Evaluación entre iguales remota y síncrona de una presentación oral

Estudio del sesgo de severidad/benevolencia

Jiménez Valverde, Gregorio

Universitat de Barcelona

Grup d'Innovació Docent d'Educació Científica, Tecnològica i per a la Sostenibilitat.

Departament d'Educació Lingüística i Literària, i Didàctica de les Ciències Experimentals i la Matemàtica. Facultat d'Educació

Passeig de la Vall d'Hebron, 171. Edifici de Llevant, 1 pl. 08035 – Barcelona (España)

gregojimenez@ub.edu

Calafell i Subirà, Genina

Universitat de Barcelona

Grup d'Innovació Docent d'Educació Científica, Tecnològica i per a la Sostenibilitat.

Departament d'Educació Lingüística i Literària, i Didàctica de les Ciències Experimentals i la Matemàtica. Facultat d'Educació

Passeig de la Vall d'Hebron, 171. Edifici de Llevant, 1 pl. 08035 – Barcelona (España)

genina.calafell@ub.edu

1. RESUMEN:

En este trabajo describimos una experiencia de evaluación entre iguales en el Máster de Formación del Profesorado de Secundaria, en la que los estudiantes tuvieron que evaluar las presentaciones orales de sus compañeros, de forma síncrona y remota. El posterior análisis estadístico de dichas evaluaciones, según el modelo de Rasch de múltiples facetas (MFRM), pudo identificar y cuantificar el sesgo de severidad o benevolencia que mostró parte del alumnado a la hora de evaluar a sus compañeros.

2. ABSTRACT:

We describe a synchronous and remote peer-assessment experience that was carried out in the Secondary Teacher Training Master, in which teacher students had to assess their peers' oral presentations. The subsequent statistical analysis of this assessment, according to the Many-Facet Rasch Measurement model, was able to identify and quantify the severity and leniency bias that some of the students showed when assessing their peers.

3. PALABRAS CLAVE: 4-6

evaluación entre iguales, docencia telemática, efecto del evaluador, modelo de Rasch de múltiples facetas, educación científica, enseñanza de las ciencias

4. KEYWORDS: 4-6

Peer-assessment, virtual teaching, rater effect, Many-Facet Rasch Measurement, scientific literacy, Science Teaching

5. DESARROLLO:

Diversos autores abogan por un mayor uso de la evaluación entre iguales en la enseñanza universitaria (Ibarra, Rodríguez y Gómez, 2012), como medio regulador del aprendizaje del estudiante. Este tipo de evaluación es especialmente importante en la formación de futuros profesores de secundaria, etapa en la que de manera más o menos explícita el currículo pide esta forma de evaluación (véase, por ejemplo, el artículo 4.3 de la orden ENS/108/2018, que regula la evaluación de la ESO en Cataluña).

Sin embargo, la implementación de la evaluación entre iguales no está exenta de problemas. Así, por ejemplo, las evaluaciones que un estudiante recibe de sus compañeros pueden estar condicionadas por varios sesgos, entre ellos, el error de generosidad o de severidad del evaluador (Myford y Wolfe, 2003). Es posible, no obstante, someter a un estudio estadístico los resultados de la evaluación entre iguales para poder identificar y cuantificar este error de severidad o de generosidad. El modelo de Rasch de múltiples facetas (Many-Facet Rasch Measurement, MFRM) permite analizar simultáneamente diferentes variables ("facetas") que pueden afectar a la calificación final de un estudiante, entre ellas el grado de severidad o benevolencia de los compañeros que le evaluaron (Eckes, 2015).

En este trabajo presentamos una experiencia de evaluación entre iguales de una producción oral en la que se han estudiado los resultados de dicha evaluación mediante MFRM. La evaluación entre iguales se llevó a cabo usando dos programas informáticos gratuitos, MOARS y Minifac, posibilitando realizarla de manera remota y síncrona, al ser telemática la docencia actual debido a la situación de emergencia sanitaria derivada de la covid19.

Descripción de la experiencia

La experiencia de evaluación remota entre iguales se ha desarrollado durante el curso 2020-21 con estudiantes de la asignatura "Didáctica de la Química" del Máster de Formación del Profesorado de Secundaria, de la Universitat de Barcelona. Participaron los 27 estudiantes matriculados en la asignatura (11 mujeres y 16 hombres).

En primer lugar, los estudiantes se agruparon en parejas con el objetivo de realizar una programación de una unidad didáctica de Química, de un tema de su elección, para cualquiera de los cursos de la ESO, siguiendo el modelo de programación competencial del Departament d'Educació (2020) de la Generalitat de Catalunya.

En segundo lugar, los estudiantes, ahora ya de manera individual, tuvieron que realizar una presentación oral de una de las actividades de su programación, que incluyera elementos de sostenibilidad o de química verde. Esta presentación fue objeto de la

evaluación entre iguales, indicando su grado de acuerdo con las siguientes afirmaciones:

1. La información que presenta sobre la actividad es adecuada y está bien organizada, de forma clara y lógica.
2. Atrae la atención del público, se muestra seguro/a, establece contacto visual (con el público/con la cámara) y mantiene el interés durante toda la exposición.
3. Habla clara y fluidamente durante toda la presentación (sin pausas ni muletillas). Su pronunciación es correcta. Su tono de voz es adecuado.
4. La exposición se acompaña de soportes visuales atractivos y de calidad, que ilustran adecuadamente la presentación.
5. Los aspectos de sostenibilidad de la actividad presentada son relevantes, significativos y coherentes dentro de la misma.

Las presentaciones se realizaron de manera síncrona, usando la plataforma Blackboard Collaborate y, después de cada presentación (que no podía durar más de 10 minutos), los estudiantes las evaluaron usando el programa MOARS (www.moars.com), que permite realizar evaluaciones entre iguales usando dispositivos móviles (Jiménez, 2021).

Justo después de acabar su exposición oral, cada estudiante obtuvo el resultado de la evaluación realizada por sus compañeros, presentada en forma de histograma, para cada uno de los 5 ítems a evaluar. Igualmente, el docente tiene acceso a todos los resultados de todos los estudiantes, accediendo a la opción de "Classroom results" de MOARS y, de hecho, puede pedir que MOARS establezca rankings de las puntuaciones globales obtenidas, para un ítem en concreto o para varios o todos los ítems a la vez. Sin embargo, si se desea identificar y cuantificar la severidad/benevolencia de cada estudiante como evaluador es necesario exportar los datos de las evaluaciones almacenados en MOARS (con la opción "Research data") e introducirlos en el programa Minifac, para realizar el análisis MFRM (www.winsteps.com/minifac.htm).

Debido a la limitación que presenta Minifac, que solo permite el análisis de 2000 datos simultáneos (para un número mayor de datos hay que utilizar la versión comercial, Facets), los datos se analizaron separadamente para cada ítem, excepto los ítems 2 y 3, que se analizaron conjuntamente, al referirse ambos a aspectos de comunicación oral. En este trabajo nos centraremos justamente en el análisis conjunto de los ítems 2 y 3 para describir el análisis MFRM realizado por Minifac.

El mapa de las medidas de las facetas analizadas (figura 1) es la tabla generada por Minifac que ofrece un resumen del análisis MFRM, una vez eliminados los valores *outliers*, de las evaluaciones entre iguales registradas:

- La primera columna ("Measr") muestra la escala en la que se han medido todas las facetas: el lógito, o logaritmo del cociente entre la probabilidad de que una persona reciba una calificación en un ítem (por ejemplo, 3) y la probabilidad de que reciba la calificación inmediatamente inferior (2). La escala de lógitos puede oscilar entre 0 (fijado en el nivel medio de las facetas) y ±∞.

- La segunda columna ("Students") distribuye a los estudiantes según la primera faceta analizada, su rendimiento: valores superiores o inferiores a 0 lógitos indican mayor o menor rendimiento de los estudiantes, para los ítems analizados. Simultáneamente, en la quinta columna ("Scale") aparece la puntuación media que correspondería a los estudiantes según la escala utilizada en la rúbrica de evaluación (en nuestro caso, escala Likert del 1 al 5, en la que 1=totalmente en desacuerdo y 5=totalmente de acuerdo). En este ejemplo, el estudiante 22 es el que mayor rendimiento obtiene conjuntamente en los ítems 2 y 3 (3,5 lógitos y 3,8 de valoración media) mientras que los estudiantes 11 y 20 son los que peores evaluaciones han recibido (0,2 lógitos y 2,7 de valoración media).

- La tercera columna ("Raters") ordena a los estudiantes según la segunda faceta analizada: su grado de severidad como evaluadores. Valores mayores de 1,0 lógitos indican severidad significativa (y ésta es mayor cuanto más alto sea este valor), mientras que valores inferiores a -1,0 lógitos indican benevolencia significativa a la hora de evaluar a sus compañeros (y mayor benevolencia cuanto menor sea este valor). Destacan 3 estudiantes por su benevolencia (lógitos<-1,0): 12, 25 y 26 y el grupo de estudiantes cuyo grado de severidad es similar al del profesor.

- La cuarta columna ordena la tercera faceta analizada, esto es, la dificultad de los ítems a evaluar: en este caso observamos que el ítem 2 y el 3 han resultado de una dificultad similar.

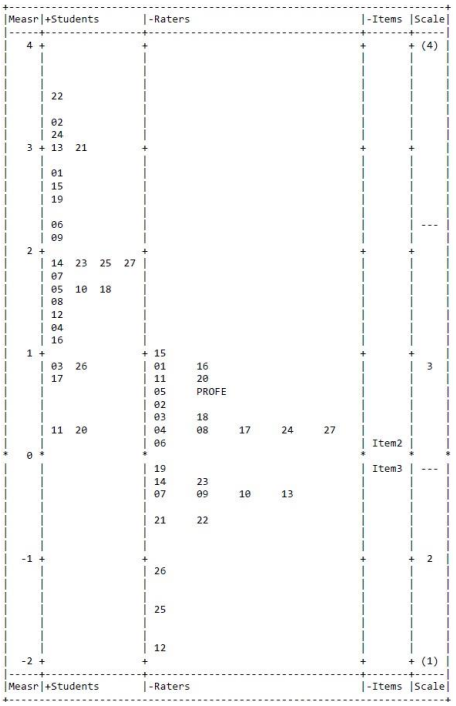
Valoración

Hemos descrito una experiencia de evaluación entre iguales, remota y síncrona, en la que el uso conjunto de MOARS y Minifac no solo nos ha permitido llevarla a cabo, sino que además nos ha ayudado a identificar y cuantificar el sesgo de severidad o benevolencia de los estudiantes.

En lo que respecta a la valoración que ha hecho el alumnado de esta experiencia, a través de una encuesta anónima que completaron al final de la actividad (N=22), el 54,5% de los estudiantes manifestó que se deberían hacer más actividades de evaluación entre iguales y solo un 13,6% del alumnado indicó que hubiese preferido no realizar esta actividad. No obstante, el mayor grado de acuerdo se consiguió cuando el 100% del alumnado valoró

como “positiva” o “muy positiva” esta experiencia de evaluación entre iguales.

5.1. FIGURA O IMAGEN 1



6. REFERENCIAS BIBLIOGRÁFICAS (según normativa APA)

Departament d'Educació (2020). *Programar per competències a l'educació secundària obligatòria*, 2ª ed. Barcelona: Gabinet Tècnic del Departament d'Educació.

Eckes, T. (2015*). Introduction to Many-Facet Rasch Measurement*, 2ª ed. Fráncfort: Peter Lang.

Ibarra, M. S., Rodríguez, G. y Gómez, M. A. (2012). La evaluación entre iguales: beneficios y estrategias para su práctica en la universidad. *Revista de Educación*, 359, 206-231.

Jiménez, G. (2021). Evaluación entre iguales representativa e inmediata con dispositivos móviles en el aula de ciencias: MOARS*. En 29 Encuentros de Didáctica de las Ciencias Experimentales* (pp. 28-35). Córdoba: Universidad de Córdoba y APICE.

Myford, C.M. y Wolfe, E. W. (2003). Detecting and Measuring Rater Effects Using Many-Facet Rasch Measurement: Part I*. Journal of Applied Measurement, 4*(4), 386-422.