



Psicología Educativa

<https://journals.copmadrid.org/psed>



Propiedades Psicométricas de la Versión Española del Cuestionario de Creencias sobre la Evaluación por Pares

Nicolás Ruiz-Robledillo, Violeta Clement-Carbonell, Rosario Ferrer-Cascales, Borja Costa-López, Cristian Alcocer-Bruno, Juan Vela-Bermejo y Natalia Albaladejo-Blázquez

Universidad de Alicante, España

INFORMACIÓN DEL ARTÍCULO

Historia del artículo:
Recibido 8 March 2024
Aceptado 11 February 2025
Online el 10 de marzo de 2025

Palabras clave:
Evaluación por pares
Alumnos universitarios
Propiedades psicométricas
Validación
Cuestionario

Keywords:
Peer-feedback assessment
University students
Psychometric properties
Validation
Questionnaire

RESUMEN

La evaluación por pares, en la que el alumnado evalúa el trabajo de sus compañeros, potencia las habilidades personales y profesionales en educación. Sin embargo, faltan herramientas para analizar las creencias en esta práctica en España. El *Beliefs about Peer Feedback Questionnaire* (BPFQ) evalúa las creencias de los estudiantes universitarios con buenas propiedades psicométricas. Este estudio adapta el cuestionario al español y examina sus propiedades psicométricas, además de las diferencias entre áreas de conocimiento. En el estudio instrumental han participado 607 alumnos de la Universidad de Alicante. Los resultados confirman la estructura de cuatro dimensiones, además de un modelo bifactorial, que incluye tanto el factor general como las subescalas del cuestionario, y la consistencia interna del cuestionario en español. Estudiantes de Educación y Ciencias mostraron mayor aceptación de la evaluación por pares que los de Ciencias Sociales y Jurídicas. El BPFQ es válido y fiable para evaluar creencias estudiantiles sobre la revisión por pares, siendo útil para analizar la introducción de nuevos sistemas de evaluación de compañeros en la universidad.

Psychometric properties of the Spanish version of the Beliefs about the Peer Feedback Questionnaire

ABSTRACT

Peer evaluation, in which students assess their peers' work, promotes personal and professional skills in education. However, there is a lack of tools to analyze beliefs about this practice in Spain. The Beliefs about Peer Feedback Questionnaire (BPFQ) assesses attitudes and beliefs in university students with good psychometric properties. This study adapts the questionnaire to Spanish and examines its psychometric properties, as well as the differences between knowledge fields. A total of 607 students from the University of Alicante participated in this instrumental study. The results substantiated the four-dimensional structure and a bifactor model, consisting of both the general factor and the four subscales, and the internal consistency of the questionnaire in Spanish. Students from Education and Science showed greater acceptance of peer evaluation than students from Social and Legal Sciences. The BPFQ is a valid and reliable questionnaire for assessing student beliefs about peer review, being useful for evaluating the introduction of new peer review systems at the university.

La evaluación por pares es un enfoque educativo que se caracteriza por la retroalimentación, la comunicación, la reflexión y la colaboración entre compañeros (Boud, 1990; Boud et al., 1999; Darvishi et al., 2022). Estudios recientes señalan que la evaluación entre compañeros es un modelo de evaluación sustancial dentro del proceso de enseñanza-aprendizaje basado en la autorregulación del aprendizaje (Erkan et al., 2021; Filius et al., 2018). Las estrategias de aprendizaje entre iguales están en sintonía con la teoría moderna del

aprendizaje cognitivo (Barkley et al., 2014), puesto que involucra a los alumnos en el proceso de evaluación (Schellekens et al., 2021), fomenta un enfoque de aprendizaje profundo (Pinheiro et al., 2023) y potencia algunas de las ideas clave de la motivación intrínseca, como la "gamificación" (del inglés *gaming*, juego) del aprendizaje, el reconocimiento y la cooperación, además de promover las habilidades metacognitivas y la autorreflexión (Hunzer, 2014; Nicol et al., 2014; Sadler, 1989). La evaluación por pares se puede utilizar en entornos

Para citar este artículo: Ruiz-Robledillo, N., Clement-Carbonell, V., Ferrer-Cascales, R., Costa-López, B., Alcocer-Bruno, C., Vela-Bermejo, J. y Albaladejo-Blázquez, N. (2025). Propiedades psicométricas de la versión Española del cuestionario de creencias sobre la evaluación por pares. *Psicología Educativa*. Avance online. <https://doi.org/10.5093/psed2025a12>

Financiación. Esta investigación ha sido financiada por el Programa de REDES-13CE de Investigación en Docencia Universitaria (Programa REDES) del Vicerrectorado de Transformación digital y el Instituto de Ciencias de la Educación de la Universidad de Alicante (2021-2022). Ref.: (5505). Correspondencia: violeta.clement@ua.es (V. Clement-Carbonell).

ISSN: 1133-0740/© 2025 Colegio Oficial de la Psicología de Madrid. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

académicos y profesionales como una estrategia para mejorar la participación del estudiantado en su propio aprendizaje (Ashenafi, 2017). De esta manera, en una actividad de evaluación por pares, el estudiantado asume la responsabilidad de evaluar el trabajo de sus compañeros, generando y recibiendo retroalimentación sobre un tema específico (Boud et al., 1999; Carlsson Hauff y Nilsson, 2022; Su y Huang, 2022). En este sentido, les permite considerar los resultados de aprendizaje de sus compañeros y reflexionar sobre su propio aprendizaje (Hwang et al., 2023). Este enfoque sitúa a los educadores como diseñadores de un entorno de retroalimentación y a los alumnos como agentes clave en este proceso (Carless, 2022).

Se ha demostrado que la evaluación por pares potencia el desarrollo de la capacidad de juicio, la capacidad crítica y la autoconciencia en estudiantes, así como su conocimiento de los criterios de evaluación (Fang et al., 2021; van Blankenstein et al., 2024). En el contexto de la educación superior, la inclusión de la evaluación por pares ha supuesto una transformación del proceso de enseñanza-aprendizaje, al fomentar el cambio desde un enfoque individualista y dirigido por el equipo docente a un enfoque de evaluación más colaborativo y centrado en el estudiantado, en consonancia con los principios del constructivismo social (Vygotsky, 2012; Kasch et al., 2021). En este contexto de interacción y colaboración social, el estudiantado puede ampliar sus conocimientos, determinar sus fortalezas y debilidades y desarrollar habilidades personales y profesionales (Deneen y Hoo, 2021; Topping, 1998), evaluando la competencia profesional de sus compañeros. De hecho, numerosos estudios describen los diversos beneficios del aprendizaje que derivan de la participación de los estudiantes en la evaluación de sus compañeros (Alemdag y Yildirim, 2022; Ibarra-Sáiz et al., 2020). La revisión por pares sitúa al estudiantado como agente proactivo y reflexivo con su propio aprendizaje, lo que fomenta el pensamiento crítico y las habilidades cognitivas de orden superior (Hoo et al., 2022; Ibarra-Sáiz et al., 2020). Además, supone una ayuda en la identificación de los objetivos, cómo alcanzarlos, así como una guía para la mejora de su trabajo (Burgess y Mellis, 2015; Davis et al., 2022).

Por otra parte, también se fomentan las habilidades de resolución de problemas al analizar el trabajo de los demás y tratar de identificar áreas de mejora aportando críticas constructivas (Dochy et al., 1999; Schellekens et al., 2021; Somervell, 1993). En el proceso, el estudiantado interactúa de manera significativa con sus iguales, con lo que se fomentan habilidades como la negociación y se contribuye a la formación de un entorno de aprendizaje más colaborativo y participativo (Cheng y Warren, 1997; Lerchenfeldt y Taylor, 2020). La literatura científica demuestra claramente que la evaluación por pares afecta al aprendizaje real, pero quizás el efecto es aún más acusado en el aprendizaje percibido (Hoo et al., 2022). Es más, como estrategia de aprendizaje colaborativo, la evaluación por pares ha demostrado ser de gran utilidad en contextos de educación superior, mostrando a su vez una eficacia diferencial en función del área de conocimiento o grado universitario cursado por el estudiantado (Huisman et al., 2020). Por este motivo, la investigación previa propone estudiar ese tipo de estrategias de evaluación por pares a través de la percepción del alumnado procedente de diferentes disciplinas universitarias (Huisman et al., 2018).

A pesar de las ventajas de incluir la evaluación por pares y de que todo el personal docente e investigador universitario esté íntimamente familiarizado con este proceso asociado a la publicación científica, en el que otros profesionales evalúan críticamente los manuscritos y proporcionan comentarios que pueden utilizarse para mejorar el trabajo (Alqassab et al., 2023), no es habitual que el estudiantado universitario tenga la oportunidad de aprovecharse de un proceso formativo similar y de introducir mejoras en su trabajo antes de presentarlo para su evaluación final (Zong et al., 2023). La evaluación entre iguales tiende a ser ineficaz o a ser rechazada por los estudiantes si no tienen la adecuada capacidad de retroalimentación (Han y Xu, 2020). Por lo tanto, es importante preparar a los

estudiantes para la evaluación por pares dotándolos de las competencias y disposiciones para superar los desafíos que puedan encontrar en este proceso (Zhan, 2021). Aunque investigaciones recientes hayan debatido acerca de la preparación de los estudiantes universitarios para afrontar la evaluación por pares, (Carless, 2022; Wood, 2021; Yu y Liu, 2021), las discusiones se han limitado a la perspectiva de los académicos y rara vez han considerado la perspectiva de los estudiantes (Zhan, 2021).

A pesar de que en las encuestas de calidad de la enseñanza universitaria la retroalimentación es un aspecto vinculado positivamente a la satisfacción del estudiantado (Estriegana et al., 2024) y que la evidencia empírica demuestra sus múltiples beneficios, menos de un tercio considera que han recibido una valoración útil de forma regular (Krause et al., 2005). Varias investigaciones han hallado que las creencias y el punto de vista del alumnado sobre la evaluación colaborativa son factores determinantes de la aplicación satisfactoria de esta estrategia de evaluación en contextos universitarios (Tornwall et al., 2022). Se debe fundamentalmente a las preocupaciones relacionadas con la validez, la fiabilidad, el sesgo, la equidad y la objetividad de los criterios al evaluar el trabajo o recibir la evaluación de otra persona que se encuentra en el mismo nivel, por lo que se prefiere que la evaluación continúe siendo responsabilidad del equipo docente (Carlsson Hauff y Nilsson, 2022; Gaynor, 2020; Panadero y Alqassab, 2019; Schellekens et al., 2021). Conocer las creencias del estudiantado a cerca de esta modalidad de evaluación permitirá mejorar el conocimiento general sobre cómo percibe el estudiantado la evaluación por pares y el desarrollo de estrategias que mejoren su compromiso y así lograr una implementación satisfactoria de la misma (Gaynor, 2020; Huisman et al., 2018).

A pesar de su importancia, la investigación sobre las creencias del estudiantado acerca de la revisión por pares es escasa. La mayoría de los estudios realizados hasta la fecha se han centrado en la evaluación sumativa entre pares, en particular la contribución individual a los proyectos de grupo, y en cuestiones relativas a la fiabilidad y validez de tales métodos de evaluación (Kaufman y Schunn, 2011; Shui Ng y Yu, 2021). En cambio, la percepción y las creencias del estudiantado a cerca de la evaluación por pares siguen siendo relativamente poco estudiadas, tal vez debido a la escasez de instrumentos de evaluación desarrollados para tal fin. Uno de los cuestionarios desarrollados recientemente para la evaluación de las creencias del estudiantado sobre la evaluación colaborativa es el *Beliefs about Peer Feedback Questionnaire* (BPFQ) (Huisman et al., 2020). Este instrumento de evaluación está compuesto por 10 ítems con una escala de respuesta tipo Likert de cinco opciones de respuesta (1 = *totalmente en desacuerdo*, 5 = *totalmente de acuerdo*) y tiene como objetivo evaluar las creencias del alumnado sobre la evaluación colaborativa mediante cuatro dimensiones: (1) la valoración de la retroalimentación por pares como método instructivo (VIM), (2) la confianza en la propia calidad de la retroalimentación por pares (CO), (3) la confianza en la calidad de la retroalimentación por pares recibida (CR) y (4) la valoración de la retroalimentación por pares como una habilidad esencial (VPS). Las puntuaciones en este cuestionario presentan propiedades psicométricas adecuadas, con una fiabilidad que oscila entre .67 y .76. Considerando la escasez de instrumentos válidos y fiables en español para evaluar las creencias del estudiantado sobre la retroalimentación por pares como estrategia de evaluación en el contexto universitario, y teniendo en cuenta las adecuadas propiedades psicométricas que presenta la versión original del BPFQ, el objetivo principal del presente trabajo es la adaptación transcultural del cuestionario al español, así como el análisis de sus propiedades psicométricas. Además, el estudio también tiene como objetivo analizar las posibles diferencias en las creencias hacia este tipo de evaluación entre el estudiantado de diferentes áreas de conocimiento.

Método

Diseño

Se realizó un estudio instrumental (Carretero-Dios y Pérez, 2007; García-Celay y León, 2005). La adaptación del instrumento BPFQ al español se llevó a cabo en dos fases. En la primera fase se desarrolló un proceso estándar para la validación lingüística del cuestionario al contexto español y a continuación, en la segunda fase, se determinó la equivalencia factorial de la versión española de BPFQ con el cuestionario original en inglés, evaluándose finalmente sus propiedades psicométricas.

Fase I: Validación Lingüística

El proceso de traducción-retrotraducción y de validación lingüística del BPFQ se llevó a cabo en cinco fases, como se indica en la Guía de Traducción y Adaptación de Tests (Gregoire, 2018; Hernández et al., 2020). Primero, dos traductores bilingües llevaron a cabo una traducción conceptual de la versión original del BPFQ al español. A continuación, se combinaron ambas traducciones para crear una única versión preliminar del cuestionario en español. En tercer lugar, dos nuevos traductores bilingües tradujeron la versión preliminar en español a la lengua original del cuestionario, creando una versión final en inglés. En el cuarto paso de este proceso, los dos últimos traductores compararon las dos versiones retro-traducidas con la original para comprobar la equivalencia lingüística, teniendo en cuenta las diferencias culturales. En quinto lugar, un comité formado por cuatro expertos españoles en evaluación educativa analizó la versión final al español asegurando la precisión cultural y lingüística del cuestionario traducido. Las posibles discrepancias fueron resueltas consultando los diferentes pasos llevados a cabo en el proceso. Finalmente, este comité no consideró necesario modificar ningún ítem y la evaluación general de la versión española del instrumento fue positiva.

Fase II: Evaluación de las Propiedades Psicométricas

Participantes

La muestra de este estudio estaba constituida por 607 estudiantes de la Universidad de Alicante, España. Los criterios de inclusión para participar en el estudio fueron: (1) ser estudiante en la Universidad de Alicante, tanto en estudios de Grado como de Postgrado, (2) ser capaz de leer y completar los cuestionarios por sí mismo y (3) haber firmado el consentimiento informado para participar en el estudio. No hubo restricciones en cuanto a la especialidad académica. De los 607 estudiantes, el 75.95% eran mujeres, con una edad media de 21.59 años ($DT = 6.96$). Alrededor del 11% pertenecía al área de conocimiento de Ciencias ($n = 67$), el 70.51% eran estudiantes de Ciencias Sociales y Jurídicas ($n = 428$) y el 18.45% del área de Educación ($n = 112$).

Instrumentos

Cuestionario Sociodemográfico *ad hoc*. Se elaboró un cuestionario sociodemográfico para recopilar información sobre la edad de las personas participantes (en años), el sexo (hombre/mujer), el nivel educativo alcanzado (grado/postgrado) y el área de conocimiento al que pertenecen sus estudios.

Beliefs about Peer-Feedback Questionnaire (BPFQ). Se trata de un cuestionario autoinformado compuesto por 10 ítems que se utiliza para evaluar las creencias del estudiantado sobre la evaluación por pares como estrategia en el contexto universitario (Huisman et al., 2020). Concretamente, esta herramienta se divide en cuatro dimensiones: (1) valoración de la retroalimentación por pares como método instructivo (VIM; ítems 1-3), (2) confianza en la propia calidad de

la retroalimentación por pares (CO; ítems 4-5), (3) confianza en la calidad de la retroalimentación por pares recibida (CR; ítems 6-7) y (4) valoración de la retroalimentación por pares como una habilidad esencial (VPS; ítems 8-10). Todos los ítems se responden a través de una escala Likert de 5 puntos; para las escalas VIM y VPS, las etiquetas van de 1 (*totalmente en desacuerdo*) a 5 (*totalmente de acuerdo*); para las escalas CO y CR, las etiquetas van de 1 (*en absoluto aplicable a mí*) a 5 (*totalmente aplicable a mí*). El ítem 3, "La retroalimentación únicamente debe ser proporcionada por el personal docente", se considera inverso. Puntuaciones elevadas indican creencias positivas y unas puntuaciones más bajas representan unas creencias negativas sobre la evaluación por pares. Las propiedades psicométricas del BPFQ fueron originalmente analizadas utilizando una muestra de 340 estudiantes en Holanda (Huisman et al., 2020). Los resultados del análisis de componentes principales revelaron bajos valores en las communalidades (.624-.718), por lo que los ítems fueron correctamente distribuidos en cuatro dimensiones. Asimismo, un análisis factorial confirmatorio sugirió la idoneidad de las cuatro dimensiones en el cuestionario (Huisman et al., 2020). Las puntuaciones del cuestionario original también presentaron una adecuada fiabilidad, con valores de alfa de Cronbach que oscilan entre .67 y .76.

Procedimiento

En primer lugar se contactó con el profesorado de diferentes titulaciones de la Universidad de Alicante para informarles del propósito de la investigación e invitarles a participar. Dicho profesorado seleccionó las clases que iban a participar explicando al estudiantado el objetivo de la investigación. Así, administraron el protocolo de evaluación, que incluye el consentimiento informado, el cuestionario sociodemográfico y la versión española del BPFQ. Los estudiantes dispusieron de un tiempo estimado de entre 10 y 20 minutos, suficiente para la cumplimentación de la evaluación. La participación fue anónima y voluntaria y no recibieron remuneración económica alguna. Se administró y cumplimentó un total de 607 cuestionarios.

El estudio fue aprobado por el Comité de Ética de la Universidad de Alicante (UA-2022-02-16) y para su desarrollo se siguieron las recomendaciones éticas recogidas en los Principios Éticos para Investigaciones Médicas (Declaration of Helsinki; World Medical Association, 2013). Se informó a todas las personas que participaron sobre el objetivo del estudio y la confidencialidad de los datos recogidos, firmando el consentimiento informado previamente a su participación, explicando que su autorización podía ser revocada o cancelada en cualquier momento según su propia voluntad.

Análisis de Datos

Se calcularon estadísticos descriptivos, como media, desviación típica, rango, percentiles 25 y 75, además del efecto suelo/techo de los ítems. Para la interpretación del efecto suelo/techo se consideró significativo el 15% o más de participantes que mostraran puntuaciones altas (efecto techo) o puntuaciones bajas (efecto suelo) (Gulledge et al., 2019).

Se calcularon correlaciones bivariadas de Pearson entre los ítems del test y entre las cuatro subescalas y el factor general del cuestionario para comprobar la colinealidad entre dimensiones y la propia utilidad de las mismas. Para la interpretación de estas correlaciones, según Hernández-Lalinde et al. (2018), se considerarían correlaciones nulas los valores iguales o menores a .10, débiles entre .11 y .30, moderadas entre .31 y .50 y correlaciones fuertes cuando los valores son iguales o mayores a .51.

El análisis factorial confirmatorio se llevó a cabo con el programa R (Browne y Cudeck, 1992; Epskamp et al., 2019; Green y Yang, 2009; Kline, 2015; Rosseel et al., 2012; The R Project for Statistical Computing, 2021), utilizando el método más robusto de máxima

probabilidad (ML) para lidiar con la distribución no normal de los datos de esta muestra (Satorra y Bentler, 1994). Concretamente, se realizaron tres AFC, el primero compuesto por un único factor general del cuestionario, el segundo siguiendo la estructura factorial del cuestionario original y su adaptación española de Ruiz-Robledillo et al. (2022), que consta de cuatro factores: VIM, CO, CR y VPS, y el tercero formado por un modelo bifactorial en el que se incluyen tanto el factor general del cuestionario como sus cuatro subescalas. El ajuste del modelo se consideró satisfactorio si χ^2_{s-b} era no significativo ($p > .05$), el índice de ajuste (CFI) era .90 o superior, la raíz del error cuadrático medio (RMSEA) se encontraba entre .05 y .08, y el índice de la raíz cuadrada media residual estandarizada (SRMR) era .08 o menor (Browne y Cudeck, 1992; Green y Yang, 2009; Kline, 2015).

También se calculó la consistencia interna de la puntuación total del test con los indicadores alfa de Cronbach (α) y omega de McDonald (ω) en R (The R Project for Statistical Computing, 2021). A pesar de que el α de Cronbach es el coeficiente más común para evaluar y estimar la consistencia interna omega (ω) (McDonald, 1999) está basada en un modelo factorial que no se sustenta en los supuestos de equivalencia tau (es decir, unidimensionalidad, igualdad de varianzas y covarianzas de las puntuaciones esperadas para los ítems) (Rodríguez et al., 2016). Según Taber (2018), en Ciencias de la Educación valores de alfa de Cronbach entre .67 y .87 podrían considerarse también razonables. Consideramos los α de Cronbach de .60 o inferiores como bajos, entre .60 y .70 como aceptables y .70 o superiores como buenos (Leary, 2008). Para ω , consideramos los valores de .60 o inferiores como bajos, entre .60 y .70 como aceptables y .70 o superior como buenos (McDonald, 1999). Además, para analizar la fiabilidad del modelo bifactorial se calcularon los indicadores de omega jerárquico de McDonald tanto para el factor único general (ω_H) como para las subescalas (ω_{HS} ; McDonald, 1999). El estimador de ω_H representa la proporción de varianza de la puntuación total del BPFQ explicado por el

factor general, controlando a los factores específicos. Asimismo, el valor de ω_{HS} muestra la proporción de varianza explicada por las subescalas del BPFQ, controlando el factor general del cuestionario (Rodríguez et al., 2016). Para la interpretación de los indicadores de consistencia interna relacionados con el modelo bifactorial, se establece que cuanto mayores son los valores de ω_H la interpretación del cuestionario se debe a las puntuaciones totales del factor general del BPFQ; de esta manera, cuanto mayores sean los valores de ω_{HS} más se debe la interpretación del cuestionario a las puntuaciones totales de las subescalas del BPFQ. Asimismo, se calcularon los indicadores relacionados con el porcentaje de correlaciones no contaminadas (PUC) por la multidimensionalidad, donde los valores $> .80$ son considerados adecuados, así como los indicadores asociados con la varianza común explicada que se debe al factor general, mostrando favorablemente la unidimensionalidad con valores $> .60$, cuando el PUC refleja también valores elevados $> .80$ y el indicador de consistencia interna del modelo bifactorial $> .70$ (Reise et al., 2013; Rodríguez et al., 2016). También se calcularon los porcentajes de la varianza verdadera de cada ítem que es explicada por el factor general del instrumento (ECV-I), en el que valores mayores a .80 muestran una influencia significativa del factor general (Stucky y Edelen, 2015; Stucky et al., 2013). Los cálculos fueron realizados con el módulo IndicesBifactor.xls (Domínguez-Lara y Rodríguez, 2017).

Asimismo, se analizaron las diferencias entre las áreas de conocimiento a las que pertenecen el estudiantado evaluado (Ciencias, Educación y Ciencias Sociales y Jurídicas). Para la elección de la prueba de contraste de hipótesis más idónea, se comprobaron los supuestos estadísticos de normalidad, independencia y homocedasticidad (Field, 2009). Los supuestos estadísticos no se cumplieron para la muestra de estudio, por lo que se aplicó la prueba no paramétrica H de Kruskal-Wallis (ver Tabla 7), en la que se determinaron diferencias estadísticamente significativas cuando $p < .05$. Además, se realizaron pruebas *post hoc* a través de la prueba U

Tabla 1. Análisis descriptivo de los ítems del BPFQ

	R	M (DE)	P25	P75	Efecto suelo (%)	Efecto techo (%)
Ítem 1. Involucrar a los/as estudiantes mediante el uso de la retroalimentación por pares es importante.	1-5	3.572 (0.964)	3	4	10.9%	14.3%
Ítem 2. La retroalimentación por pares es útil.	1-5	3.619 (0.926)	3	4	10.1%	13.7%
Ítem 3. La retroalimentación únicamente debe ser proporcionada por el personal docente.	1-5	3.148 (0.996)	2	4	5.1%	6.8%
Ítem 4. En general, estoy seguro/a de que la retroalimentación que proporciono a otr@s estudiantes es de buena calidad.	1-5	3.399 (0.793)	3	4	10.2%	5.9%
Ítem 5. En general, estoy seguro/a de que la retroalimentación que proporciono a otr@s estudiantes les ayuda a mejorar su trabajo.	1-5	3.326 (0.844)	3	4	13.2%	6.8%
Ítem 6. En general, estoy seguro/a de que la retroalimentación que recibo de otros/as estudiantes es de buena calidad.	1-5	3.338 (0.840)	3	4	13.5%	6.6%
Ítem 7. En general, estoy seguro/a de que la retroalimentación que recibo de otros/as estudiantes me ayuda a mejorar mi trabajo.	1-5	3.395 (0.918)	3	4	14.8%	9.7%
Ítem 8. Ser capaz de proporcionar una retroalimentación constructiva a los/as compañeros/as es una habilidad importante.	1-5	4.058 (0.835)	4	4	22.9%	32.6%
Ítem 9. Ser capaz de afrontar la retroalimentación crítica por parte de los/as compañeros/as es una habilidad importante.	1-5	4.005 (0.882)	3	4	4.1%	31.6%
Ítem 10. Ser capaz de mejorar el trabajo realizado por uno/a mismo/a en función de la retroalimentación recibida por parte de los/as compañeros/as es una habilidad importante.	1-5	4.054 (0.824)	4	4	22.2%	31.8%
VIM	-	3.446 (0.759)	3	4	-	-
CO	-	3.362 (0.741)	3	4	-	-
CR	-	3.367 (0.819)	3	4	-	-
VPS	-	4.039 (0.768)	3.667	4.667	-	-
BPFQ general factor	-	3.592 (0.569)	3.200	4.000	-	-

Nota. VIM = valoración de la retroalimentación por pares como método instructivo; CO = confianza en la propia calidad de la retroalimentación por pares; CR = confianza en la calidad de la retroalimentación por pares recibida; VPS = valoración de la retroalimentación por pares como una habilidad esencial; BPFQ = Beliefs in Peer Feedback Questionnaire; R = rango, M = media, DE = desviación estándar, P25 = percentil 25, P75 = percentil 75.

Tabla 2. Correlaciones Bivariadas de Pearson entre los Ítems del BPFQ

	1	2	3	4	5	6	7	8	9	10					
It1	-														
It2	.675***	-													
It3	.343***	.292***	-												
It4	.284***	.281***	.082	-											
It5	.302***	.307***	.123**	.639***	-										
It6	.373***	.342***	.179***	.431***	.505***	-									
It7	.389***	.414***	.194***	.402***	.442***	.733***	-								
It8	.410***	.447***	.218***	.371***	.374***	.377***	.428***	-							
It9	.356***	.409***	.180***	.311***	.357***	.370***	.417***	.746***	-						
It10	.416***	.462***	.222***	.303***	.388***	.424***	.445***	.664***	.786***	-					
VIM	.848***	.821***	.702***	.270***	.307***	.375***	.418***	.451***	.396***	.461***	-				
CO	.324***	.325***	.114**	.899***	.911***	.519***	.467***	.411***	.370***	.383***	.319***	-			
CR	.410***	.408***	.200**	.447***	.507***	.924***	.937***	.433***	.424***	.467***	.427***	.528***	-		
VPS	.434***	.484***	.227***	.362***	.411***	.430***	.474***	.886***	.934***	.899***	.480***	.427***	.486***	-	
BPFQ GF	.666***	.693***	.255***	.602***	.648***	.704***	.732***	.738***	.734***	.750***	.675***	.691***	.772***	.817***	-

Nota. VIM = valoración de la retroalimentación por pares como método instructivo; CO = confianza en la propia calidad de la retroalimentación por pares; CR = confianza en la calidad de la retroalimentación por pares recibida; VPS = valoración de la retroalimentación por pares como una habilidad esencial; BPFQ GF = Beliefs in Peer Feedback Questionnaire General Factor.

*** $p < .001$.

de Mann-Whitney y se utilizó también la corrección de Bonferroni para prevenir el error tipo 1. Además, se calcularon los tamaños del efecto para conocer la magnitud de esas diferencias. Se consideró un tamaño del efecto pequeño cuando $r_{bis} < .30$, mediano cuando $r_{bis} = .30-.50$ y grande cuando $r_{bis} > .50$ (Field, 2009). Estos análisis se llevaron a cabo con el programa Jamovi versión 1.6.23.

Resultados

Análisis Descriptivo de los Ítems

En relación a los análisis descriptivos, se calculó la media, desviación típica, rango y percentil de cada uno de los ítems del BPFQ. Se utilizó una escala de respuesta tipo de Likert de 5 opciones, y se observó un efecto techo en tres de los diez ítems. Los resultados se muestran en la [Tabla 1](#).

Análisis Correlacional entre los Ítems, las Dimensiones y el Factor General del Cuestionario

En cuanto a las correlaciones bivariadas de Pearson entre los ítems del cuestionario, la relación existente entre los ítems dentro de una misma dimensión es más fuerte, siendo la correlación más débil entre ítems pertenecientes a diferentes dimensiones ([Tabla 2](#)).

En relación a las correlaciones entre las subescalas del BPFQ, la dimensión VIM presenta correlaciones positivas, moderadas y significativas con CO ($r = .319***$), CR ($r = .427***$) y VPS ($r = .480***$). Asimismo, CO presenta una correlación positiva, fuerte y significativa con CR ($r = .528***$) y positiva, moderada y significativa con VPS ($r = .427***$). Por último, CR y VPS también presentan una correlación positiva, moderada y significativa ($r = .486***$). Finalmente, el factor general del cuestionario correlaciona de forma positiva, fuerte y significativa con el resto de subescalas: VIM ($r = .675***$), CO ($r = .691***$), CR ($r = .772***$) y VPS ($r = .817***$).

Análisis Factorial Confirmatorio

Se llevaron a cabo tres análisis factoriales confirmatorios (AFC) para comprobar si la versión española del BPFQ cumplía con la

estructura factorial del cuestionario original. El primer AFC está compuesto por un único factor general. Los valores de ajuste del modelo no son satisfactorios, $\chi^2(35) = 964.342$, CFI = .688, TLI = .599, RMSEA = .209, SRMR = .099, 95% IC (.198, .221). Atendiendo a la versión original de [Huisman et al. \(2020\)](#), se realizó un segundo AFC compuesto por cuatro dimensiones. Los resultados corresponden a un modelo tetrafactorial: (1) valoración de la retroalimentación por pares como método instructivo (VIM), (2) confianza en la propia calidad de la retroalimentación por pares (CO), (3) confianza en la calidad de la retroalimentación por pares recibida (CR) y la (4) valoración de la retroalimentación por pares como una habilidad esencial (VPS). Los datos muestran un ajuste satisfactorio del modelo de cuatro factores idéntico al original, $\chi^2(29) = 85.939$, CFI = .980, TLI = .970, RMSEA = .057, SRMR = .026, 95% IC (.043, .071). Un tercer AFC se llevó a cabo con un modelo bifactorial, en el que se incluye tanto el factor general del cuestionario como sus subescalas. Los valores de ajuste para este modelo son satisfactorios, siendo $\chi^2(15) = 25.129$, CFI = .997, TLI = .990, RMSEA = .033, SRMR = .012, 95% IC (.003, .055). La [Tabla 3](#) indica los valores de PUC, ECV y ECV-I como indicadores de específicos del modelo bifactorial. El 68% de la varianza en los ítems se debe al factor general ($\omega_H = .823$). Los valores de PUC = .822 y ECV = .678 indican la influencia significativa del factor general. Además, el aporte de los factores específicos no es significativo, pues los valores ECV-I superan el .60 en su mayoría ([Tabla 3](#)).

Tabla 3. Indicadores específicos del modelo bifactorial

PUC	ECV	VIM: ECV-I	CO: ECV-I	CR: ECV-I	VPS: ECV-I
.822	.678	.510	.837	.814	.610
		.542	.880	.727	.889
		.550			.638

Nota. PUC = porcentaje de correlaciones no contaminadas; ECV = varianza común explicada; ECV-I = porcentaje de varianza verdadera de cada ítem explicada por el factor general.

Por otro lado, las cargas factoriales de la estructura de cuatro factores exceden el valor de .40, demostrando un rango desde .594 hasta .799, salvo el ítem 3 (VIM3), que tiene una carga factorial de .385 ([Figura 1](#)). Sin embargo, por motivos teóricos y metodológicos, se decidió mantener todos los ítems del test.

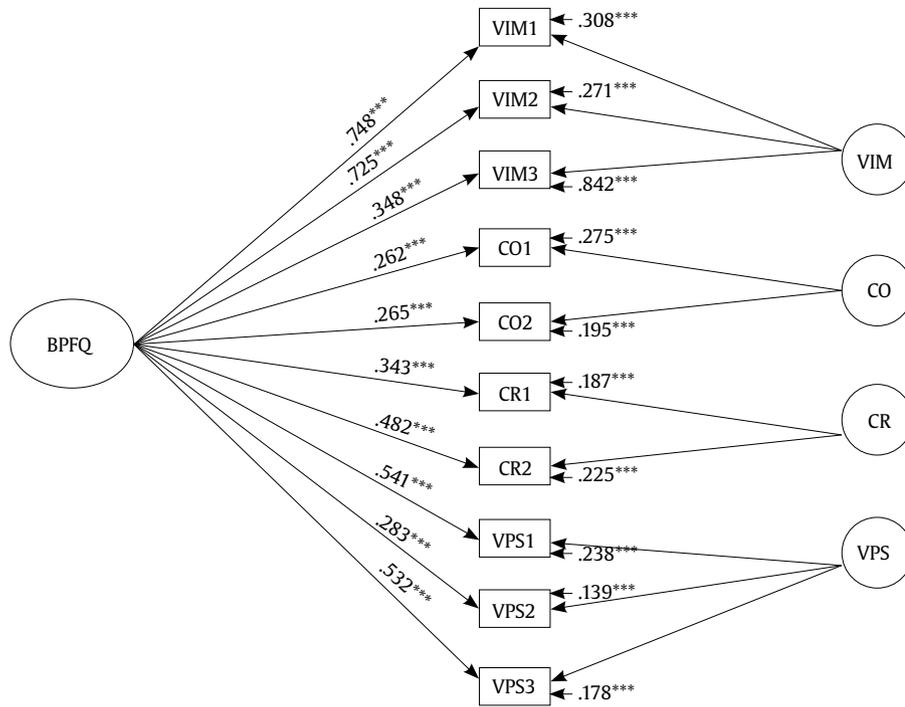


Figura 1. Cargas factoriales del modelo bifactorial compuesto por el factor general y los cuatro factores de la versión española del BPFQ ($N = 607$).

VIM = valoración de la retroalimentación por pares como método instructivo: It1, It2, e It3; CO = confianza en la propia calidad de la retroalimentación por pares: It4, e It5; CR = confianza en la calidad de la retroalimentación por pares recibida: It6, e It7; VPS = valoración de la retroalimentación por pares como una habilidad esencial: It8, It9, e It10. *** $p < .001$.

Tabla 4. Características psicométricas de la versión española del BPFQ

Ítem	ω de McDonald si se elimina el ítem	α Cronbach si se elimina el ítem	Indicador de discriminación ($r_{it,c}$)
VIM1	.842	.848	.590
VIM2	.839	.846	.605
VIM3	.868	.874	.293
CO1	.847	.855	.497
CO2	.843	.851	.552
CR1	.837	.846	.614
CR2	.834	.843	.635
VPS1	.844	.841	.668
VPS2	.845	.843	.644
VPS3	.843	.840	.683

Nota. VIM = valoración de la retroalimentación por pares como método instructivo; CO = confianza en la propia calidad de la retroalimentación por pares; CR = confianza en la calidad de la retroalimentación por pares recibida; VPS = valoración de la retroalimentación por pares como una habilidad esencial; BPFQ = *Beliefs in Peer Feedback Questionnaire*; $r_{it,c}$ = correlación del ítem con el total del test.

Fiabilidad

Todos los valores de correlación ítem-resto del test como indicador de discriminación de los ítems se encuentran por encima de .30, excepto el ítem 3 ($r_{it,c} = .293$) (Tabla 4). El valor del coeficiente de consistencia interna de fiabilidad para el factor general se encuentra por encima de .80 (siendo $\alpha = .862$, 95% IC [.845, .878] y $\omega = .857$, 95% IC [.840, .874]). Además, tan solo este indicador de fiabilidad general aumentaría ligeramente en caso de eliminar el ítem 3 (“La retroalimentación únicamente debe ser proporcionada por el personal docente”).

Además, se obtuvieron los valores de fiabilidad a través del coeficiente de consistencia interna alfa de Cronbach y omega de McDonald para cada dimensión del instrumento BPFQ, siendo todos ellos adecuados (Tabla 5). En relación a los valores de Omega Jerárquico, el factor general explica alrededor de un 87% de la varianza total del cuestionario, controlando a las subescalas. Además, las subescalas CR y VPS parecen explicar más de un 84% de la varianza total del cuestionario que, con respecto a las otras dos

Tabla 5. Coeficientes de consistencia interna de fiabilidad para las dimensiones del BPFQ

Dimensión	α de Cronbach	95% IC	ω de McDonald	95% IC	ω de McDonald jerárquico (ω_{HS}/ω_H)
VIM	.696	.650, .737	.730	.694, .766	.725
CO	.779	.740, .811	.778	.175, .825	.786
CR	.844	.817, .867	.844	.805, .875	.846
VPS	.891	.875, .906	.896	.882, .910	.899
FG BPFQ	.862	.845, .878	.865	.845, .882	.823

Nota. VIM = valoración de la retroalimentación por pares como método instructivo; CO = confianza en la propia calidad de la retroalimentación por pares; CR = confianza en la calidad de la retroalimentación por pares recibida; VPS = valoración de la retroalimentación por pares como una habilidad esencial; FG BPFQ = Factor General del *Beliefs in Peer Feedback Questionnaire*; IC = Intervalo de confianza.

Tabla 6. Prueba H de Kruskal-Wallis. Diferencias en el BPFQ entre áreas de conocimiento (Ciencias, Educación y Ciencias Sociales y Jurídicas)

Dimensión	Área de conocimiento	Rango promedio	H	p
VIM	Ciencias	364.49	58.955	< .001
	Educación	399.54		
	Ciencias Sociales y Jurídicas	269.53		
CO	Ciencias	388.81	40.772	< .001
	Educación	360.33		
	Ciencias Sociales y Jurídicas	275.85		
CR	Ciencias	393.53	49.957	< .001
	Educación	369.90		
	Ciencias Sociales y Jurídicas	272.74		
VPS	Ciencias	371.28	39.520	< .001
	Educación	372.38		
	Ciencias Sociales y Jurídicas	275.57		

Nota. VIM = valoración de la retroalimentación por pares como método instructivo; CO = confianza en la propia calidad de la retroalimentación por pares; CR = confianza en la calidad de la retroalimentación por pares recibida; VPS = valoración de la retroalimentación por pares como una habilidad esencial; BPFQ = *Beliefs in Peer Feedback Questionnaire*.

Tabla 7. Pruebas *post-hoc*. Prueba U de Mann-Whitney para dos muestras independientes

Dimensión	Área de conocimiento	z	p	r_{bis}
VIM				
Ciencias	Educación	-1.383	.167	.056
	Ciencias Sociales y Jurídicas	-5.004	< .001	.203
Educación	Ciencias Sociales y Jurídicas	-6.597	< .001	.268
CO				
Ciencias	Educación	-0.912	.362	.004
	Ciencias Sociales y Jurídicas	-5.134	< .001	.208
Educación	Ciencias Sociales y Jurídicas	-4.648	< .001	.189
CR				
Ciencias	Educación	-1.239	.215	.050
	Ciencias Sociales y Jurídicas	-5.313	< .001	.216
Educación	Ciencias Sociales y Jurídicas	-5.483	< .001	.223
VPS				
Ciencias	Educación	-0.214	.830	.009
	Ciencias Sociales y Jurídicas	-4.275	< .001	.174
Educación	Ciencias Sociales y Jurídicas	-5.248	< .001	.213

Nota. VIM = valoración de la retroalimentación por pares como método instructivo; CO = confianza en la propia calidad de la retroalimentación por pares; CR = confianza en la calidad de la retroalimentación por pares recibida; VPS = valoración de la retroalimentación por pares como una habilidad esencial; r_{bis} = rango biserial de Glass como indicador del tamaño del efecto para la U de Mann-Whitney, prueba no paramétrica para el análisis de diferencias entre diferentes muestras.

dimensiones, controlando el factor general (Tabla 5). A pesar de la influencia significativa del factor general, los valores de ω_{HS} de los factores específicos son adecuados (> .40).

Diferencias en BPFQ según el Área de Conocimiento

En las Tablas 6 y 7 se recogen los resultados obtenidos de las pruebas no paramétricas para comprobar las diferencias existentes entre las áreas de conocimiento. Se encontraron diferencias estadísticamente significativas en todas las dimensiones del cuestionario adaptado al español entre, al menos, dos de los grupos pertenecientes a las áreas de conocimiento del estudiantado. Concretamente, en las áreas de Ciencias y Ciencias Sociales y Jurídicas se hallaron diferencias estadísticamente significativas en todas las dimensiones, presentando puntuaciones más altas el área de Ciencias (Tablas 6 y 7). Por último, también se encontraron diferencias estadísticamente significativas en todas las dimensiones del cuestionario en las áreas de Educación y Ciencias Sociales y Jurídicas, mostrando una mayor puntuación en todas las dimensiones el área de Educación (Tabla 6).

Discusión

A pesar de que la literatura científica previa ha demostrado que la evaluación por pares juega un rol esencial en el aprendizaje y aporta grandes beneficios a los estudiantes, como son la capacidad autorreflexiva (Carlsson Hauff y Nilsson, 2022; Schellekens et al., 2021), la inclusión de este tipo de metodologías en el contexto universitario sigue siendo relativamente escasa, entre otras razones por la falta de instrumentos específicos para la evaluación de su aceptación por parte de los estudiantes (van Popta et al., 2017). En este contexto cabe destacar la necesidad de disponer de herramientas de evaluación válidas y fiables (Ramos, 2018) para identificar las creencias del estudiantado acerca de este tipo de evaluación, así como generalizar su inclusión generalizada en el contexto universitario a la vista de los beneficios que aporta. A ello se debe que el presente estudio haya tenido como objetivo adaptar transculturalmente el cuestionario elaborado por Huisman et al. (2020), *Beliefs about Peer-Feedback Questionnaire* (BPFQ), que evalúa las creencias sobre la evaluación por pares entre los estudiantes. Además, se han analizado sus propiedades psicométricas y las posibles diferencias en estas creencias entre el estudiantado de diferentes áreas de conocimiento.

Los análisis correlacionales entre los ítems, las subescalas y el factor general del BPFQ han demostrado también un potencial de discriminación entre sí, además de presentar validez de contenido en las dimensiones, al mostrar correlaciones más elevadas los ítems que pertenecen a las mismas dimensiones. Asimismo, las correlaciones entre las dimensiones parecen ir en la misma dirección que en la versión original del cuestionario (Huisman et al., 2020). De hecho, estas subescalas también presentan correlaciones elevadas y positivas con el factor general del cuestionario, lo que podría significar validez de constructo y unidad del instrumento para evaluar las creencias sobre la evaluación por pares en el estudiantado universitario español.

En relación a los análisis factoriales confirmatorios, los resultados muestran un claro índice de ajuste de la estructura factorial del cuestionario en español, de acuerdo con el análisis factorial confirmatorio, replicando la estructura original de la misma propuesta por Huisman et al. (2020). De hecho, este estudio ha revelado un ajuste muy satisfactorio de una estructura compuesta por cuatro factores, así como una gran precisión del instrumento a través de los indicadores de consistencia interna y fiabilidad para todas las dimensiones y el factor general en las puntuaciones de la versión española del BPFQ. Sin embargo, el que mejores indicadores de ajuste presenta es el modelo bifactorial, que incluye tanto el factor general del cuestionario como las subescalas. Sin embargo, cabe destacar también la influencia significativa del factor general en los ítems que componen cada dimensión, obtenida a raíz de los valores PUC y ECV (Reise et al., 2013). Para el estudiantado universitario, el factor general del BPFQ explica su percepción de la evaluación por pares y refuerza su evaluación a través de puntuaciones directas para la medida del constructo general. De hecho, las puntuaciones de la versión española también permiten que sea interpretado el constructo no solo a través de su factor general sino a través de las cuatro dimensiones que replica de la versión original del cuestionario: (1) valoración de la retroalimentación por pares como método instructivo (VIM), (2) confianza en la propia calidad de la retroalimentación por pares (CO), (3) confianza en la calidad de la retroalimentación por pares recibida (CR) y (4) valoración de la retroalimentación por pares como una habilidad esencial (VPS). Estos resultados son coherentes con la versión original propuesta por Huisman et al. (2020).

Con respecto a los indicadores de fiabilidad obtenidos en las puntuaciones de la versión española, tanto el factor general del BPFQ como el de cada subescala son adecuados para la muestra de estudiantado universitario, por lo que parece que el instrumento tiene una buena precisión a la hora de evaluar el constructo en cuestión. De hecho, los indicadores de fiabilidad mostrados a través de los coeficientes de consistencia interna, alfa de Cronbach y omega de McDonald, tanto del factor general como de cada subescala, son similares a los del cuestionario original (Huisman et al., 2020). Además, el cálculo de los valores de omega de McDonald jerárquico también indica que el constructo del cuestionario también puede ser interpretable a través de las puntuaciones del factor general, con independencia de las subescalas, y viceversa.

Respecto a las diferencias en las creencias respecto a la evaluación por pares en función del área de conocimiento, se han observado diferencias estadísticamente significativas en el estudiantado. Concretamente, aunque las diferencias de las puntuaciones entre el estudiantado de Ciencias de la Educación no sean significativas, siendo las de los primeros más elevadas en las dimensiones VIM y VPS, el estudiantado del área de Educación es el grupo que alcanza puntuaciones significativamente más elevadas que con respecto al grupo de Ciencias Sociales en todas las dimensiones del cuestionario. Por tanto, de acuerdo con nuestros resultados, el estudiantado de Educación parece valorar más positivamente la evaluación por pares. Estos resultados además son congruentes con el estudio de Huisman et al. (2020), en el que el estudiantado de Ciencias Sociales y Educación presenta puntuaciones dispares, siendo las de este último las más elevadas. Entre los estudios realizados hasta la fecha en relación a la evaluación por pares,

la mayor parte se han llevado a cabo en el área de Educación, ya que parece ser de gran interés para la mejora de los procesos de toma de decisiones, construcción de conocimiento y juicios evaluativos del área (van Popta et al., 2017). Esto podría explicarse por la percepción de los estudiantes de la autoevaluación y su conocimiento de la misma, que indica que podría influir en la percepción del valor que se otorga a las evaluaciones recibidas del propio trabajo (Carlsson Hauff y Nilsson, 2022). Por otra parte, la creencia de los estudiantes sobre la falta de formación y experiencia de sus compañeros en la materia y en la realización de este tipo de evaluaciones (Han y Xu, 2020; Zhan, 2021) podría explicar las puntuaciones más bajas en los estudiantes de las áreas de conocimiento de Ciencias y Ciencias Jurídicas en comparación con los de Educación. Estos últimos estarían más formados en sistemas de evaluación novedosos y alternativos como parte de los contenidos propios de su formación y más abiertos a nuevas prácticas educativas, por lo que es más probable que perciban a sus compañeros como más capaces de realizar este tipo de evaluaciones y, por tanto, les concedan mayor valor. Este proceso dentro del aprendizaje puede motivar al estudiantado a desarrollar las habilidades asociadas al juicio valorativo y de interacción social de los estudiantes, que puede influir directamente en la calidad de su rendimiento académico (Hoo et al., 2022; van Popta et al., 2017). Teniendo en cuenta todos los resultados mencionados, nuestra investigación sugiere que el cuestionario autoinformado para la evaluación de las creencias sobre la retroalimentación por pares en el estudiantado universitario es rápido, adecuado y fiable para una medida general y de cada una de las subescalas relacionadas con estas creencias.

Fortalezas, Limitaciones y Líneas Futuras de Investigación

En cuanto a las fortalezas, en este estudio contamos con estudiantado de diferentes áreas de conocimiento, lo cual nos puede proporcionar información valiosa acerca de la utilidad que tiene este tipo de metodologías en el proceso de enseñanza-aprendizaje en diferentes áreas dentro de la universidad. Además, proporcionar este cuestionario a la comunidad científica y universitaria podría contribuir al fomento del conocimiento en estas áreas, así como a ampliar el concepto relacionado con diversos aspectos de la evaluación por pares y la influencia en las creencias que el estudiantado tiene sobre el mismo. Asimismo, la elaboración y adaptación de este tipo de instrumentos podría facilitar el formato de evaluación de estas creencias y cómo influye en la interacción entre el estudiantado (Huisman et al., 2020). Sin embargo, existen diversas limitaciones asociadas al presente estudio. En cuanto a la estructura del cuestionario, a pesar de que presenta adecuados indicadores de fiabilidad, sería interesante analizar en futuros estudios la validez interna del instrumento a través del cálculo de la validez discriminante y convergente. Otra de las limitaciones hace referencia al pequeño tamaño de la muestra, que podría plantear problemas a la hora de generalizar resultados de este estudio. Además, el estudiantado tiene una mala experiencia en relación a esta metodología de evaluación, lo que podría suponer un sesgo de respuesta si no la conocen, por lo que, aunque se encuentren en un nivel de educación superior, el conocimiento que poseen es muy escueto, lo que implica interpretar los resultados con cautela. Por tanto, en futuros estudios sería interesante la creación de grupos focales para acercarse aún más al alumnado y precisar la información relativa a esta metodología de evaluación por pares con objeto de incrementar el conocimiento del estudiante al respecto y mejorar la fiabilidad de los hallazgos de futuras investigaciones, así como la validez del constructo en cuestión. Además, en futuras investigaciones se podrían calcular las diferencias entre las diversas áreas de conocimiento con el objetivo de comprobar el nivel de creencias sobre esta metodología en el estudiantado.

En conclusión, podemos afirmar que las puntuaciones de la versión española del BPFQ son válidas y fiables para evaluar las creencias acerca de la evaluación por pares en estudiantado universitario al

replicar la mayoría de las propiedades psicométricas encontradas en la versión original: (I) una estructura factorial con un factor general sobre creencias acerca de la evaluación por pares y cuatro dimensiones específicas y (II) una adecuada consistencia interna del cuestionario en general. Asimismo, el estudio también presenta un modelo bifactorial que muestra indicadores de ajuste satisfactorios, a través de los cuales se puede afirmar que el constructo que evalúa este cuestionario para la versión en español puede ser interpretado tanto a través de las puntuaciones del factor general con independencia de las subescalas y viceversa. Además, los resultados también avalan que el estudiantado del área de Educación es el que más positivamente valora la evaluación por pares, por lo que este instrumento resulta de gran utilidad a la hora de introducir nuevas estrategias en el aprendizaje colaborativo de los estudiantes. De este modo, el fomento y la facilidad para evaluar la percepción de los estudiantes sobre la evaluación por pares repercute positivamente en la mejora del proceso de enseñanza-aprendizaje en este nivel educativo, construyendo así un conocimiento sobre los juicios valorativos y potenciando las habilidades de interacción entre estudiantes y su calidad de vida académica.

Extended Summary

Peer assessment, characterized by feedback, communication, reflection, and collaboration among peers, has emerged as a prominent educational approach (Darvishi et al., 2022). Recent studies have suggested that peer-feedback assessment is considered as an essential assessment model within the teaching-learning process based on the learning self-regulation (Erkan et al., 2021; Filius et al., 2018). Aligned with modern cognitive learning theory, peer learning strategies engage students in the assessment process, fostering deep learning approaches and intrinsic motivation (Barkley et al., 2014; Pinheiro et al., 2023). This collaborative approach not only enhances student engagement but also facilitates judgment skills, critical abilities, and self-awareness (Fang et al., 2021).

In higher education, integrating peer assessment changes the teaching-learning process from a teacher-led to a collaborative, student-centered approach, in line with social constructivism principles (Kasch et al., 2021; Vygotsky, 2012). Students benefit by expanding their knowledge, identifying strengths and weaknesses, and developing personal and professional skills (Deneen y Hoo, 2021). Despite its benefits, research on students' beliefs toward peer assessment remains limited, needing valid and reliable assessment tools (Huisman et al., 2020; Kaufman y Schunn, 2011).

The Beliefs about Peer Feedback Questionnaire (BPFQ) addresses this gap, assessing students' attitudes about collaborative assessment (Huisman et al., 2020). This study aims to cross-culturally adapt the BPFQ into Spanish and analyze its psychometric properties, considering potential differences in beliefs among students from different fields of study.

Method

Study Design

An instrumental study was conducted to adapt the BPFQ into Spanish and evaluate its psychometric properties. The adaptation took place in two phases: linguistic validation and psychometric properties evaluation.

Phase I: Linguistic Validation

The process of translating and validating the BPFQ involved five phases. Initially, two bilingual translators translated the questionnaire into Spanish, which was then combined into a preliminary version. Subsequently, two other translators translated this version

back into English, resulting in a final English version. Linguistic equivalence was checked in the fourth step, and a committee of Spanish experts reviewed the final Spanish version for cultural and linguistic accuracy. Discrepancies were resolved through consultation, resulting in a positively evaluated Spanish version without any item modifications.

Phase II: Psychometric Properties Evaluation

Participants. A total of 607 students participated from the University of Alicante, Spain. The sample represented diverse academic specializations.

Instruments. The BPFQ, comprising four dimensions, assessed students' beliefs about peer assessment. A sociodemographic questionnaire gathered additional information.

Procedure

Faculty members from various programs at the University of Alicante were contacted and briefed about the research. They selected participating classes and explained the study to students. Participants received information about the study's purpose, and data confidentiality. The assessment, including informed consent, a sociodemographic questionnaire, and the Spanish BPFQ, was administered. Participation was voluntary, anonymous, and uncompensated. Ethics approval was obtained from the University of Alicante Ethics Committee (UA-2022-02-16), adhering to the World Medical Association's (2013) Declaration of Helsinki.

Data Analysis

Descriptive analyses were firstly carried out to examine the performance of the test. Pearson's bivariate correlations were performed among confirmatory factor analysis verified the questionnaire's factorial structure. Internal consistency was assessed using Cronbach's alpha and McDonald's omega. Differences between fields of study were analyzed using the Kruskal-Wallis test.

Results

Descriptive Item Analysis

Descriptive analyses calculated means, standard deviations, ranges, and percentiles for each BPFQ item. A 5-point Likert scale was used, with a ceiling effect observed in three of the ten items (Table 1).

Correlation Analysis among Items, Subscales, and the General Factor of the BPFQ

Stronger correlations were found among the items which belonged to the same dimension. Positive, moderate, and significant correlations were also found among dimensions of the questionnaire. Moreover, the general factor of the BPFQ was strongly, positively, and significantly correlated to all the dimensions (Table 2).

Confirmatory Factor Analysis

Three confirmatory factor analyses were performed. The bifactor model showed the best fit indicators and it verified a structure composed of both the general factor and the subscales of the Spanish version of the BPFQ (Ruiz-Robledillo et al., 2022), and the original questionnaire (Huisman et al., 2020). These subscales were: (1) valuing peer feedback as an instructional method (VIM),

(2) confidence in one's own quality of peer feedback (CO), (3) confidence in the quality of received peer feedback (CR), and (4) valuing peer feedback as an essential skill (VPS). Data indicated the best satisfactory fit for the bifactor model, $\chi^2(15) = 25.129$, CFI = .997, TLI = .990, RMSEA = .033, SRMR = .012, 95% CI [.003, .055] (Figure 1). For theoretical and methodological reasons, all test items were retained.

Reliability

Item-total correlations as indicators of discrimination were above .30 for all items except item 3 ($r_{itc} = .293$) (Table 3). The coefficient of internal consistency reliability for the overall factor exceeded .80 ($\alpha = .862$, 95% CI [.845, .878], and $\omega = .857$, 95% CI [.840, .874]). Additionally, only the general reliability index would slightly increase upon removing item 3 ("Feedback should only be provided by teaching staff"). Regarding the hierarchical omega values, the general factor questionnaire scores explained more than the 80% of the construct without the influence of the subscales, and vice versa (Table 4).

Differences in BPFQ by Field of Study

Results of non-parametric tests assessing differences between fields of study are presented in Table 5. Statistically significant differences were found in all dimensions of the Spanish-adapted questionnaire across at least two student field groups. Specifically, between Science and Social Sciences and Law, significant differences were observed in all dimensions, with higher scores in the Science field (Tables 5 and 6). Lastly, significant differences were also found in all questionnaire dimensions between Education and Social Sciences and Law, with higher scores in all dimensions in the Education field (Table 5).

Discussion

This paper discusses the shortage of peer assessment in university contexts despite its recognized benefits, attributing this partly to the lack of appropriate assessment tools. To address this gap, the study aimed to adapt and validate the Beliefs about Peer-Feedback Questionnaire (BPFQ) for assessing student beliefs towards peer assessment. Results showed that the Spanish version of the BPFQ replicated the original structure and demonstrated good psychometric properties, including factorial structure, internal consistency, and reliability. However, the best fit is for a bifactor model, in which both the general factor and the subscales are included. Significant differences in beliefs about peer assessment were found among students from different fields, with those in education valuing it more positively. This is in line with previous research highlighting the importance of education in peer assessment. The study acknowledges strengths such as its diverse sample but also recognizes limitations like a small sample size and students' limited experience with peer assessment, suggesting the need for cautious interpretation. Despite limitations, the BPFQ is deemed a reliable instrument for evaluating beliefs about peer assessment, particularly among Education students. This supports its utility in enhancing collaborative learning strategies.

The study concludes by emphasizing the positive impact of assessing student perceptions of peer assessment on teaching and learning process, highlighting the importance of understanding evaluative judgments and improving student interaction skills. Future research directions include exploring peer assessment methodologies further through focus groups to refine understanding and improve the reliability and validity of research findings. Overall, the study contributes to the advancement of peer

assessment practices in higher education, facilitating knowledge enhancement and promoting more effective teaching and learning environments.

Conflicto de Intereses

Los autores de este artículo declaran que no tienen ningún conflicto de intereses.

Referencias

- Alemdag, E. y Yildirim, Z. (2022). Effectiveness of online regulation scaffolds on peer feedback provision and uptake: A mixed methods study. *Computers & Education*, 188, Article 104574. <https://doi.org/10.1016/j.compedu.2022.104574>
- Alqassab, M., Srijbos, J.-W., Panadero, E., Ruiz, J. F., Warrens, M. y To, J. (2023). A Systematic review of peer assessment design elements. *Educational Psychology Review*, 35(1), Article 18. <https://doi.org/10.1007/s10648-023-09723-7>
- Ashenafi, M. M. (2017). Peer-assessment in higher education - twenty-first century practices, challenges and the way forward. *Assessment & Evaluation in Higher Education*, 42(2), 226-251. <https://doi.org/10.1080/02602938.2015.1100711>
- Barkley, E. F., Cross, K. P. y Major, C. H. (2014). *Collaborative learning techniques: A handbook for college faculty*. John Wiley & Sons.
- Boud, D. (1990). Assessment and the promotion of academic values. *Studies in Higher Education*, 15(1), 101-111. <https://doi.org/10.1080/030750790.12331377621>
- Boud, D., Cohen, R. y Sampson, J. (1999). Peer learning and assessment. *Assessment & Evaluation in Higher Education*, 24(4), 413-426. <https://doi.org/10.1080/0260293990240405>
- Browne, M. W. y Cudeck, R. (1992). Alternative ways of assessing model fit. *Sociological Methods & Research*, 21(2), 230-258. <https://doi.org/10.1177/0049124192021002005>
- Burgess, A. y Mellis, C. (2015). Feedback and assessment for clinical placements: Achieving the right balance. *Advances in Medical Education and Practice*, 6, 373-381. <https://doi.org/10.2147/AMEP.S77890>
- Carless, D. (2022). From teacher transmission of information to student feedback literacy: Activating the learner role in feedback processes. *Active Learning in Higher Education*, 23(2), 143-153. <https://doi.org/10.1177/1469787420945845>
- Carlsson Hauff, J. y Nilsson, J. (2022). Students' experience of making and receiving peer assessment: The effect of self-assessed knowledge and trust. *Assessment & Evaluation in Higher Education*, 47(6), 959-971. <https://doi.org/10.1080/02602938.2021.1970713>
- Carretero-Dios, H. y Pérez, C. (2007). Standards for the development and review of instrumental studies: Considerations about test selection in psychological research. *International Journal of Clinical and Health Psychology*, 7(3), 863-882.
- Cheng, W. y Warren, M. (1997). Having second thoughts: Student perceptions before and after a peer assessment exercise. *Studies in Higher Education*, 22(2), 233-239. <https://doi.org/10.1080/03075079712331381064>
- Darvishi, A., Khosravi, H., Sadiq, S. y Gašević, D. (2022). Incorporating AI and learning analytics to build trustworthy peer assessment systems. *British Journal of Educational Technology*, 53(4), 844-875. <https://doi.org/10.1111/bjet.13233>
- Davis, S., Duane, B., Loxley, A. y Quigley, D. (2022). The evaluation of an evidence-based model of feedback implemented on an undergraduate dental clinical learning environment. *BMC Medical Education*, 22(1), Article 588. <https://doi.org/10.1186/s12909-022-03630-1>
- Deneen, C. C. y Hoo, H.-T. (2021). Connecting teacher and student assessment literacy with self-evaluation and peer feedback. *Assessment & Evaluation in Higher Education*, 0(0), 1-13. <https://doi.org/10.1080/02602938.2021.1967284>
- Dochy, F., Segers, M. y Sluijsmans, D. (1999). The use of self-, peer and co-assessment in higher education: A review. *Studies in Higher Education*, 24(3), 331-350. <https://doi.org/10.1080/03075079912331379935>
- Domínguez-Lara, S. y Rodríguez, A. (2017). Statistical indices from bifactor models. *Interacciones*, 3(2), 59-65. <https://doi.org/10.24016/2017.v3n2.51>
- Epskamp, S., Stuber, S., Nak, J., Veenman, M. y Jorgensen, T. D. (2019). *semPlot: Path diagrams and visual analysis of various SEM packages' output (Version 1.1.2)*. [Computer software].
- Erkan, E., Dimitriadis, Y. y Gašević, E. (2021). A collaborative learning approach to dialogic peer feedback: A theoretical framework. *Assessment & Evaluation in Higher Education*, 46(4), 586-600. <https://doi.org/10.1080/02602938.2020.1786497>
- Estriegana, R., Teixeira, A. M., Robina-Ramirez, R., Medina-Merodio, J.-A. y Otón, S. (2024). Impact of communication and relationships on student satisfaction and acceptance of self- and peer-assessment. *Education and Information Technologies*, 29, 14715-14731. <https://doi.org/10.1007/s10639-023-12276-5>

- Fang, J.-W., Chang, S.-C., Hwang, G.-J. y Yang, G. (2021). An online collaborative peer-assessment approach to strengthening pre-service teachers' digital content development competence and higher-order thinking tendency. *Educational Technology Research and Development*, 69(2), 1155-1181. <https://doi.org/10.1007/s11423-021-09990-7>
- Field, A. (2009). *Discovering statistics using SPSS*. SAGE Publications.
- Filius, R. M., de Kleijn, R. A., Uijl, S. G., Prins, F. J., van Rijen, H. V. y Grobbee, D. E. (2018). Strengthening dialogic peer feedback aiming for deep learning in SPOCs. *Computers & Education*, 125, 86-100. <https://doi.org/10.1016/j.compedu.2018.06.004>
- García-Celay, I. M. y León, O. G. (2005). Sistema de clasificación del método en los informes de investigación en Psicología. *International Journal of Clinical and Health Psychology*, 5(1), 115-127.
- Gaynor, J. W. (2020). Peer review in the classroom: Student perceptions, peer feedback quality and the role of assessment. *Assessment & Evaluation in Higher Education*, 45(5), 758-775. <https://doi.org/10.1080/02602938.2019.1697424>
- Green, S. B. y Yang, Y. (2009). Commentary on coefficient alpha: A cautionary tale. *Psychometrika*, 74(1), 121-135. <https://doi.org/10.1007/s11336-008-9098-4>
- Gregoire, J. (2018). ITC guidelines for translating and adapting tests (2 ed.). *International Journal of Testing*, 18(2), 101-134. <https://doi.org/10.1080/15305058.2017.1398166>
- Gulledge, C. M., Smith, D. G., Ziedas, A., Muh, S. J., Moutzourous, V. y Makhni, E. C. (2019). Floor and ceiling effects, time to completion, and question burden of PROMIS CAT domains among shoulder and knee patients undergoing nonoperative and operative treatment. *JBJS Open Access*, 4(4), Article e0015. <https://doi.org/10.2106/JBJS.OA.19.00015>
- Han, Y. y Xu, Y. (2020). The development of student feedback literacy: The influences of teacher feedback on peer feedback. *Assessment & Evaluation in Higher Education*, 45(5), 680-696. <https://doi.org/10.1080/02602938.2019.1689545>
- Hernández, A., Hidalgo, M. D., Hambleton, R. K. y Gómez-Benito, J. (2020). International Test Commission guidelines for test adaptation: A criterion checklist. *Psicothema*, 32(3), 390-398. <https://doi.org/10.7334/psicothema2019.306>
- Hernández-Lalinde, J. D., Espinosa Castro, J. F., Peñaloza Tarazona, M. E., Fernández González, J. E., Chacón Rangel, J. G., Toloza Sierra, C. A., Arenas Torrado, M. K., Carrillo Sierra, S. M. y Bermúdez Pirela, V. J. (2018). Sobre el uso adecuado del coeficiente de correlación de Pearson: definición, propiedades y suposiciones. *AVFP Archivos Venezolanos de Farmacología y Terapéutica*, 37(5). http://saber.ucv.ve/ojs/index.php/rev_aavft/article/view/16160
- Hoo, H.-T., Deneen, C. y Boud, D. (2022). Developing student feedback literacy through self and peer assessment interventions. *Assessment & Evaluation in Higher Education*, 47(3), 444-457. <https://doi.org/10.1080/02602938.2021.1925871>
- Huisman, B., Saab, N., van Driel, J. y van den Broek, P. (2018). Peer feedback on academic writing: Undergraduate students' peer feedback role, peer feedback perceptions and essay performance. *Assessment & Evaluation in Higher Education*, 43(6), 955-968. <https://doi.org/10.1080/02602938.2018.1424318>
- Huisman, B., Saab, N., van Driel, J. y van den Broek, P. (2020). A questionnaire to assess students' beliefs about peer-feedback. *Innovations in Education and Teaching International*, 57(3), 328-338. <https://doi.org/10.1080/014703297.2019.1630294>
- Hunzer, K. M. (2014). *Collaborative learning and writing: Essays on using small groups in teaching English and composition*. McFarland.
- Hwang, G.-J., Zou, D. y Wu, Y.-X. (2023). Learning by storytelling and critiquing: A peer assessment-enhanced digital storytelling approach to promoting young students' information literacy, self-efficacy, and critical thinking awareness. *Educational Technology Research and Development*, 71(3), 1079-1103. <https://doi.org/10.1007/s11423-022-10184-y>
- Ibarra-Sáiz, M. S., Rodríguez-Gómez, G. y Boud, D. (2020). Developing student competence through peer assessment: The role of feedback, self-regulation and evaluative judgement. *Higher Education*, 80(1), 137-156. <https://doi.org/10.1007/s10734-019-00469-2>
- Kasch, J., van Rosmalen, P., Löhr, A., Klemke, R., Antonaci, A. y Kalz, M. (2021). Students' perceptions of the peer-feedback experience in MOOCs. *Distance Education*, 42(1), 145-163. <https://doi.org/10.1080/01587919.2020.1869522>
- Kaufman, J. H. y Schunn, C. D. (2011). Students' perceptions about peer assessment for writing: Their origin and impact on revision work. *Instructional Science*, 39(3), 387-406. <https://doi.org/10.1007/s11251-010-9133-6>
- Kline, R. B. (2015). *Principles and practice of structural equation modeling* (4th ed.). Guilford Publications.
- Krause, K.-L., Hartley, R., James, R. y McInnis, C. (2005). *The first year experience in Australian universities: Findings from a decade of national studies*.
- Leary, M. R. (2008). *Introduction to behavioral research methods* (5th ed.). Pearson.
- Lerchenfeldt, S. y Taylor, T. A. H. (2020). Best practices in peer assessment: Training tomorrow's physicians to obtain and provide quality feedback. *Advances in Medical Education and Practice*, 11, 571-578. <https://doi.org/10.2147/AMEP.S250761>
- McDonald, R. P. (1999). *Test theory: A unified treatment*. Psychology Press. <https://doi.org/10.4324/9781410601087>
- Nicol, D., Thomson, A. y Breslin, C. (2014). Rethinking feedback practices in higher education: A peer review perspective. *Assessment & Evaluation in Higher Education*, 39(1), 102-122. <https://doi.org/10.1080/02602938.2013.795518>
- Panadero, E. y Alqassab, M. (2019). An empirical review of anonymity effects in peer assessment, peer feedback, peer review, peer evaluation and peer grading. *Assessment & Evaluation in Higher Education*, 44(8), 1253-1278. <https://doi.org/10.1080/02602938.2019.1600186>
- Pinheiro, A., Rolim, V., Gasevic, D. y Ferreira, R. (2023). A comparative analysis between good feedback descriptors on online courses. *Símpoio Brasileiro de Informática na Educação*, 34, 1512-1523. <https://doi.org/10.5753/sbie.2023.235316>
- Ramos, Z. (2018). *Psicometría básica*. Fundación Universitaria del Área Andina. Bogotá.
- Reise, S. P., Scheines, R., Widaman, K. F. y Haviland, M. G. (2013). Multi-dimensionality and structural coefficient bias in structural equation modeling: A bifactor perspective. *Educational and Psychological Measurement*, 73(1), 5-26. <https://doi.org/10.1177/0013164412449831>
- Rodríguez, A., Reise, S. P. y Haviland, M. G. (2016). Evaluating bifactor models: Calculating and interpreting statistical indices. *Psychological Methods*, 21(2), 137-150. <https://doi.org/10.1037/met0000045>
- Rosseel, Y. (2012). Lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, 48(2), 1-36. <https://doi.org/10.18637/jss.v048.i02>
- Ruiz-Robledillo, N., Albaladejo-Blázquez, N., Clement-Carbonell, V., Vela-Bermejo, J., Rubio-Aparicio, M., Portilla-Tamarit, I., Costa-López, B. y Ferrer-Cascales, R. (2022). Exploratory factor analysis of the Spanish version of the Beliefs About Peer-Feedback Questionnaire. *EDULEARN22 Proceedings*, 8787-8793. IATED.
- Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science*, 18(2), 119-144. <https://doi.org/10.1007/BF00117714>
- Satorra, A., y Bentler, E. M. (1994). Corrections to test statistics and standard errors in covariance structure analysis. En A. von Eye y C. C. Clogg (Eds.), *Latent variables analysis: Applications for developmental research* (pp. 399-419). Sage.
- Schellekens, L. H., Bok, H. G. J., de Jong, L. H., van der Schaaf, M. F., Kremer, W. D. J. y van der Vleuten, C. P. M. (2021). A scoping review on the notions of Assessment as Learning (AaL), Assessment for learning (Afl), and Assessment of learning (Aol). *Studies in Educational Evaluation*, 71, Article 101094. <https://doi.org/10.1016/j.stueduc.2021.101094>
- Shui Ng, W. y Yu, G. (2021). Students' attitude to peer assessment process: A critical factor for success. *Interactive Learning Environments*, 0(0), 1-19. <https://doi.org/10.1080/10494820.2021.1916762>
- Somervell, H. (1993). Issues in assessment, enterprise and higher education: The case for self-peer and collaborative assessment. *Assessment & Evaluation in Higher Education*, 18(3), 221-233. <https://doi.org/10.1080/0260293930180306>
- Stucky, B. D., y Edelen, M. O. (2015). Using hierarchical IRT models to create unidimensional measures from multidimensional data. En S. P. Reise y D. A. Revicki (Eds.), *Handbook of item response theory modeling: Applications to typical performance assessment* (pp. 183-206). Routledge.
- Stucky, B. D., Thissen, D., y Edelen, M. O. (2013). Using logistic approximations of marginal trace lines to develop short assessments. *Applied Psychological Measurement*, 37(1), 41-57. <https://doi.org/10.1177/0146621612462759>
- Su, W. y Huang, A. (2022). More enjoyable to give or to receive? Exploring students' emotional status in their peer feedback of academic writing. *Assessment & Evaluation in Higher Education*, 47(7), 1005-1015. Scopus. <https://doi.org/10.1080/02602938.2021.2004389>
- Taber, K. S. (2018). The use of Cronbach's alpha when developing and reporting research instruments in science education. *Research in Science Education*, 48, 1273-1296. <https://doi.org/10.1007/s11165-016-9602-2>
- The R Project for Statistical Computing. (2021). <https://www.r-project.org/index.html>
- Topping, K. J. (1998). Peer assessment between students in colleges and universities. *Review of Educational Research*, 68(3), 249-276. <https://doi.org/10.2307/1170598>
- Tornwall, J., McGaughy, M. y Schubert, C. (2022). Peer review: Factors that motivate students to provide supportive peer feedback. *Nurse Educator*, 47(2), 114-119. <https://doi.org/10.1097/NNE.0000000000001043>
- van Blankenstein, F. M., Dirx, K. J. y de Bruycker, N. M. (2024). Ask your peer! How requests for peer feedback affect peer feedback responses. *Educational Research and Evaluation*, 30(1-2), 36-57. <https://doi.org/10.1080/13803611.2024.2376832>
- van Popta, E., Kral, M., Camp, G., Martens, R. L. y Simons, P. R.-J. (2017). Exploring the value of peer feedback in online learning for the provider. *Educational Research Review*, 20, 24-34. <https://doi.org/10.1016/j.edurev.2016.10.003>
- Vygotsky, L. S. (2012). *Thought and Language, revised and expanded edition*. MIT Press.
- Wood, J. (2021). A dialogic technology-mediated model of feedback uptake and literacy. *Assessment & Evaluation in Higher Education*, 46(8), 1173-1190. <https://doi.org/10.1080/02602938.2020.1852174>

- World Medical Association. (2013). World Medical Association Declaration of Helsinki: Ethical principles for medical research involving human subjects. *JAMA*, *310*(20), 2191 -2194. <https://doi.org/10.1001/jama.2013.281053>
- Yu, S. y Liu, C. (2021). Improving student feedback literacy in academic writing: An evidence-based framework. *Assessing Writing*, *48*, Article 100525. <https://doi.org/10.1016/j.asw.2021.100525>
- Zhan, Y. (2021). Are they ready? An investigation of university students' difficulties in peer assessment from dual perspectives. *Teaching in Higher Education*, *0*(0), 1-18. <https://doi.org/10.1080/13562517.2021.2021393>
- Zong, Z., Schunn, C. D. y Wang, Y. (2023). When do students provide more peer feedback? The roles of performance and prior feedback experiences. *Instructional Science*, *51*(6), 977-1003. <https://doi.org/10.1007/s11251-023-09640-w>