroquímica.



Hli 04''Ksf c +Rwpwcelop'rtqo gfk'f'g'ecf c''lgo lrtgi wpc'0''Ftej c +Fkurgtulop'f'g'ix'rwpcwcelop'f'g'ecf c''lgo lrtgi wpc'0''

También se observa, en la Fig. 2 (Drcha), que algunos estudiantes alcanzan la puntuación máxima, 0.4 puntos, en la mayoría de ítems excepto en las preguntas nº 12 y 13 del bloque de termodinámica, y que la mayor dispersión en la puntuación se detecta en los ítems 7, 11, 14 y 18, los cuales tienen un mayor número de distractores, de 13 y 7 respuestas en el ítem. Los ítems con solo 4 distractores como son, por ejemplo, los números 3, 10, 12, 15, 16 y 19 también producen cierta dispersión en la puntuación aunque mucho más polarizada entre valores positivos y negativos. Por tanto, se puede concluir que ítems con un elevado número de distractores producen una dispersión en la puntuación dentro de un rango positivo y con valores

menos negativos que en el caso de aquellos ítems donde el número de distractores es menor. En este último caso, la puntuación se concentra en un valor determinado positivo y en valor más negativo, ya que tiene una mayor contribución en la puntuación. Esto es así siempre y cuando todos los ítems estén construidos sobre una misma base de puntuación (1 punto) y que además, contribuyan con la misma proporción en la puntuación del cuestionario (0.4 puntos), como es este caso.

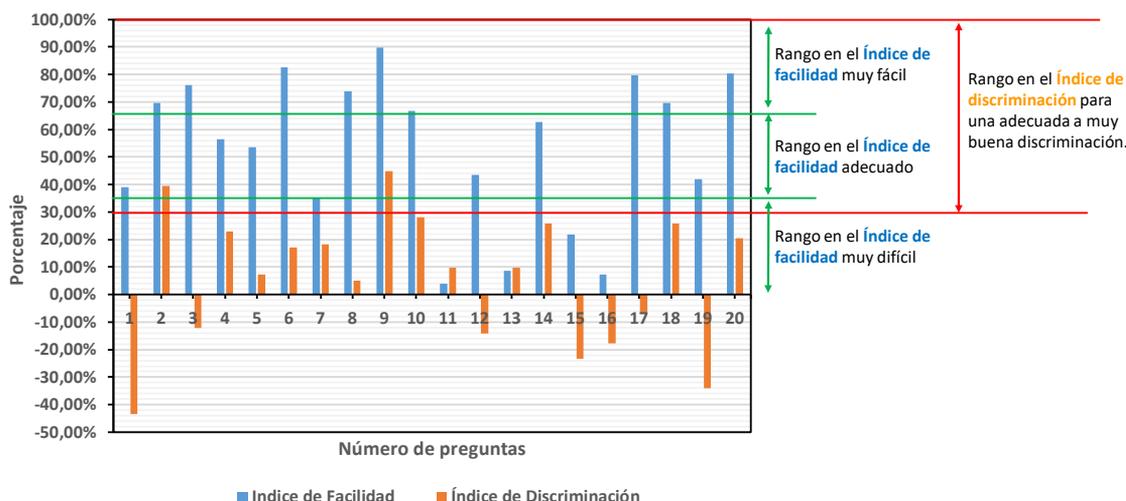
Los dos indicadores psicométricos que se analizan son, el índice de facilidad, definido como el valor medio de lo fácil o difícil que es un ítem con respecto al resto de las preguntas, y el de discriminación, definido como la correlación entre la calificación promedio de la pregunta y las del resto del examen, es decir, es un indicador del desempeño en cada ítem, por separado, de los estudiantes competentes frente a los menos competentes.

Las expresiones matemáticas de estos indicadores son: Índice de Facilidad= $X_{\text{promedio}}/X_{\text{máx}}$, donde X_{promedio} es la media obtenida en el ítem y $X_{\text{máx}}$ es la puntuación máxima obtenida en el ítem. El índice de discriminación se calcula dividiendo el grupo de estudiantes por tercios, teniendo en cuenta su puntuación con referencia al cuestionario global. Para los grupos superior e inferior se obtiene la puntuación media del ítem analizado (continuando el orden de ejecución de arriba hacia abajo); finalmente, del valor anterior se resta el promedio de la puntuación. De forma simplificada, el Índice de Discriminación= $(X_{\text{superior}}-X_{\text{inferior}})/N$ donde, X_{superior} es la suma de la fracción alcanzada (obtenida/máxima) para este ítem, para un tercio de los estudiantes con calificaciones más altas en todo el cuestionario (este es el número de respuestas correctas en este grupo) y X_{inferior} es la suma análoga para los estudiantes ubicados en el tercio inferior del cuestionario. N es la población de estudiantes que responden al ítem. Más detalle sobre estos parámetros se pueden consultar en otros trabajos (Gamage y col., 2019; Gómez-Soberón y col., 2013).

Los valores de referencia que propone la plataforma Moodle para estos indicadores en un cuestionario sumativo son: para el índice de facilidad, un valor entre 35-65% se considera correcta para el estudiante promedio. Valores superiores/inferiores indica que la dificultad de la pregunta disminuye/aumenta, respectivamente. Para el índice de discriminación se considera adecuado un valor entre el 30-50%. Valores superiores/inferiores indican una buena/mala discriminación, respectivamente. No obstante, estos valores pueden variar dependiendo de los datos obtenidos y de la finalidad de la prueba (Lord y Novik, 1968; Rodríguez, 2005). Algunos autores proponen un valor superior al 20% como un buen resultado tanto en el índice de facilidad como de discriminación (Guilford, 1975; Cano, 2004).

Los índices de facilidad y de discriminación para cada ítem se recogen en la Fig. 3 junto con los valores de referencia. Respecto al índice de facilidad, la mayoría de los ítems/preguntas se mueven en el rango adecuado, si bien hay preguntas que han resultado fáciles (nº 6, 9, 17 y 20) mientras que otras han resultado difíciles (nº 7, 11, 13, 15 y 16). Respecto al índice de discriminación, solamente dos ítems (nº 2 y 9) entran dentro del rango de valores adecuados, el resto tienen valores muy bajos, incluso negativos. En el caso de cuestionarios de tipo sumativo, estos valores negativos indicarían que los estudiantes del grupo bajo rindieron más en el ítem que los del grupo alto, y deberían de eliminarse por su escasa fiabilidad ya que reducen la precisión de la puntuación global del cuestionario.

Sin embargo, estos valores de referencia no se pueden extrapolar en nuestro caso ya que se trata de un cuestionario no sumativo, con ítems que tienen respuestas puntuadas negativamente, lo que hace que se obtengan valores más bajos a los establecidos e incluso, que no se pueda definir una diferencia clara si el índice de discriminación negativo se debe a los estudiantes mejores, más competentes, o los peores, menos competentes. Además, se tiene el inconveniente que la población de estudiantes que realizan el cuestionario (un total de 23) es relativamente bajo para el cálculo de este parámetro, donde la población es dividida en tercios.



Hli 05'Xcniqt gu'gp'gn'p'f'keq'f'g'hw'eknf'cf'f'f'g'f'kuek'lo kpece'p'r'ctc'ecf'c'f'go lrt gi wpx'f'gn'ewgunkapctkq'Oqqf'ng'0'

4.3. Análisis estadístico y psicométrico global del cuestionario

Como información general y valoración global del cuestionario, la plataforma Moodle proporciona otros parámetros estadísticos que se recogen en la Tabla 2. La calificación media del cuestionario o mediana indica el valor central de la distribución de las calificaciones y nos indica, en este caso, que la mitad de los estudiantes obtuvieron una calificación inferior al 53.39% mientras que la otra mitad alcanzó una calificación superior. Para cuestionarios discriminantes con retroalimentación diferida, se considera un examen eficiente cuando se obtiene una calificación entre 50-75%, como es en esta prueba.

Vcdrc'40Rct^a o gvt qu'guncf'f'nequ'f'rukeqo² vkequ'f'gn'ewgunkapctkq'Oqqf'ng'0'

Mediana	Desviación estándar	Asimetría (Sesgo)	Curstosis	Coefficiente de consistencia interna	Error estándar	Ratio de error
53.39%	10.83%	-0.44	-0.19	21.22%	9.61%	88.76%

La dispersión de las calificaciones alrededor del valor promedio es pequeña, con una desviación estándar del 10.83% y la asimetría de las distribución de las puntuaciones, llamado sesgo, es un valor negativo de -0.44, estando dentro de las recomendaciones para este indicador entre -1 a 1. Este valor ligeramente negativo, indica una cola de los datos hacia la izquierda, es decir, la presencia de unas pocas calificaciones con valores bajos y un mayor número de puntuaciones concentradas en valores altos de la escala. La curstosis indica qué tan plana es la distribución, que con un valor negativo de -0.19, indica una distribución prácticamente plana. El coeficiente de consistencia interna, también llamado alfa de Cronbach, es un parámetro psicométrico que estima la fiabilidad de una prueba. Aquí es deseable alcanzar valores elevados ya que un valor bajo puede indicar dos cosas: primero, que las preguntas no son muy buenas para discriminar entre estudiantes con diferentes niveles de conocimiento, por lo que las diferencias entre las

puntuaciones totales se deben al azar, y segundo, algunas de las preguntas están funcionando con una calidad diferente al resto de los ítems haciendo que el cuestionario en su conjunto no sea homogéneo. En nuestro caso, el valor es bajo del 21.22%, habría que revisar los ítems que lo integran, aunque dada las características de esta prueba no es de esperar valores muy altos y posiblemente, se necesita una mayor población de estudiantes para calcular este parámetro. También hay que tener en cuenta que los tests no miden con la misma precisión a todas las personas, ya que ésta depende en gran medida del nivel de la persona en la variable medida (Muñiz, 2010).

La incertidumbre en la calificación de cualquier estudiante viene dado por el error estándar. En definitiva, se trata si ese mismo estudiante resuelve un cuestionario equivalente en la misma situación, se esperaría una calificación dentro de ese error estándar. Cuanto menor sea el error, mejor es el cuestionario. Se considera un valor aceptable alrededor de 8-9%, en nuestro caso un 9.61%, ya que un valor superior no indicaría con exactitud sus habilidades verdaderas y posiblemente estén erróneamente calificados. Por último, la tasa de error es un parámetro vinculado al coeficiente de consistencia interna y estima el porcentaje de la desviación estándar que se debe a efectos aleatorios en lugar de diferencias de la habilidad entre los estudiantes. Un valor del 70% se considera como adecuado, aunque en nuestro caso se obtiene un valor algo mayor del 88.76%.

5. Conclusiones

El análisis de los parámetros estadísticos y psicométricos de un cuestionario Moodle permite conocer el efecto de las ítems/preguntas que lo integran en la calificación final, así como detectar qué temas han sido bien asimilados por los estudiantes y cuáles han presentado mayor dificultad para su comprensión. También ha permitido conocer el comportamiento de los clásicos indicadores psicométricos analizados en los tests, como son el índice de facilidad, el índice de discriminación, el coeficiente de consistencia interna, etc, en un tipo de cuestionario con unas características particulares donde se emplean ítems de opción múltiple, que contienen respuestas que puntúan negativamente. Se ha comprobado que estos indicadores psicométricos, que proporciona directamente la plataforma Moodle, presentan valores que se desvían de los propuestos por la propia plataforma Moodle para cuestionarios de tipo sumativo. Mientras que el índice de facilidad presenta valores aceptables dentro del rango de referencia, el índice de discriminación presenta valores bajos, fuera de rango, así como el coeficiente de consistencia interna, debido al diseño específico de este cuestionario. Se ha comprobado, también, que la mayor dispersión en la puntuación de los ítems se produce en aquellos que tienen un mayor número de distractores, aunque los valores negativos que pueden alcanzar éstos son más pequeños que en aquellos ítems, con un menor número de distractores, donde la puntuación está polarizada en un valor positivo y en un mayor valor negativo, ya que cada respuesta tiene una mayor contribución en la puntuación, al tener menos distractores.

No obstante, la filosofía de construcción de tests en el ámbito educativo tienen como objetivo determinar si los estudiantes dominan un campo de conocimiento y no tanto discriminar entre las personas, como la mayoría de los test psicológicos. Además, hay otras variables que influyen en los propios resultados de la prueba, y por tanto, en la fiabilidad del mismo como son, el número de estudiantes que realizan la prueba, el tipo de estudiante (repetidor, primera matrícula), el nivel de conocimiento de partida (si han cursado asignaturas similares en cursos anteriores, en secundaria y bachillerato), lugar de realización (fuera del aula o en el aula de informática), etc.

Por tanto, los cuestionarios Moodle como herramienta educativa formativa es muy útil para el estudio y el aprendizaje de una asignatura, ya que promueve un esfuerzo para el estudiante por el simple hecho de realizar la prueba. Su uso como herramienta evaluadora es discutible, ya que los resultados estadísticos y

psicométricos que se derivan de la prueba dependen del propio diseño del cuestionario, del número de participantes y de la diferente población. Resulta difícil establecer la fiabilidad de un cuestionario en base al índice de discriminación y al coeficiente de consistencia interna, dos parámetros que son sensibles a las características específicas de la prueba. A pesar de ello, el cuestionario que se ha diseñado para la asignatura de Química Física General tiene resultados estadísticos y psicométricos aceptables, con ciertas debilidades que deben de ser revisadas y analizadas comprobando el efecto de este mismo cuestionario en otra población diferente de estudiantes. Por tanto, en principio, no es conveniente eliminar aquellos ítems/preguntas del cuestionario que tengan valores del índice de facilidad o de discriminación alejados de los de referencia.

6. Referencias

- BAELO ÁLVAREZ, R. (2009). "El e-learning , una respuesta educativa a las demandas de las sociedades del siglo XXI". *RkzgtDk0' Tgxkac" fg" Ogf'ku" {" Gfwecekp* 35, 87-96. Disponible en <https://recyt.fecyt.es/index.php/pixel/article/view/61354>
- BLANCO, M. y GINOVART, M. (2012). "Los cuestionarios del entorno Moodle: su contribución a la evaluación virtual formativa de los alumnos de matemáticas de primer año de las titulaciones de Ingeniería" en *TWUEOWpkxgt ulokgu" cpf'Mpqy r'f'ig'Uqekgw'Lqwtpcn* vol. 9, issue 1, p.166-183.
- CANO, F. (2004). "Construcción de pruebas de conocimiento" en el *Ugo k'pctkq'k'p'v'p'ce'k'p'c'n'E'q'o'rt'q'o'ku'f'g'r'w'g'x'c'w'e'k'p'q'd'l'g'w'x'c'eq'p'g't'o'g'l't'c'o'k'p'w'q'f'g'r'w'ec'k'f'c'f'f'g'r'w'g'f'w'e'c'k'p'w'w'g't'k'q't*. Bogotá. ACOFI-Asociación Latinoamericana de Psicología.
- CREWS, T.B. y CURTIS, D.F. (2011). "Online course evaluations: Faculty perspective and strategies for improved response rates" en *Cuuguo g'p'v'('G'x'c'w'e'k'p'k'p'J'k'i'j'g't'G'f'w'e'c'k'p*, vol 36, issue 7, p.865-878.
- DELGADO GARCÍA, A.M. y OLIVER CUELLO, R. (2006). "La evaluación continua en un nuevo escenario docente" en *TWUEOWpkxgt ulokgu"cpf'Mpqy r'f'ig'Uqekgw'Lqwtpcn* vol. 3, issue 1, p.1-13.
- ENQA (2015). *T'g'r'q't'v'q'p'U'c'p'f'c't'u'c'p'f'I'w'k'f'g'r'k'p'g'u'l'q't'S'w'c'k'f'('C'u'w't'c'p'eg'k'p'j'g'G'w't'q'r'g'c'p'J'k'i'j'g't'G'f'w'e'c'k'p'c't'g'c'0'* Brussels. Belgium. ISBN: 978-9-08-168672-3.
- FERRAO, M. (2010). "E-assessment within the Bologna paradigm: evidence from Portugal" en *Cuuguo g'p'v'('G'x'c'w'e'k'p'k'p'J'k'i'j'g't'G'f'w'e'c'k'p*, vol. 35, issue 7, p. 819-830.
- GAMAGE, S.H.P.W., AYRES, J.R., BEHREND, M.B., SMITH, E.J. (2019). "Optimising Moodle quizzes for online assessment". *k'p'v'g't'p'c'w'k'p'c'n'l'q'w't'p'c'n'l'q'h'U'V'G'O'G'f'w'e'c'k'p*, vol. 6, p.1-14.
- GNU, General Public License (2013). Informe del cuestionario en https://docs.moodle.org/all/es/Significado_de_las_estad%C3%ADsticas_del_examen_de_Moodle. [Consulta: 15 de Febrero de 2021].
- GÓMEZ-SOBERÓN, J., GÓMEZ-SOBERÓN, M.C., CORRAL-HIGUERA, R., ARREDONDE-REA, S.P., ALMARAL-SÁNCHEZ, J.L. y CABRERA-CAVARRUBIAS, F.G. (2013). "Calibrating Questionnaires by Psychometric Analysis to Evaluate Knowledge" en SAGE Open. Disponible en DOI: 10.1177/2158244013499159.
- GUILFORD, J.P. (1975). *Ru'f'eq'o'g'v'k'e'o'g'j'q'f'u*. Bombay, Neva Delhi. Editorial Tata McGraw-Hill.
- HECK, A., y VAN GASTEL, L. (2006). "Mathematics on the threshold". *k'p'v'g't'p'c'w'k'p'c'n'l'q'w't'p'c'n'l'q'h'O'c'y'g'o'c'w'k'c'n'l'G'f'w'e'c'k'p'k'p'U'k'g'p'eg'c'p'f'V'g'ej'p'q'q'i'f*, 37, 925-945.
- JURADO-NÚÑEZ, A., FLORES-HERNÁNDEZ, F., DELGADO-MALDONADO, L., SOMMER-CERVANTES, H., MARTÍNEZ-GONZÁLEZ, A., SANCHÉZ-MENDIOLA, M. (2013) "Distractores en preguntas de opción múltiple para estudiantes de medicina: ¿cuál es su comportamiento en un examen sumativo de altas consecuencias?" *k'p'x'g'u'k'i'c'e'k'p'g'p'G'f'w'e'c'k'p'O'2'f'k'ec* 2, 202-210.

LÓPEZ TOCÓN, I. (2016). "Las TIC como elemento innovador en el proceso de enseñanza-aprendizaje de la asignatura de Química Física General" en *EIO KG38. "Kpxgukí cek»p"Gf wecvkx"eqp"Kó rcevq"Uqekc10Ugxlw:0* Disponible en http://amieedu.org/actascimie16/wp-content/uploads/2016/06/232_A3.pdf [Consulta: 15 de Febrero de 2021].

LORD, F. y NOVICK, M. (1968). *Ucwkakcrl'j gqtkgu'qhl'ò gpxc'n'vgu'ueqt gu* USA, Editorial Addison-Wesley.

MUÑIZ, J. (2010). "Las Teorías de los Tests: Teoría clásica y Teoría de respuesta a los ítems". *Rcr grgu'f grlRuk»rqi q*, 31, 57-66.

MUÑIZ, J. (2017). *Vgqt "Eñ ukc'f g'hqu'Vguu* Madrid: Editorial Pirámide.

MUÑIZ, J. y FONSECA-PEDRERO, E. (2019). "Diez pasos para la construcción de un test". *Rukeqj go c*, 31, 7-16.

PARSHALL, C.G., HARMES, J.C., DAVEY, T. y PASHLEY, P. (2010). "Innovative items for computerized testing", en W.J. van der Linden y C.A. Glas, *Grgo gpul'qhl'cf crvki "guakpi* (pp 215-230). Londres: Springer.

REVUELTA, J., XIMÉNEZ, M. C., y OLEA, J. (2003). "Psychometric and psychological effects of item selection and review on computerized testing". *Gf wecvkqpcrl'cpf "Ru'ej qrii kccn'O gcumt go gpx*, 63, 791-808.

RICE, W.H. (2010). *Oqqf rg'f guctt qru'f g'ewtuqu'G'rgctplpi*. Madrid: Editorial Anaya Multimedia.

RODRÍGUEZ, O., CASAS, P.P. y MEDINA, Y. (2005). "Análisis psicométrico de los exámenes de evaluación de la calidad de la educación superior (ECAES) en Colombia". *Cxcpegu'gp"O gf kek»p*, 3, 153-172.