

Análisis de validez y fiabilidad del modelo de encuesta a los estudiantes para la evaluación de la calidad de la docencia.

*Beatriz Porras Pomares. Directora del Área de Calidad.
Universidad de Cantabria*

*Pedro Gil Sopeña. Técnico de Organización y Calidad.
Universidad de Cantabria*

Palabras clave: Encuesta estudiantes, calidad docente, validez, fiabilidad, evaluación.

Resumen: *El modelo de encuesta de opinión a los estudiantes sobre la actividad docente del profesorado utilizada por la Universidad de Cantabria ha estado cuatro años vigente. Se puso en marcha como adecuación de la anterior encuesta sobre la docencia al contexto del EEES en la etapa de adaptación de los antiguos títulos mediante planes piloto. En el curso 2008-09, la encuesta pasó a formar parte de los procedimientos del Sistema de Garantía Internos de Calidad para la evaluación de la calidad de la enseñanza y del profesorado, y se ha aplicado en los nuevos títulos de Grado y de Máster. A partir del curso 2011-12, se incorporará como un instrumento de información fundamental para la evaluación de la actividad docente del profesorado (PEAD), que desarrolla en la Universidad de Cantabria el Programa DOCENTIA de ANECA.*

La importancia de estos procesos hizo necesaria una revisión en profundidad de este modelo de encuesta, que se ha abordado desde el punto de vista del diseño y contenido del cuestionario empleado. La metodología seguida para analizar la bondad del nuevo cuestionario se ha resuelto, por un lado, con un estudio de su validez, entendida como el hecho de que el instrumento mida aquello que realmente queremos medir, y por otro lado su fiabilidad, entendida como la ausencia de errores aleatorios e independencia de desviaciones producidas por errores causales.

Los resultados de las pruebas han sido altamente positivos, garantizando la validez y fiabilidad del cuestionario. Los procedimientos de análisis que se describen se han aplicado en la Universidad de Cantabria, pero son aplicables en cualquier otro contexto docente, puesto que no se apoya en propiedades de la muestra o del procedimiento de realización del test, sino en su diseño y contenido.

Antecedentes:

Las encuestas a los estudiantes sobre la calidad de la enseñanza son desde hace muchos años un instrumento de diagnóstico sobre la calidad de la enseñanza en el ámbito universitario. Sin embargo, en muy pocos casos han trascendido sus resultados más allá de un uso interno como información para el propio profesor, limitando su objetivo a una hipotética mejora de la calidad de la docencia, que queda bajo la responsabilidad del docente.

La adaptación de la Universidad al EEES ha impuesto la necesidad de un sistema de garantía de calidad enfocado especialmente a la calidad de la enseñanza, que ha incorporado las encuestas a los estudiantes como uno de los procedimientos clave (Porras B., Arízaga B. 2009,) Sin embargo esta incorporación se ha hecho sin un análisis detallado de su validez y su fiabilidad para medir la calidad de la enseñanza, y esta falta de análisis ha sido uno de los argumentos para discutir la utilidad de las encuestas para la toma de decisiones hacia la mejora de la calidad de la enseñanza y del profesorado.

De la misma manera, las encuestas se han incorporado automáticamente a los diseños de los procedimientos para la evaluación de la actividad docente, que es uno de los programas fundamentales para la calidad del sistema universitario, el programa DOCENTIA de ANECA (Agencia Nacional de Evaluación de la Calidad y Acreditación), e igualmente discutido en su utilidad para evaluar la calidad del profesorado.

Se ha abordado en muchas ocasiones la dificultad de utilizar los resultados de las encuestas en relación con la escasa participación de los estudiantes, y con las diversas formas de realización de las encuestas

(en papel, on-line, de forma presencial o no,...), o al menos se plantea como un problema para aceptar las encuestas como un método de valoración fiable. Hay también estudios detallados sobre el efecto de posibles sesgos debidos a factores implicados en la valoración del alumnado, como el publicado por Casero, A (2010), o el trabajo realizado en la Universidad Politécnica de Cataluña por García-Berro, E. García-Berro, E., Colom, X., Martínez, E., Sallares, J., Roca, S. (2011)

Sin embargo no hay muchos estudios que aborden el problema en el marco de la teoría de la validez y la fiabilidad de los test. Como referencia para el análisis que se describe en este artículo destacamos el trabajo realizado por Martínez Gómez, M. (2005) en su Tesis Doctoral "Estudio del cuestionario de evaluación del profesorado en la UPV mediante opinión de los estudiantes. Tratamiento estadístico".

En esta situación, el Área de Calidad de la Universidad de Cantabria se plantea una revisión del modelo de la encuesta sobre calidad docente, que de lugar a la elaboración de un nuevo cuestionario con los siguientes objetivos:

- Menor número de ítems.
- Redacción más escueta y clara de los enunciados.
- Adaptación a los criterios y directrices establecidos en el marco del Espacio Europeo de Educación Superior.

Se propone seguir un procedimiento de diseño que garantice la validez del cuestionario, en el sentido que se describe a continuación, y la realización de pruebas de fiabilidad, que avalen la confianza en el sistema para la evaluación de la calidad de la docencia.

Validez

La validez de un cuestionario no se puede realizar mediante una prueba estadística o matemática, sino que debe realizarse mediante un análisis del contenido de los ítems y la verificación de hipótesis sobre el su significado. En muchas ocasiones, la validez de un cuestionario se realiza mediante la valoración de expertos en la materia. En el caso de la evaluación de la calidad de la docencia la validez del cuestionario está sometida también, en un contexto menos técnico pero imprescindible, a su aceptación por los órganos con competencia en materia de evaluación del profesorado, que incluye tanto a los representantes sindicales como a los propios órganos de gobierno de la Universidad.

El antiguo cuestionario reproducía el modelo propuesto en la documentación del programa DOCENTIA de ANECA (ANECA 2007). Este argumento se utilizó en su momento como criterio de validez para su utilización en la Universidad de Cantabria, a falta de otros recursos más adecuados o del tiempo necesario para otro análisis.

Para nuestro cuestionario, la comprobación de la validez se realizó por medio de una valoración de expertos que atendieron a las siguientes consideraciones:

1. Se considera que un ítem es válido si estimula información exacta y relevante, por ese motivo se presta especial atención en la selección y redacción de los mismos.
2. Cuanto menos tenga que reflexionar el sujeto sobre el significado de un ítem, más válida será la respuesta puesto que no dará lugar a interpretaciones.
3. Algunas preguntas que son válidas para evaluar a un grupo de personas, pueden no serlo para otro grupo, por eso las preguntas deben ser válidas para el conjunto del profesorado de la Universidad y deben atender aspectos que la experiencia del estudiante durante el periodo de docencia sea capaz de percibir.

La comprobación de validez del cuestionario siguió tres etapas:

En primer lugar, el Área de Calidad realizó un estudio a partir de las evidencias recogidas durante los cuatro años de utilización del modelo de encuesta anterior, las sugerencias aportadas durante este tiempo por los distintos grupos de interés y los modelos utilizados por otras universidades nacionales e internacionales, presentando una propuesta para modificar el formulario de encuesta.

Esta propuesta fue presentada a los miembros de la Comisión de Calidad e Innovación Educativa para su estudio y aportación de nuevas sugerencias. La Comisión de Calidad e Innovación Educativa está compuesta por miembros de las comisiones de calidad de los títulos que se imparten en la universidad, representantes de distintos órganos de gobierno con responsabilidad en materia de calidad docente (ordenación académica, calidad e innovación educativa, profesorado y estudiantes, gerencia), y representantes de estudiantes y egresados.

El modelo propuesto, junto con las nuevas aportaciones, fueron analizadas y debatidas por la comisión permanente, que planteó algunas modificaciones.

El modelo revisado y acordado en la comisión permanente se elevó a la Comisión de Calidad e Innovación Educativa para su estudio y aprobación. Esta Comisión también propuso nuevas modificaciones que se tuvieron en cuenta.

Este proceso dio lugar a un modelo final, que fue presentado a los representantes sindicales en la Mesa de Negociación, como parte del PEAD, y aprobado finalmente en Consejo de Gobierno.

Fiabilidad

En el análisis de la fiabilidad hemos empleado dos métodos y tres pruebas diferentes. Los métodos son, el análisis del coeficiente de consistencia interna y el test – retest. Las tres pruebas realizadas han sido, en primer lugar y previo al lanzamiento del nuevo formulario, el análisis del coeficiente de consistencia interna de una selección de trece ítems del modelo de formulario precedente que mantienen en el nuevo modelo una redacción equivalente. En segundo lugar, el análisis del coeficiente de consistencia interna del nuevo formulario de encuesta. Y en tercer y último lugar, la repetición del test a un mismo grupo de sujetos bajo las mismas condiciones.

COEFICIENTE DE CONSISTENCIA INTERNA

Cuando se planteó el nuevo modelo de encuesta carecíamos de series de datos de respuestas del mismo, pero sí realizamos una primera prueba de fiabilidad que nos permitiese seguir avanzando en el diseño e implantación de la nueva encuesta. Uno de los objetivos que se propusieron cuando nos planteamos la revisión del anterior formulario fue que existiese una línea de comparación entre el nuevo y el antiguo modelo de encuesta que nos permitiese contrastar los datos y ampliar, en lugar de perder, la serie de resultados de los últimos cuatro años.

Prueba 1: Como primera aproximación al problema, se tuvieron en cuenta aquellos ítems del modelo de encuesta anterior que han mantenido un significado equivalente en el nuevo modelo. Los resultados obtenidos del coeficiente alfa de Cronbach ofrecían un buen criterio de fiabilidad, mostrando además unos resultados independientes del nivel de participación en las encuestas.

Tabla 1: Primer análisis del coeficiente de consistencia interna.

CÓDIGO IDENTIFICADOR DE ASIGNATURA	Nº DE ENCUESTAS REALIZADAS	COEFICIENTE α DE CRONBACH
A1	11	0.7798
A2	7	0.7971
A3	99	0.9946
A4	25	0.8642
A5	48	0.9486
A6	48	0.9751
A7	49	0.9502
A8	17	0.9902
A9	19	0.9519

Si bien los valores de α más bajos se corresponden con las asignaturas que menor participación han tenido, los valores obtenidos se encuentran en todo momento por encima del umbral de validez (superior a 0.7) aceptado generalmente en la literatura especializada (Nunnally J.C. ,1978)

Estos resultados nos permitieron seguir adelante con la propuesta, a la espera de completar el estudio de validez y fiabilidad con pruebas realizadas, esta vez, a las series de datos obtenidas con el nuevo formulario.

Prueba 2: La segunda prueba de fiabilidad ha consistido en el cálculo del coeficiente alfa de Cronbach para el nuevo cuestionario en una selección de asignaturas, elegidas al azar entre las que se han impartido en el primer cuatrimestre del curso 2011-12 en títulos de Grado. Los resultados obtenidos son aún mejores que los primeros: el nuevo modelo arroja valores de α muy por encima del umbral de validez (0,7) y en todos los casos salvo uno, por encima de 0,91.

Tabla 2: Segundo análisis del coeficiente de consistencia interna.

CÓDIGO IDENTIFICADOR DE ASIGNATURA	PROFESOR	Nº DE ENCUESTAS REALIZADAS	COEFICIENTE α DE CRONBACH
B1	P1	30	0,9700
B2	P1	70	0,9321
B3	P2	68	0,9771
B4	P1	77	0,9410
B5	P1	80	0,9658
B6	P2	20	0,9576
B7	P3	17	0,9298
B8	P1	32	0,9377
B9	P2	26	0,9186
B10	P3	48	0,9567
B11	P4	40	0,9549
B12	P5	97	0,9611
B13	P1	78	0,9562
B14	P1	47	0,8818
B15	P1	45	0,9140
B16	P1	21	0,9339
B17	P2	16	0,9273

TEST - RETEST

Los sujetos responden dos veces al mismo test, dejando entre las dos veces un intervalo de tiempo. El coeficiente de correlación entre las dos ocasiones es lo que se denomina coeficiente de fiabilidad test – retest. El intervalo de tiempo puede ser de días, semanas o meses, pero no tan grande como para que los sujetos puedan alterar la percepción que tienen de aquello que queremos medir.

Este método corresponde a un concepto más intuitivo de fiabilidad: un instrumento es fiable si en veces sucesivas aporta los mismos resultados. En este caso se ha seguido un método de contraste test-retest, con una metodología de estudio novedosa para los cuestionarios de calidad docente.

Prueba 3: Para la realización de la prueba test-retest, se ha seleccionado a un grupo de 22 sujetos de segundo curso del Grado en Economía que estuvieron matriculados de la asignatura C1 el curso anterior, a los que, a primeros de noviembre y en un intervalo de una semana, realizamos dos encuestas de opinión sobre la actividad docente del profesorado. La asignatura se impartió durante el segundo cuatrimestre del curso 2010 – 11. De este modo, los sujetos no alterarían la percepción que ya se habían formado del profesor que les dio docencia en su momento, al no mediar más aprendizaje ni contacto con dicho profesor entre una encuesta y otra.

A los sujetos se les advirtió de que estaban realizando una prueba estadística, pero no en qué consistía dicha prueba, ni que una semana después iban a tener que repetirla.

Por último, la encuesta se realizó en las mismas condiciones en las que se va a formalizar para el resto de estudiantes y docentes.

El estudio de los resultados obtenidos con la prueba del Test – Retest se centra en los análisis de los promedios, la desviación estándar, la frecuencia absoluta y por último la correlación entre ambas encuestas analizada en cada ítem.

El análisis de los promedios obtenidos en ambas encuestas (Tabla 3) nos revela que son muy similares ítem a ítem, lo que es reflejo de que los sujetos de la muestra entendieron los enunciados de la misma forma en las veces sucesivas en las que respondieron a la encuesta. El promedio final, obtenido de los promedios de cada ítem, desvela una desviación entre una encuesta y otra de tan solo 7 centésimas.

Tabla 3: Promedios por ítem.

ITEM	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	Promedio
Encuesta 1	4,36	4	4,09	4,05	4,23	4,05	4,86	4,36	4,68	4,05	4,38	4,73	4,41	4,41	4,55	4,5	4,64	4,37
Encuesta 2	4,36	3,82	4,18	4,14	4,41	4,18	4,55	4,27	4,36	4,14	4,19	4,48	4,19	4,32	4,41	4,36	4,68	4,30

Tabla 4: Promedios por sujeto.

ENCUESTA 1	ENCUESTA 2
5,4706	5,5294
5,4118	5,1765
5,1176	5,1176
4,8750	5,1176
4,8235	4,9412
4,5882	4,8824
4,5294	4,8667
4,5294	4,5882
4,5294	4,5294
4,5294	4,2353
4,4118	4,0000
4,3529	4,0000
4,2941	4,0000
4,2353	3,9412
4,1765	3,9412
4,0000	3,9412
3,9412	3,8235
3,8824	3,6875
3,7647	3,5882
3,6471	3,5882
3,5882	3,5294
3,5294	3,5294

Un análisis de los promedios obtenidos, esta vez estudiando los valores resultantes por sujeto (Tabla 4), muestra también semejanzas muy altas entre ambas encuestas, indicativo de la claridad y estabilidad de los enunciados.

Las desviaciones estándar obtenidas en cada ítem de ambas encuestas muestran que los resultados están muy relacionados entre las dos y discurren de forma pareja (Tabla 5). Los valores más próximos los encontramos entre los ítems 1, 4, 7, 13, 14, 15, cuya desviación en ambos casos no difiere más allá del 0,04. La desviación más alta comparando ambas encuestas las encontramos en los ítems 2, 8, 11, con valores de desviación superiores a 0,18.

Tabla 5: Desviación estándar.

ITEM	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
Encuesta 1	0,79	0,98	0,75	0,84	0,69	0,95	0,71	1,00	0,84	0,90	1,16	0,70	0,96	0,91	0,74	0,96	0,85
Encuesta 2	0,79	0,66	0,85	0,83	0,73	1,05	0,67	1,20	0,90	1,04	0,98	0,81	0,98	0,95	0,73	0,85	0,89

Por último, para el análisis de correlación hemos empleado el coeficiente de correlación de Pearson (Page, M.A., 1993): en cada ítem se ha calculado el coeficiente de correlación entre los valores obtenidos en la primera (X) y la segunda (Y) encuesta, y lo mismo con la correlación entre las medias.

Tabla 6: Coeficientes de correlación.

ITEM	1	2	3	4	5	6	7	8	9
r _{xy}	1,0000	0,8077	0,9406	0,8031	0,8475	0,9436	0,7631	0,9412	0,8525
ITEM	10	11	12	13	14	15	16	17	Media
r _{xy}	0,8106	0,8992	0,8232	0,8838	0,9504	0,8862	0,8744	0,8457	0.8001

De los valores obtenidos con este índice podemos concluir que existe una correlación positiva muy alta entre las dos encuestas y que hay por tanto un coeficiente de fiabilidad test – retest muy alto. De hecho todos los valores excepto el correspondiente al ítem 7 (0,7631), están por encima de 0,8 y en varios casos, incluso, por encima de 0,9. Cabe destacar el ítem 1 donde la correlación ha sido perfecta.

La prueba test-retest es totalmente satisfactoria, y permiten asegurar la bondad del cuestionario.

Conclusión

La validez del cuestionario para los objetivos de evaluación se considera establecida por la revisión de expertos, al aportarse:

- Las sugerencias recibidas de profesores y estudiantes a lo largo de cuatro años de funcionamiento del cuestionario anterior.
- El análisis de otros modelos de cuestionarios utilizados en el mismo contexto de evaluación: encuestas de calidad docente de universidades nacionales e internacionales.
- El análisis de la comisión permanente de calidad e innovación educativa y del pleno de la comisión.

En cuanto a la fiabilidad, los resultados obtenidos con el coeficiente α de Cronbach son indicadores de la homogeneidad de los ítems, es decir, de que todos miden lo mismo y de que existe una gran consistencia interna, por ello podemos suponer que si las respuestas están relacionadas es porque los ítems expresan o son indicadores del mismo rasgo.

Los resultados obtenidos en la prueba del test – retest, los entendemos como un indicador de la estabilidad y no ambigüedad de la encuesta, en la medida en que en ambas ocasiones los resultados han sido muy parecidos. Los sujetos entendieron lo mismo, de la misma manera y respondieron de forma idéntica o casi idéntica.

Referencias:

- ANECA (2007) Orientaciones para la elaboración del procedimiento de evaluación. http://www.aneca.es/var/media/215773/docentia_orientaciones_070302.pdf
- Casero, A. (2010) Factores moduladores de la percepción de la calidad docente. *RELIEVE*, v. 16, n. 2, p. 1-17. http://www.uv.es/RELIEVE/v16n2/RELIEVEv16n2_3.htm
- Cronbach, L. J. (1943). On estimates of test reliability. *The Journal of Educational Psychology*, 34, 485-494.
- Cronbach L. J. (1951). Coefficient alpha and the internal structure of test. *Psychometrika*, 16, 297-334.
- García-Berro, E., Colom, X., Martínez, E., Sallares, J., Roca, S. (2011) La encuesta al alumnado en la evaluación de la actividad docente del profesorado. *Aula Abierta* 2011, Vol. 39, núm. 3, pp. 3-14 ICE. Universidad de Oviedo http://www.uniovi.net/ICE/publicaciones/Aula_Abierta/numeros_anteriores/i17/12_AA_vol39_n3_sept_2011
- Martínez Gómez, M. (2005) Tesis Doctoral Estudio del cuestionario de evaluación del profesorado en la UPV mediante opinión de los estudiantes. Tratamiento estadístico. Universidad Politécnica de Valencia. <http://hdl.handle.net/10251/1891>
- Nunnally, J.C. (1978) *Psychometric Theory*. McGraw-Hill (New York) ISBN 00-7047-4656
- Page, M.A. (1993) *Elementos de Psicometría*. Eudema Universidad ISBN 84-7754-165-5
- Porrás Pomares, B., Arízaga Bolumburu, B. (2009) Sistema de Garantía Interno de Calidad. *Universidad de Cantabria* ISBN: 978-84-692-0331-8
- Porrás Pomares, B., Arízaga Bolumburu, B. (2009) Sistema de Garantía Interno de Calidad. Manual general de procedimientos. *Universidad de Cantabria* ISBN: 978-84-692-0312-5