

La evaluación entre pares mediada por tecnología

Delgado Celis, Zaira Yael

Facultad de Estudios Superiores Iztacala

UNAM¹

zaira.delgado@ired.unam.mx

Miranda Díaz, Germán Alejandro

Facultad de Estudios Superiores Iztacala

UNAM²

amiranda@ired.unam.mx

Meza Cano, José Manuel

Facultad de Estudios Superiores Iztacala

UNAM³

manuel.meza@ired.unam.mx

Resumen

El siguiente trabajo tuvo como objetivo identificar el nivel de concordancia que los estudiantes obtuvieron cuando evalúan un trabajo escrito a partir del índice Kappa. Se llevó a cabo en dos grupos de la asignatura Introducción a la Filosofía de Psicología SUAyED, las herramientas utilizadas fueron un taller dentro de la plataforma *Moodle* donde se presentaron las instrucciones para realizar la actividad y una rúbrica para evaluar los trabajos. A cada estudiante se le asignaron dos trabajos para evaluar.

Los resultados mostraron que no hay concordancia entre los evaluadores respecto a la evaluación de los trabajos, sólo se presentó un caso en el que si hubo concordancia alto, un caso con un nivel medio alto, medio bajo y bajo. Por lo que el diseño de evaluación entre pares requiere ser reimplementado como parte del proceso formativo de los estudiantes, así como investigar la percepción que tienen los estudiantes sobre esta actividad.

Palabras clave

Evaluación entre pares, taller virtual, concordancia entre evaluadores, Kappa.

Introducción

En la educación formal, una forma de asegurarse de que el estudiante ha aprendido los contenidos de las asignaturas es a partir de las evaluaciones, puesto que permite valorar al estudiante en cuanto a los conocimientos adquiridos de determinado periodo para finalmente emitir un juicio, dicha evaluación es realizada en la mayoría de los casos a partir de un examen en la que se consideran preguntas y conceptos descontextualizados de toda instrucción y actividades planteadas (Dochy, Segers y Sluijmans, 1999).

Con relación a lo anterior Santos Guerra (1997) menciona que la evaluación es vista como la medición del aprendizaje, por lo que su función es el control, selección, comprobación, clasificación, acreditación y jerarquización en los estudiantes.

¹ Universidad Nacional Autónoma de México

² Universidad Nacional Autónoma de México

³ Universidad Nacional Autónoma de México

Sin embargo la evaluación vista de esta manera no permite considerar otros elementos intervinientes en el proceso de aprendizaje de los estudiantes, que también tienen relevancia porque dan cuenta de éste.

En los últimos años las instituciones de educación, buscan desarrollar habilidades y competencias en sus estudiantes necesarias para su campo de formación, es así que la evaluación antes mencionada no permite dar respuesta a esta necesidad y por ende no posibilita que el estudiante reflexione sobre su propia actividad (Sluujmans, Dochy y Moerkerke, 1998).

No obstante, si la evaluación se plantea como comprensión, su óptica se transforma al considerar una dimensión crítica y reflexiva, sus funciones se modifican para enfatizar en el diagnóstico, diálogo, comprensión y retroalimentación. En este sentido Américo (2003) menciona que este planteamiento de evaluación permite también el diálogo entre los diferentes agentes de la educación, permite la reflexión sobre la evaluación y por tanto dirige a una comprensión profunda lo que permite plantear la mejora de la actividad (entendida actividad como el conjunto de procesos, actores, acciones y objetos a la cual está orientada).

De esta manera la evaluación por pares apareció como una alternativa a la evaluación sumativa, actualmente el interés que se tiene respecto a su implementación va encaminado en términos de una evaluación formativa. De acuerdo a Topping (2009) existe evidencia que demuestra que la evaluación por pares puede dar lugar a mejoras en la efectividad y la calidad del aprendizaje.

En esta misma línea Falchicov (2001) lo define como el proceso a través del cual grupos o individuos evalúan a sus pares. En este caso puede o no implicar una discusión previa sobre los criterios, lo que puede involucrar el uso de instrumentos creados por otros anteriormente o diseñados por el grupo para contemplar las necesidades específicas. Topping (2009) por su parte argumenta que la evaluación entre pares comprende un arreglo para los estudiantes en el que se considera y especifica el nivel, valor o calidad del producto o función de otro estudiante de igual estatus (puede ser flexible en el caso de los años). En este caso la evaluación por pares toma la forma de retroalimentación. Es así que la valoración que se emiten en este proceso puede implicar desde la asignación de calificaciones o puntajes hasta la emisión de juicios y comentarios para proporcionar una retroalimentación cualitativa al compañero que se evalúa.

A su vez, se han llevado diversas investigaciones sobre evaluación entre pares (Topping, 2009; Falchicov, 2001) respecto a los puntajes que los estudiantes emiten a sus compañeros observar si existen una correlación entre las puntuaciones que ellos emiten y del profesorado.

Por otra parte, las aportaciones de los autores antes citados han destacado algunas ventajas de la implementación de la evaluación entre pares, algunas de ellas refieren a:

- Al momento de realizar una valoración de los trabajos de los compañeros se contribuye a su proceso de aprendizaje.
- Se puede manifestar una motivación por parte de los estudiantes por el aprendizaje ya que, entre otras cosas, la evaluación de los estudiantes es una forma de autoevaluación.
- La evaluación entre pares forma parte de su proceso de aprendizaje.

- Mejora la percepción acerca de la calidad de sus trabajos.
- Aumenta la responsabilidad y satisfacción en el proceso de aprendizaje.
- Que los estudiantes compartan la responsabilidad de la evaluación.

Por lo que estas ventajas contribuyen a la característica de la evaluación formativa como reflexiva, crítica y comprensiva, donde se promueven la colaboración y la metacognición.

Sin embargo también se reconocen algunas desventajas de este proceso, pues los estudiantes pueden:

- Tener pocas o nulas habilidades para discriminar los rubros que debe evaluar.
- Carecer de motivación para realizar la actividad.
- Tener actitudes negativas ante una evaluación de estas características.
- Considerar la evaluación como una carga adicional.

La evaluación entre pares mediada por la tecnología

En el caso de la educación en línea, donde el proceso de aprendizaje se encuentra mediado por la tecnología (Rosenberg, 2001), la tarea del docente no solo implica dar tutoría, evaluar tareas, resolver dudas de manera individual, sino además, ayudar al estudiante a apropiarse de habilidades para afrontar la demanda que se le genera en su campo de formación.

Tsai y Liang (2009) mencionan que los entornos de aprendizaje en línea permiten tener una nueva administración y de la evaluación entre pares, pues han mostrado beneficios para estudiantes y docentes, por ejemplo en un estudio de Tseng y Tsai (2007) al realizar una evaluación entre pares en un entorno en línea encontraron que los estudiantes ofrecían una retroalimentación caracterizada por ser correctiva o didáctica.

Al respecto, Liang y Tsai (2010) abordan una investigación en estudiantes de nivel superior para realizar escritos de biología mediado por tecnología, sus escritos fueron evaluados entre pares y por un experto en tres rondas, los resultados mostraron que los puntajes otorgados por los compañeros y por el experto muestran una consistencia positiva. Por otra parte el análisis de contenido mostró que si hay mejoras significativas graduales y mejor cobertura y organización, como resultado de una actividad de evaluación entre pares.

Tsai y Liang (2009) se implementó un módulo de aprendizaje evaluación por pares en línea para ayudar a 36 estudiantes universitarios en el desarrollo de actividades para enseñar ciencia en nivel preescolar. Se pidió a cada estudiante presentar un proyecto de actividades en ciencias y posteriormente realizar tres rondas de evaluación por pares. Se encontró que los compañeros de los estudiantes mostraron una puntuación válida que fue coherente con las puntuaciones de un experto.

Es así que los estudios que se han implementado de tipo empírico buscan la validez de la implementación entre pares, puesto que implican la comparación de los puntajes otorgados por los estudiantes y los emitidos por los docentes (expertos), por otra parte se refleja una mejora en cuanto a la elaboración de los escritos que realizan después de realizar comentarios a los trabajos de sus compañeros (Topping, 2009).

Sin embargo una línea que vale la pena considerar es la comparación de las puntuaciones que emiten los propios estudiantes a un mismo trabajo.

Desarrollo

El objetivo del presente trabajo es identificar el nivel de concordancia que los estudiantes tienen cuando evalúan un trabajo escrito.

Población

Dos grupos de primer semestre de la asignatura Introducción a la filosofía de la psicología de la carrera de psicología en línea SUAyED, conformado por 40 estudiantes de los cuales se encontraron activos 15 participantes para la realización del trabajo escrito, sin embargo solo se contemplaron 13 casos válidos puesto que dos de ellos no obtuvieron el total de evaluaciones requeridas.

Herramientas

Para llevar a cabo la actividad se implementó en la plataforma Moodle la actividad Taller, ésta permite implementar la evaluación entre pares, posibilita implementar el uso de rúbricas.

También se utilizó una rúbrica (anexo 1) que facilita realizar la calificación del desempeño del estudiante que se conformó de seis criterios a contemplar al momento de evaluar el trabajo (introducción, desarrollo, organización, citas y paráfrasis, conclusiones y bibliografía) y cada uno se presentó en cuatro niveles (deficiente, malo, bien y excelente).

Procedimiento

Se les presentó a los estudiantes como actividad final de la unidad 4 correspondiente al tema “noción de la comprensión teleológica de la psicología” el desarrollo de un trabajo escrito de manera colaborativa llamado ensayo, puesto que la elaboración de escritos tiene dos ventajas: a partir de éstos se transforma el conocimiento y no se reproduce, por tanto la finalidad de escribir es para aprender. Por otra parte posibilita la evaluación entre pares donde se revisan y se proporcionan una retroalimentación o revisión del texto para una posterior redacción final.

Cada equipo estuvo conformado por cinco integrantes, el número total de equipos fue de ocho. Sin embargo tres equipos se dieron de baja, quedando un total de cinco. Dentro de cada equipo no todos participaron, por lo que los equipos quedaron conformados de la siguiente manera:

Tabla 1. Número de integrantes por equipo

Integrante	Equipo A	Equipo B	Equipo C	Equipo D	Equipo E	Equipo F	Equipo G	Equipo H
1	A	-	E	G	J	-	-	M
2	B	-	F	H	K	-	-	N
3	C	-	-	I	L	-	-	Ñ

4	D	-	-	-	-	-	-	
5	-	-	-	-	-	-	-	-

Una vez que los equipos se enteraron de la tarea tuvieron dos semanas para realizarla, por lo que se organizaron para cumplir con cada uno de los apartados que conforman al ensayo, al finalizar la entrega del trabajo escrito se realizó en la herramienta Taller.

Esta herramienta se compone de cuatro apartados:

La primera fase se llama configuración; permite presentar todas las instrucciones a partir de las cuales se especifica qué es lo que debe realizar el estudiante, así como configurar la rúbrica para la evaluación.

La fase 2 corresponde al envío; primero se les presenta a los estudiantes las instrucciones para que suban la actividad a esta herramienta, posteriormente, permite asignar a cada estudiante los trabajos que deberá evaluar.

La fase 3 llamada fase de evaluación le muestra a cada estudiante los trabajos a evaluar, la rúbrica con la que emitirá un valor a cada rubro que se le presenta y le solicita emitir una retroalimentación.

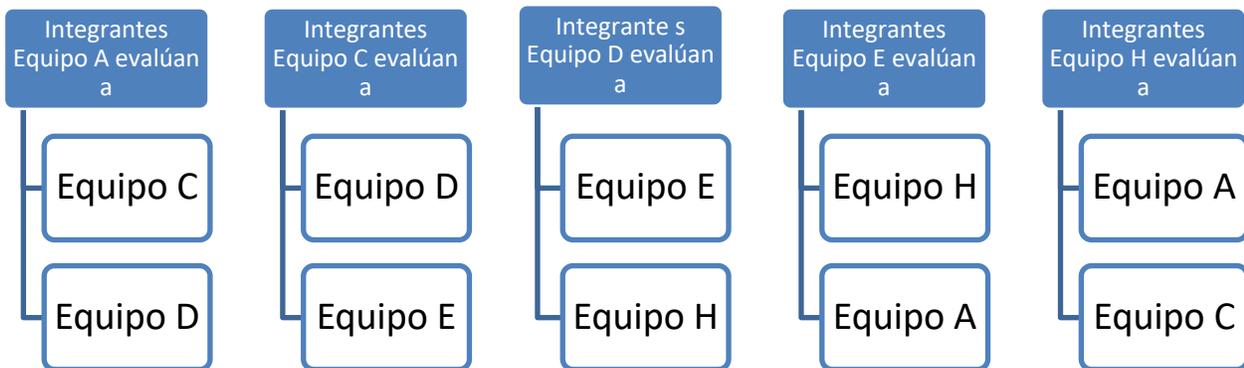
La fase 4 correspondiente a la evaluación, le permite al docente calcular las calificaciones de acuerdo a las puntuaciones otorgadas por los estudiantes.

Al término del período de desarrollo del trabajo escrito, cada integrante de equipo subió el trabajo al taller de manera individual, debido a que la condición para que se le asignara un trabajo a evaluar es que realizara su entrega.

Se les solicitó que de manera individual cada uno de los integrantes de cada equipo subieran el ensayo a la herramienta Taller de manera individual, con la finalidad de tener un control respecto a quien asignarle trabajos a evaluar.

A cada estudiante que envió el ensayo se le asignaron dos trabajos de compañeros de otros equipos, el siguiente esquema refleja la asignación de los trabajos:

Imagen 1. Distribución de los trabajos a evaluar por integrantes de equipo



Como se puede apreciar, cada integrante de equipo evaluó dos ensayos diferentes, además el proceso de asignación no permite que haya reciprocidad, es decir, que Equipo A evalúe a Equipo C y viceversa, de esta manera se reduce la posibilidad de que se beneficien en la asignación de calificación.

Una vez que los estudiantes tuvieron sus asignaciones, realizaron la evaluación a partir de la rúbrica que se les proporcionó. La dinámica consiste en que el estudiante ingresa a la herramienta Taller y puede ver los archivos de sus compañeros, selecciona uno e inmediatamente accede al texto y a la rúbrica donde elige los rubros según se encuentren o no en el trabajo. Al finalizar la rúbrica accede a un espacio donde emite una retroalimentación al ensayo de su compañero, y guarda los datos para que posteriormente su compañero pueda visualizarlos.

Al término del período de evaluación, el taller se cambió a la fase tres, donde se calculó la calificación final del trabajo, contemplando las dos evaluaciones que obtuvo cada estudiante. Finalmente se cierra el taller para que los estudiantes tengan acceso a las evaluaciones.

Resultados

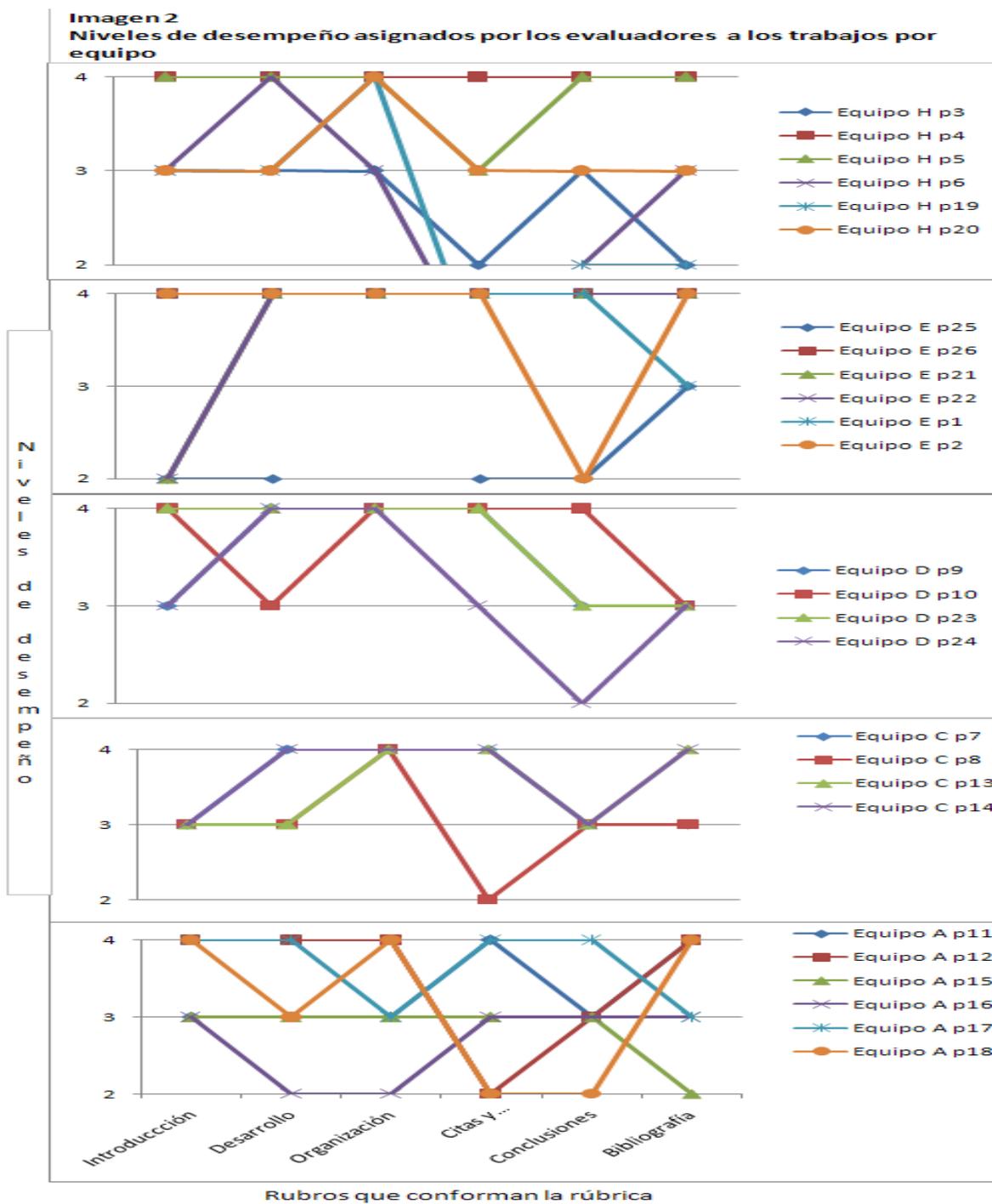
De los 15 estudiantes que trabajaron en el desarrollo del trabajo escrito, solo se contemplaron 13 casos, puesto que dos de ellos no obtuvieron las dos evaluaciones. En total se registraron 26 evaluaciones.

En la imagen 2 se presentan los puntajes registrados por los estudiantes en cada rubro de la rúbrica respecto al trabajo realizado por el equipo.

Como se puede observar no todos los equipos contaron con el mismo número de evaluaciones debido a la forma en cómo se llevó a cabo la distribución de los trabajos. Por su parte el eje de la Y representa el nivel de desempeño que se le asignó según el rubro, donde los valores van de de 1 a 4 observados en el trabajo escrito.

El eje X representa los rubros que fueron evaluados: introducción, desarrollo, organización, citas y paráfrasis, conclusiones y bibliografía.

Los colores de las líneas representan a cada evaluador del trabajo asignado. Como se puede apreciar, hay una varianza en los puntajes emitidos por los evaluadores respecto al mismo trabajo, aunque existen diversos casos en los que los estudiantes coinciden en un nivel de desempeño de un rubro respecto a un trabajo.



Para identificar el nivel de concordancia que tuvieron los dos evaluadores respecto a un mismo trabajo se utilizó el índice Kappa, que mide los grados de acuerdo entre dos jueces respecto a un mismo caso.

Los datos obtenidos en la imagen 2 fueron tratados en el programa SPSS para obtener el índice Kappa. En la tabla 3 se muestran los resultados:

Tabla 2 Índice Kappa por pareja de jueces

Pareja evaluadora	Juez 1	Juez2	Nivel de Índice Kappa
1	1	2	-0,91
2	3	4	0
3	5	6	-0,71
4	7	8	0,42
5	9	10	0
6	11	12	0
7	13	14	0,66
8	15	16	-0,28
9	17	18	-0,36
10	19	20	-0,28
11	21	22	1
12	23	24	0,18
13	25	26	0

Como se puede observar, solo existen tres índices positivos de los cuales solo uno (pareja 11) obtuvo el índice de 1 que implica un nivel de concordancia alto, es decir los dos evaluadores concordaron en que el trabajo cumplió con todos los elementos solicitados, aunque no de manera excelente, pues en el rubro de introducción los dos evaluadores consideraron como “malo” dicho apartado.

Se presentó otro caso (pareja 7) considerado como medio alto por obtener un índice de 0.66, pues concordaron en cinco de los seis rubros evaluados, fue en el rubro de desarrollo donde un juez lo consideró como “bien” y el segundo como “malo”.

El tercer índice obtenido considerado como un nivel medio bajo es el caso de la pareja evaluadora 4 con 0.42 de concordancia, donde los evaluadores concordaron en cuatro rubros, fueron los rubros de desarrollo y citas y paráfrasis que no tienen la misma percepción.

Finalmente se presenta un caso (pareja 12) con un nivel bajo de concordancia entre los jueces. Es decir solo coincidieron en tres rubros, los cuales fueron desarrollo, citas y conclusiones.

Por otra parte se obtuvieron índices negativos (parejas evaluadoras 1, 3, 8, 9 y 10). Por otra parte, se encontraron dos casos (parejas evaluadoras 2 y 13) como datos que no pueden ser calculados debido a que uno de los evaluadores emitió calores constantes, en dichos casos uno de los evaluadores puntuó como excelente todos los rubros.

Conclusiones y aportes al trabajo

Como se puede observar en los resultados obtenidos de la evaluación entre pares, se presentaron dos casos desde un inicio en los que uno de los evaluadores no realizó la revisión de los trabajos que le fueron asignados, lo que implica que en la actividad la mayoría de los estudiantes se mantuvieron dispuestos a realizar la tarea propuesta, aunque esto no necesariamente implicó que los estudiantes llevaran a cabo la actividad de manera crítica y reflexiva. Esto puede verse reflejado en los variados puntajes emitidos a los trabajos de sus compañeros, en específico aquellos casos que emitieron puntajes de 4 en todos los rubros, sin embargo solo dos casos pueden considerarse como válidos respecto al índice Kappa.

Asimismo, esta evaluación fue la cuarta dentro de la planeación del semestre, se esperaba que los estudiantes desarrollaran alguna habilidad relacionada a la evaluación sin embargo, los casos de concordancia válidos solo fueron dos.

Este trabajo permite identificar que la evaluación entre pares como proceso formativo permite que los estudiantes se vuelvan miembros activos y posiblemente valorar las retroalimentaciones que emitieron a los trabajos que evaluaron tuvo mayor pertinencia y concordancia, lo cual es conveniente contemplar. Asimismo este proceso es un ejercicio que requiere que sea llevado a cabo a lo largo de su formación, pues ellos se encuentran en primer semestre y hay diversos factores que pudieron influir en sus evaluaciones.

Por otra parte otra vía de indagación sería la percepción que tienen los estudiantes respecto a esta actividad, como se menciona, son estudiantes de primer semestre y generalmente son personas que después de varios años se reincorporan a un contexto educativo, por lo que posiblemente los estudiantes no se sienten cómodos al realizar esta actividad y solo la llevan a cabo porque forma parte de su evaluación en la asignatura, esta hipótesis es un elemento que se considera pertinente indagar en otro estudio a partir de indagar su opinión respecto a los ejercicios de evaluación entre pares para adaptar la actividad para mantener un balance entre el objetivo de la actividad y sus intereses y comodidad en el aula.

Finalmente, la herramienta utilizada llamada Taller, si bien no coloca de manera anónima los nombres de los estudiantes al momento de realizar su envío, la dinámica que se implementó para la asignación de los trabajos evitó que se presentara la reciprocidad en la asignación de los puntajes. Por otra parte se resalta la relevancia de esta herramienta para poder implementar la evaluación entre pares en la educación en línea como proceso formativo ya que posibilita mostrar paso a paso y de manera organizada cómo llevar a cabo el procedimiento que lo conforma.

Es así que dentro de las ventajas de la evaluación por pares se encuentra que los estudiantes comparten la responsabilidad de la evaluación, desde valorar cada elemento del ensayo hasta emitir una retroalimentación al trabajo de sus compañeros.

Referencias

- Américo, S. (octubre, 2003). Herramienta NTIC para la coevaluación, evaluación entre pares o evaluación paritaria en cátedras. IX Congreso Argentino de Ciencias de la Computación. Argentina.
- Dochy, F. Segers, M. y Sluujmans, D. (1999) The use of self-, peer and co- assessment in higher education: a review. *Studies in higher education*. 24 (3) pp. 331-350 Recuperado de: <http://dx.doi.org/10.1080/03075079912331379935>
- Falchicov, N. (2001). *Learning Together: Peer Tutoring in Higher Education* (1ª ed.). Routledge. Recuperado de: https://books.google.com.mx/books?hl=es&lr=&id=6liO0ynJM1oC&oi=fnd&pg=PP1&dq=Learning+Together:+Peer+Tutoring+in+Higher+Education+&ots=CQL3rysom9&sig=8SS9fYfWGI_nl_woLs26vsJhMpl#v=onepage&q=Learning%20Together%3A%20Peer%20Tutoring%20in%20Higher%20Education&f=false
- Liang Jyh-Chong y Tsai Chin-Chung (2010) Learning through science writing via online peer assessment in a college biology course. *ELSEVIER*. 13. pp. 242-247. Recuperado de: <http://www.sciencedirect.com/science/article/pii/S1096751610000357>
- Santos Guerra, M. Á. (1996). *Evaluación educativa*. Magisterio del Río de la Plata: Argentina
- Sluujmans, D. Dochy, F. y Moerkerke, G. (1998) Creating learning environment by using self-,peer- and co-assessment. *Learning environment research*. 1 Pp. 293-319. Recuperado de <https://link.springer.com/article/10.1023/A:1009932704458>
- Topping, K. J. (2009). Peer assessment. *Theory into practice*, 48(1), 20-27. Recuperado de <http://www.tandfonline.com/doi/abs/10.1080/00405840802577569>
- Rosenberg, M. (2001). *E-learning: Estrategias para transmitir conocimiento en la era digital*. Bogotá: McGraw-Hill Intramericana.
- Tsai C-C, y Liang J-C (2009) The development of science activities via on-line peer assessment: the role of scientific epistemological views. *Springer*. 37 pp. 293–310. Recuperado de <https://link.springer.com/article/10.1007/s11251-007-9047-0>
- Tseng, S. C., y Tsai, C. C. (2007). On-line peer assessment and the role of the peer feedback: A study of high school computer course. *Computers & Education*, 49(4), 1161-1174. Recuperado de <http://www.sciencedirect.com/science/article/pii/S0360131506000297>

Anexo 1 Rúbrica de evaluación

	Excelente (4)	Bueno (3)	Regular (2)	Deficiente (1)
Portada	Tiene la información institucional así como los nombres de los participantes del ensayo.	Contiene los datos de la institución y de los participantes del ensayo pero están incompletos	Solo contiene los nombres de los participantes O Solo contiene la información institucional	No hay portada
Introducción	Contiene un objetivo, explica con claridad y de manera sintética de qué trata el ensayo.	Incluye exposición general del tema. Los objetivos están un poco confusos.	No se presenta la exposición general del tema o las subdivisiones principales. El objetivo requiere clarificación o no se presentan de forma objetiva.	La introducción está incompleta, es confusa o está ausente. No incluye exposición general del tema, no son relevantes. El objetivo no está claro.
Desarrollo	Retoma todos los temas del índice, presentando ideas personales generando una conexión lógica entre las ideas. Uso de citas y paráfrasis estilo APA	Falta un tema de los revisados, presenta ideas personales pero no son claras. Hay algunas conexiones entre ideas. Hace citas y paráfrasis pero no estilo APA	Faltan varios temas, presentan varias ideas retomadas de las lecturas que se repiten no presenta ideas personales y hay poca conexión entre ideas. El documento presenta muy poca originalidad. No hace citas ni paráfrasis	Las ideas que se presentan tienen poca o ninguna relación con el tema, están pobremente definidas, no son claras. Muchas ideas se repiten. Utilizó el Copiar y pegar en una sola ocasión. No hace citas ni paráfrasis
Conclusiones	Hay proposiciones al final de un argumento	Sólo incluye opiniones personales	Sólo incluye un resumen del resto del ensayo	Es demasiado corta (menor a 3 líneas)
Bibliografía	Citada estilo APA (ver documento del bloque general nuestra aula)	Incluye bibliografía pero faltan uno o dos elementos de estilo APA	Se presentan la bibliografía no tiene estilo APA	Falta bibliografía