



Procedimiento de fiabilización de la herramienta de Observación por Pares en la Educación Superior (He-PoT)

Peer Observation Tool Reliability Procedure on Higher Education (HE-PoT)

Autores

José Antonio Domínguez-Montes¹
Javier Gálvez-González¹
Fernando Manuel Otero-Saborido¹
Roos Van Gasse²

¹ Universidad Pablo de Olavide
(España)

² KU Leuven (Bélgica)

Autor de correspondencia:
José Antonio Domínguez-Montes
jadommon@alumno.upo.es

Recibido: 23-03-26
Aceptado: 30-05-26

Como citar en APA

Domínguez-Montes, J. A., Gálvez-González, J., Van Gasse, R., & Otero-Saborido, F. M. (2026). Procedimiento de fiabilización de la herramienta de Observación por Pares en la Educación Superior (HE-PoT). *Retos*, 82, 715-725. <https://doi.org/10.47197/retos.v82.119089>

Resumen

Introducción: La observación por pares de la enseñanza (POT) se ha vuelto cada vez más popular en las instituciones de Educación Superior de todo el mundo. No obstante, es escaso del número instrumento que hayan sido sometidos a procesos de validación y fiabilización.

Objetivo: Analizar la fiabilidad de la herramienta POT-Higher Education (HE-PoT) para su posterior aplicación y mejora de la Enseñanza Universitaria desde un enfoque colaborativo.

Metodología: Un diseño observacional nomotético, puntual y multidimensional fue utilizado para analizar 50 sesiones docentes mediante la herramienta validada HE-PoT y el software Lince Plus. Por un parte, 3 docentes universitarios fueron observadores. Por otra, 4 docentes de la Facultad de Ciencias del Deporte fueron observados. La fiabilidad se garantizó mediante un entrenamiento de expertos, evaluada con el coeficiente Kappa de Cohen (intra e inter-observador) y un análisis de generalizabilidad de las facetas observadores, ítems y sesiones.

Resultados: La herramienta mostró una fiabilidad inter-observador excelente (promedio = 0.867) y una alta generalizabilidad para la faceta ítems ($e^2=0.876$). Los resultados destacan que la mayor varianza (61,2%) reside en la interacción ítem-sesión, confirmando la estabilidad de los observadores frente a la heterogeneidad de las conductas pedagógicas.

Conclusiones: Los niveles de concordancia y generalizabilidad muestran que la herramienta HE-PoT es un instrumento fiable para la observación por pares entre el profesorado universitario.

Palabras clave

Análisis de fiabilidad; educación superior; herramientas de observación; metodología observacional; observación por pares.

Abstract

Introduction: Peer Observation of Teaching (POT) has gained increasing prominence in Higher Education institutions worldwide. However, there is a paucity of instruments that have undergone rigorous validation and reliability testing.

Objective: To analyze the reliability of the POT-Higher Education (HE-PoT) tool for its subsequent application in enhancing university teaching through a collaborative approach.

Methodology: A nomothetic, point, and multidimensional observational design was employed to analyze 50 teaching sessions using the validated HE-PoT tool and Lince Plus software. The study involved three university instructors acting as observers and four faculty members from the Faculty of Sport Sciences being observed. Reliability was ensured through expert training and assessed using Cohen's Kappa coefficient (intra- and inter-observer), alongside a generalizability Theory analysis of the observers, items, and sessions facets.

Results: The tool demonstrated excellent inter-observer reliability (mean=0.867) and high generalizability for the items facet ($e^2=0.876$). Findings highlight that the greatest variance (61.2%) resides in the item-session interaction, confirming observer stability despite the inherent heterogeneity of pedagogical behaviors.

Conclusions: The levels of reliability and generalizability indicate that the HE-PoT tool is a reliable instrument for peer observation among university faculty.

Keywords

Keywords: Reliability analysis; higher education; observation tools; observational methodology; peer observation.

Introducción

El sistema universitario español, regulado por la Ley Orgánica 2/2023, ha ido reforzando e intensificando su integración bajo el contexto del Espacio Europeo de Educación Superior. La configuración del llamado “Plan Bolonia” orientó a las universidades a que sus modelos formativos cooperaran y convergieran en una estructura común de Educación Superior (en adelante, ES) (Aránguiz et al., 2021; Mataranz, 2020). Ibarra-Sáiz et al. (2020), consideran la ES como el motor de cambio y transformación hacia una sociedad más justa, responsable y equitativa para todos. Partiendo de la concepción de que el centro de la educación lo debe constituir el aprendizaje, el éxito de la ES no debería medirse por la calidad de la enseñanza ofrecida por los centros, departamentos, cursos o profesorado, sino por la calidad del aprendizaje del estudiantado. Por esa razón, poner el foco en el estudiantado debería ser uno de los rasgos que significara a la ES en Europa (Flores-Ferro et al., 2025).

En España la ES ha ido evolucionado con un progresivo crecimiento y con la transformación e inclusión de importantes procesos como es el centrado en la evaluación del profesorado (Aznar-Díaz et al., 2020), donde conviene seguir implementando mejoras en estrategias de evaluación con el objeto de afrontar los retos actuales y promover una formación que incorpore más espacios de evaluación dialógicos (Gallardo-Fuentes et al., 2025).

De ahí que, la observación de los docentes en el aula, como parte del proceso de evaluación del profesorado, pueda incorporarse como un mecanismo para el desarrollo de la práctica docente (Fletcher, 2018) y en el caso de la observación entre iguales, dado su valor formativo (Yiend et al., 2014), es considerada como una buena práctica para la mejora de proceso de enseñanza y aprendizaje. Además, instituciones internacionales como la Organización para la Cooperación y el Desarrollo Económicos (OCDE, 2019), señalan la importancia de las prácticas de aprendizaje entre pares como un mecanismo efectivo para el desarrollo profesional del docente.

En este sentido, la observación por pares de la enseñanza (en adelante, POT, por sus siglas en inglés), se ha vuelto cada vez más popular en instituciones de ES de todo el mundo como EE.UU., Australia y el Reino Unido (Carragher & McGaughey, 2016; Johnston et al., 2022). La POT, ha sido definida recientemente como “la evaluación entre pares de la docencia consiste en un proceso colaborativo y deliberado de observación de la enseñanza con el fin de mejorar las prácticas docentes” (Georgiou et al., 2018).

Cabe destacar que existe un sector del profesorado y de los departamentos que suelen evitar la POT, principalmente por actitudes resistentes del profesorado (Richards & Farrell, 2005; Bailey, 2006). A este respecto, Corcelles-Seuba et al. (2025), sugieren que los cambios en la calidad de la enseñanza que puede aportar la POT, dependen de que se reduzca la resistencia del profesorado a utilizarla.

Sin embargo, la POT aporta numerosos beneficios documentados (Chism, 2007; Sachs & Parsell, 2014). Hallamos numerosas investigaciones que expresan los beneficios de la POT, tales como la mejora de la moral, la motivación y el aumento de la colaboración entre los miembros del profesorado (Brancato, 2003; Huston & Weaver, 2008; Menges, 1987; Skinner & Welch, 1996). Otros de los beneficios de la POT son: promover y/o difundir prácticas docentes de calidad (Chao et al., 2006; Little, 2009; McGahan et al., 2015; Ridge & Lavigne, 2020; Yiend et al., 2014), mejorar el desarrollo profesional (Bell, 2001; Fletcher, 2018; Hammersley-Fletcher & Orsmond, 2005; Thomas et al., 2014) y fomentar el crecimiento de las relaciones profesionales a lo largo del tiempo (Shortland, 2010).

En la literatura científica existen diferentes instrumentos POT utilizados en la docencia universitaria enfocados tanto para la enseñanza presencial y como para la enseñanza en línea. Algunos de esas herramientas son el RTOP de Amrein-Beardsley & Popp (2012); el Instrument for observation of 'Project Based Learning' facilitation de Garcia et al. (2017); el Peer Evaluation Tool de Rabada et al. (1986); el Peer Observation and Evaluation Tool (POET) diseñado inicialmente por Trujillo et al. (2008) y validado posteriormente por Crabtree et al. (2016) y desarrollado Online (POET-O) por (Jia et al., 2025); la Escala de evaluación del proceso enseñanza-aprendizaje de la técnica deportiva basado en la enseñanza por pares en la Educación Superior (ETEPES) de Álvarez-Medina et al. (2020); the Practical Observation Rubric to Assess Active Learning (PORTAAL) de Eddy et al. (2015); the Classroom Observation Protocol for Undergraduate STEM de Smith et al. (2013), entre otros.



La herramienta de Observación por Pares en la Enseñanza Universitaria (en adelante, HE-PoT) de Domínguez-Montes et al. (2024), fue creada y diseñada previamente ad hoc siguiendo principios de la metodología observacional y se alinea con la categorización teórica que Peel (2005), define como segunda dimensión (D2). Esta dimensión relaciona al profesorado de un modo colaborativo y basado en procesos, y dentro de las diferentes formas que puede adoptar la POT, reconocidas como modelos de observación, se enmarca en el modelo colaborativo (Gosling, 2002; Yiend et al., 2014), o que Byrne (2010) clasifica como “modelo de desarrollo entre pares”.

El diseño HE-PoT fue concebido para observar las clases presenciales de cualquier disciplina en la ES. Aspectos formales de la sesión, habilidades comunicativas y contenidos pedagógicos son las 3 dimensiones en las que se organizan los 11 ítems de la herramienta que incluye campos para que el observador añada comentarios cualitativos destinados a un futuro feedback. En relación a la importancia de la retroalimentación, Corcelles-Seuba et al. (2023), han demostrado que el mayor desafío en la POT está relacionado con el feedback que intercambian los docentes. Además, en la investigación llevada a cabo por De la Iglesia et al. (2025), concluyen que, a más intercambio dialógico, más feedback de calidad es compartido, llevándose así la implementación de acciones asociadas a la construcción conjunta de conocimiento y aumentando las características vinculadas a la retroacción que conduce a introducir mejoras en la práctica docente. El feedback, frente a los autoinformes del profesorado, proporciona información valiosa sobre cómo los profesores construyen conocimiento de forma colaborativa Flores et al. (2025).

La revisión realizada por Otero-Saborido et al. (2024), muestra una escasez de herramientas validadas en la enseñanza universitaria. Así mismo, algunas de las herramientas validadas tienen un número excesivo de ítems. Esa hipertrofia de información facilita el instrumento para fines credencialistas, pero dificulta el uso de esta herramienta como instrumento de mejora en un contexto constructivo e igualitario entre profesorado. Por otra parte, la validación de contenido de una herramienta de observación como la validada por Domínguez-Montes et al. (2024), es un paso previo y necesario para la construcción de un instrumento de observación, pero es incompleto. Son necesarios procesos posteriores de fiabilización a través de la comprobación de la estabilidad de la medida del instrumento y la capacidad de generalización del mismo.

Por tanto y para poder consolidar la potencialidad de la HE-PoT, el objetivo del presente estudio fue analizar la fiabilidad de la herramienta POT-Higher Education (HE-PoT) para su futurible aplicación y mejora de la Enseñanza Universitaria desde un enfoque colaborativo.

Método

Diseño

La metodología observacional fue utilizada en la presente investigación. Se partió de la herramienta diseñada por Domínguez-Montes et al. (2024). La herramienta ad hoc tenía las siguientes características dentro de los diseños observacionales: nomotético, puntual y multidimensional, correspondiendo al cuadrante III de este tipo de diseños (Anguera et al., 2011); a) nomotético, porque existen diferentes docentes que observan y son observados en el marco del proceso de enseñanza-aprendizaje de sus respectivos grupos-clase; b) puntual, puesto que la observación por pares se utiliza para obtener una visión de un momento concreto y c) multidimensional, ya que se analizaron un amplio espectro de conductas relacionadas con las observaciones por pares en la Enseñanza Universitaria (dimensiones e ítems y categorías).

Participantes

Cuatro asignaturas del Grado de Ciencias de la Actividad Física y del Deporte fueron observadas durante los cursos 2024-2025 y 2025-2026 resultando un total de 50 sesiones analizadas. Dentro de los participantes diferenciamos entre dos categorías: observadores expertos y voluntarios observados. En primer lugar, 3 observadores expertos participaron en los procedimientos de observación. Los observadores recibieron previamente 3 sesiones de formación sobre la herramienta (HE-PoT). Los criterios de inclusión del profesorado para formar parte del panel de observadores fueron los siguientes: 1) ser doctor en Ciencias de la Actividad Física y el Deporte (en adelante CCAFD) o cursar un programa de doctorado

en este ámbito; 2) poseer un título de graduado o licenciado en CCAFD o equivalentes; 3) cumplir como mínimo cinco años de docencia en la Enseñanza Universitaria) en el ámbito y/o docente en otras etapas educativas con la especialidad en cuestión (E. Secundaria o E. Primaria); 4) haber publicado libros y/o artículos relacionados con la temática propuesta u otras afines. Al menos se tuvieron que cumplir tres de los cuatro criterios.

Por otra parte, 4 docentes voluntarios fueron observados por los expertos sin que recibieran ninguna información sobre la herramienta. El profesorado voluntario que se ofreció ser observado, simplemente, informó a los investigadores de sus horarios de docencia. Posteriormente, sin aviso previo, fueron observados por los expertos.

Instrumento

Se utilizó la POT-Higher Education (HE-PoT) validada por Domínguez-Montes et al. (2024). Este instrumento incluye 11 ítems que se gradúan en 3 dimensiones secuenciadas en función mayor o menor presencia de la conducta descrita en el ítem. Las tres categorías son: 'Excelente' (E), 'En proceso' (EP) e 'Inadecuado' (I). Cada ítem tiene su descripción cualitativa en las tres categorías. La herramienta incluye además las instrucciones de uso, la hoja de registro y en el Apéndice 1, de material complementario: resolución y aclaración de posibles dudas en relación al contenido de los ítems.

Procedimiento

Dos pasos se siguieron en la investigación

a) Formación de observadores

La formación tuvo como fin mostrar las fases y objetivos de la investigación y entrenar a los observadores en la herramienta de observación HE-PoT. El panel de expertos POT (observadores y observados) fueron entrenados *ad hoc* para realizar las observaciones garantizando así la estabilidad y fiabilidad intra e inter observadores.

Este proceso de entrenamiento-formación consistió en una serie de fases utilizadas en investigaciones previas (Escartí et al., 2013): 1) Familiarización con la herramienta (explicación de las dimensiones, criterios, resolución de dudas, etc.); 2) Visualización de sesiones (se visualizó una clase de forma individual); 3) Puesta en común de los resultados obtenidos por los 4) Reformulación de la redacción y aclaración de ítems de herramienta.

b) Observaciones

Esta fase se realizó mediante observación directa no participativa (Portell et al., 2015), analizando cada observador de forma independiente tanto sesiones en el aula como grabaciones. Se visualizaron un total de 50 clases (30 sesiones grabadas y 20 presenciales) distribuidas en los 4 docentes observados. 24 sesiones fueron destinadas al análisis intra-observadores y 26 al análisis inter-observador. Para la elección de las clases a observar se tuvieron en cuenta las posibilidades horarias de observadores (Stone & Shiffman, 2002). Los observadores se presentaban en el aula 15 minutos antes en el caso de las observaciones presenciales. Durante la observación codificaban cada uno de los 11 ítems rodeando la categoría descriptiva más adecuada. De forma opcional, podrían cumplimentar el apartado destinado a comentario.

Materiales

Para la visualización de las sesiones grabadas, los expertos se apoyaron en el Software de investigación LINCE PLUS Observation tool v2.1 para estudios observacionales sistemáticos (Gabin et al., 2012; Soto et al., 2019; Soto et al., 2022), el cual simplifica en gran medida el registro de los datos de observación. Una vez recogidos los datos en la HE-PoT, se volcaron en una plantilla diseñada en el Software informático de hojas de cálculo Microsoft Excel, y posteriormente del Excel, se volcó al Software estadístico IBM SPSS 29.0., para su posterior análisis estadístico. El análisis de la generalizabilidad (TG) se realizó con el programa SAGT v1.0 (Hernández-Mendo et al., 2016).

Análisis estadístico

Para analizar la fiabilidad de la herramienta HE-PoT, se exploró el grado de concordancia desde dos puntos de vista. En primer lugar, el punto de vista de un mismo observador con dos mediciones distanciadas en 15 días (intraobservador) y, en segundo lugar, la concordancia entre dos observadores inde-

pendientes. Para ambas, se utilizó el coeficiente Kappa de Cohen (K), teniendo en cuenta así la indexación de la concordancia de respuestas entre dos observaciones mediante el estadístico desarrollado por Cohen (1960). Así mismo, implementó un análisis de generalizabilidad (Cronbach, 1972), utilizando las facetas, observadores [OBS], ítems [ITE] y sesiones [SES].

Resultados

La herramienta a fiabilizar estaba configurada por 11 ítems organizados en las dimensiones temáticas 'aspectos formales', 'comunicativa' y 'pedagógica' (Tabla 1). El análisis de la fiabilidad a partir de la Kappa de Cohen fue realizado desde el punto de vista intraobservador e interobservador. Los resultados obtenidos en el primer caso reflejan una fiabilidad perfecta en las dos medidas del observador para los 11 ítems (Tabla 1). En el caso de la perspectiva del interobservador, 9 de los 11 ítems obtuvieron una concordancia casi perfecta (entre 0.81-1.00) (Landis & Koch, 1977) y los restantes resultaron con una fiabilidad considerable (entre 0.61-0.80) (Landis & Koch, 1977). El promedio de concordancia en ambos puntos de vista fue cercano a 1.

Tabla 1. Fiabilidad intra e inter observadores según el coeficiente Kappa de Cohen (K)

Dimensión	Ítem	Kintra	Kinter
Aspectos formales	1. El docente es puntual a la entrada y a la salida	1	1
	2. El volumen de voz es suficientemente alto y su entonación se modula para ser más comunicativo evitando una línea melódica plana.	1	0.890
	3. El docente formula preguntas durante la sesión al alumnado para que busquen modos alternativos de investigación o resolución de problemas.	1	0.838
Comunicativa	4. Constantemente realiza preguntas para verificar la comprensión de lo que se va exponiendo.	1	1
	5. Escucha respuestas del alumnado, las respeta y hace uso de ellas para conectar con otras ideas.	1	0.762
	6. En su exposición utiliza gestos, desplazamientos y otros movimientos que ayudan al proceso de enseñanza-aprendizaje.	1	0.882
Pedagógica	7. El docente explica los objetivos de la sesión antes de iniciar la misma.	1	0.853
	8. Conecta los contenidos y aprendizajes de la sesión con aprendizajes anteriores de otras sesiones y/o cursos/asignaturas.	1	0.815
	9. En las explicaciones, presenta modelos, ejemplos o aplicaciones prácticas para favorecer así su comprensión.	1	0.745
	10. El docente incluye algún tipo de metodología activa durante la sesión	1	0.901
	11. El docente indica durante la clase algún material/fuente complementaria (libro, artículo, película...) dónde profundizar sobre el contenido de la sesión.	1	0.879
Promedio		1	0.867

Nota: Kintra = Kappa de Cohen para fiabilidad intra-observador; Kinter = Kappa de Cohen para fiabilidad inter-observador

Como se observa en la Tabla 2, se implementó un análisis de generalizabilidad utilizando las facetas: observadores [OBS], ítems [ITE] y sesiones [SES] con un doble objetivo. Por un lado, para estimar si la codificación de los observadores tuvo un suficiente poder de generalización (e.g., adecuada calidad del dato) y, por otro, si las categorías y sesiones de observación fueron suficientemente heterogéneas. El análisis muestra una notable variabilidad asociada a las facetas ítems [sesiones] (61,20 %). Así mismo, la variabilidad fue inexistente o muy baja en la faceta observadores en su interacción con los ítems y las sesiones de observación.

Por otra parte, los índices de generalizabilidad para el modelo [ITE]/[OBS][SES] fueron de alta fiabilidad relativa (0.876) y absoluta (0.860). En el caso de del modelo [SES]/[ITE][OBS] los resultados ofrecieron una moderada fiabilidad relativa (0.625) y absoluta (0.564). En el caso del modelo [OBS]/[ITE][SES] obtuvieron coeficientes de generalizabilidad relativo y absoluto cercanos a 0. Estos resultados sugieren que la variabilidad en la evaluación proviene principalmente de los ítems, mientras que los observadores muestran consistencia, lo que respalda la fiabilidad del instrumento en contextos similares.

Tabla 2. Valores del análisis de la generalizabilidad y de los niveles de optimización

Fuentes de variación	Suma de cuadrados	GL	Cuadrado medio	Aleatorio	Mixtos	Corregidos	%	Error estándar
[OBS]	0,089	1	0,089	0	0	0	0,023	0
[ITE]	43,851	10	4,385	0,077	0,077	0,077	19,091	0,036
[OBS]/[ITE]	0,491	10	0,049	0,001	0,001	0,001	0,128	0,001
[SES]	34,793	24	1,45	0,041	0,041	0,041	10,225	0,018
[OBS][SES]	1,229	24	0,051	0,001	0,001	0,001	0,339	0,001
[ITE][SES]	126,967	240	0,529	0,246	0,246	0,246	61,201	0,024
[OBS]/[ITE][SES]	8,691	240	0,036	0,036	0,036	0,036	8,994	0,003



Modelo [OBS]/[ITE][SES]						
[OBS]	[ITE]	[SES]		Coefficiente G	Error	Desv. típica del error
(2 INF)	(11. INF)	(25. INF)	Relativo	0,281	0,010	0,15
			absoluto	0,001	0,00	0,99
Modelo [ITE]/[OBS][SES]						
[OBS]	[ITE]	[SES]		Coefficiente G	Error	Desv. típica del error
(2 INF)	(11. INF)	(25. INF)	Relativo	0,876	0,011	0,104
			absoluto	0,860	0,013	0,111
Modelo [SES]/[ITE][OBS]						
[OBS]	[ITE]	[SES]		Coefficiente G	Error	Desv. típica del error
(2 INF)	(11. INF)	(25. INF)	Relativo	0,625	0,025	0,157
			absoluto	0,564	0,032	0,158

Discusión

El objetivo del presente estudio fue analizar la fiabilidad de la herramienta POT-Higher Education (HE-PoT) con vistas a su futura aplicación y mejora de la Enseñanza Universitaria desde un enfoque colaborativo. Se partió de la herramienta diseñada por Domínguez-Montes et al., (2024) y sometida a una validez de contenido. El instrumento lo integraban 11 ítems repartidos en 3 dimensiones (aspectos formales, comunicativos y pedagógicos). Cada ítem incluía, a su vez, en tres categorías que evolucionan desde la conducta 'excelente', pasando por una 'en proceso' a una tercera y última 'inadecuada'.

El proceso de fiabilización atendió a una doble estrategia. Por un lado, se analizó el grado de concordancia entre diferentes usos de la herramienta a través de la Kappa de Cohen. Por otro lado, se analizaron los resultados a partir de la teoría de la generalizabilidad. En el primer caso se tuvo en cuenta tanto la concordancia en el registro de un observador consigo mismo (intraobservador) como la de dos observadores entre sí (interobservadores). La fiabilidad intraobservador obtenida fue la máxima en todos los ítems de la herramienta. Así mismo, aunque la fiabilidad interobservador fue sensiblemente menor que la intra, 9 de los 11 ítems obtuvieron una concordancia casi perfecta y dos los obtuvieron una fiabilidad considerable.

En este sentido, una revisión de trabajos similares que usaron la Kappa de Cohen para fiabilizar herramientas observacionales nos arroja luz sobre varios aspectos de los resultados obtenidos (Hashim & Ab Wahid, 2024; Burk et al., 2022; Var et al., 2025; Kojima et al., 2021; McNulty et al., 2020; Lachowsky et al., 2019; Strahl et al., 2018). En primer lugar, exceptuando el trabajo de Kojima et al., 2021, la mayoría de los trabajos anteriores no incluye la confiabilidad intraobservadores como medio para fiabilizar la herramienta. Solo la consistencia entre observadores es tenida en cuenta para considerar fiable la herramienta. Por ello, entendemos que la confiabilidad intraobservador es insuficiente para considerar la estabilidad de un instrumento y los resultados obtenidos deben complementarse con otras estrategias.

En el caso de la medida interobservador, el promedio de fiabilidad obtenido para los 11 ítems (0,86) es superior al alcanzado en los trabajos de Burk et al. (2022) McNulty et al. (2020), Lachowsky et al. (2019) y Strahl et al. (2018). Se encuentra alineado con los 0,82 y 0,90 obtenidos por Var et al., (2025) y Kojima et al., 2021 y, por último, se halla por debajo de la confiabilidad interobservadores perfecta hallada por el trabajo de Hashim y Ab Wahid (2024).

En cualquiera de los casos, algunos de estos trabajos, aun siendo herramientas observacionales tratan temáticas como la ansiedad, las habilidades deportivas o la educación para la salud. Sin bien sus resultados pueden ser tenidos en cuenta para referenciarse sobre criterios generales de fiabilidad, no hay una correspondencia directa con el tipo de instrumento, el contenido y los observadores de este estudio. Por esa razón, una comparativa con otros instrumentos de observación por pares en la enseñanza universitaria nos muestra dos certezas. En primer lugar, la tendencia a la proliferación del diseño y uso de esos instrumentos POT en esta etapa educativa sobre todo en Australia, EEUU e Inglaterra. En segundo lugar, la carencia de procesos de validación en el diseño de estos (Otero-Saborido et al., 2024). Los casos más recientes lo hallamos en la herramienta POT RTOP de Amrein-Beardsley y Popp (2012) y en el 'Project Based Learning' (PBL) de Garcia et al. (2017). En el RTOP, la fiabilidad y validez se investigaron en el contexto de la universidad y examinando lo que los participantes percibían que habían aprendido del proceso de evaluación del profesorado y los compañeros en el que habían participado. En el caso PBL para la observación del «aprendizaje basado en proyectos» de Garcia et al. (2017), la primera versión

del instrumento fue sometida a una prueba previa y completada por los cuatro observadores compañeros mientras veían dos sesiones de PBL grabadas en vídeo. Dos tutores grabados en vídeo también completaron el instrumento y luego fueron entrevistados para recabar sus sugerencias y comentarios.

Así mismo, los índices de generalizabilidad (TG) fueron incluidos como otra estrategia para comprobar la fiabilidad de la herramienta. Un instrumento con alta generalizabilidad asegura que la variabilidad de los datos proviene principalmente de lo que se quiere medir (las conductas docentes) y no de factores no deseados como el juicio subjetivo del observador o las particularidades de una sesión concreta. Esto es especialmente importante porque en POT suelen participar docentes observadores con

contextos diversos. Por la TG permite que el instrumento funcione en contextos docentes diversos perfiles, experiencias y disciplinas muy distintas. Sin bien los resultados obtenidos desde el punto de vista de la variabilidad de los ítems pudieran considerarse altos y generalizables; el instrumento, aún obteniendo valores de fiabilidad aceptables, muestra una fiabilidad más relativa desde el punto de vista de la variabilidad en el modelo Sesiones/ítems; Observadores. Este resultado podría explicarse por el reducido número de docentes observadores. Por ello, esta pudiera ser una de las limitaciones de este trabajo, junto con la dualidad de observar en una misma muestra sesiones en directo y sesiones grabadas, a pesar de la calidad del sistema de cámaras instalados por la propia universidad.

Por ello, como prospectiva de futuro este trabajo podría tener continuidad desde una doble perspectiva. Por un lado, diversificar los docentes y las asignaturas observadas y en diferentes centros y/o titulaciones universitarias para evaluar la fiabilidad relativa y absoluta en este modelo. Por otro, realizar estudios sobre muestras que sean observadas en una misma naturaleza, es decir, sin mezclar presencialidad y grabación.

Conclusiones

A partir de una doble estrategia de fiabilidad, se obtuvo un instrumento robusto y fiable para su aplicación en la enseñanza universitaria bajo un marco colaborativo. Los altos niveles de concordancia y generalizabilidad demuestran que la herramienta HE-PoT es un instrumento idóneo para proporcionar feedback objetivo y fomentar el desarrollo profesional entre el profesorado universitario. Además, esta herramienta cubre el vacío actual de instrumentos observacionales validados, ofreciendo un enfoque sistemático para mejorar la calidad del proceso de enseñanza-aprendizaje en diversos contextos académicos.

Agradecimientos

Los autores agradecen al profesorado participante de la Universidad Pablo de Olavide por colaborar en esta investigación.

Financiación

Esta investigación ha sido financiada por la Agencia Estatal de Investigación en el marco del V Plan Propio de Investigación y Transferencia 2018-2020 de la Universidad Pablo de Olavide (Referencia: PPI2201).

Afiliación de los autores

José Antonio Domínguez Montes (Universidad Pablo de Olavide, España), Javier Gálvez González (Profesor Titular de la Universidad Pablo de Olavide), Roos Van Gasse (Profesora adjunta de la KU Leuven, Bélgica) y Fernando Manuel Otero Saborido (Profesor Titular de la Universidad Pablo de Olavide).

Referencias

- Álvarez Medina, J., Murillo Lorente, V., Casterad Seral, J., & Nuviala, A. (2022). Valoración del aprendizaje técnico del voleibol mediante la metodología observacional por pares en estudiantes universitarios. *Cultura, Ciencia y Deporte*, 17(51), 31-39. <http://dx.doi.org/10.12800/ccd.v17i51.1662>
- Amrein-Beardsley, A., & Popp, S. E. O. (2012). Peer observations among faculty in a college of education: Investigating the summative and formative uses of the Reformed Teaching Observation Protocol (RTOP). *Educational Assessment, Evaluation and Accountability*, 24(1), 5-24. <https://doi.org/10.1007/s11092-011-9135-1>
- Anguera-Argilaga, M. T., Blanco Villaseñor, Á., Hernández Mendo, A., & Losada López, J. L. (2011). Diseños Observacionales: Ajuste y aplicación en psicología del deporte. *Cuadernos de Psicología del Deporte*, 11(2), 63-76. <https://revistas.um.es/cpd/article/view/133241>
- Aránguiz, C., Rivera-Vargas, P. y Imbernón, F. (2021). A una década del Plan Bolonia: posibilidades y límites de su implementación en la Universidad de Barcelona. *Revista de la educación superior*, 50(200), 85-104. <https://doi.org/10.36857/resu.2021.200.1891>
- Aznar Díaz, I., Hinojo-Lucena, F. J., Cáceres-Reche, M. P. y Romero-Rodríguez, J. M. (2020). Pedagogical Approaches in the Knowledge Society: The Flipped Classroom Method for the Development of Creativity and Dialogical Learning. *International Journal of Emerging Technologies in Learning*, 15(3), 4-14. <http://dx.doi.org/10.3991/ijet.v15i03.11664>
- Bailey, K.M. (2006). *Language Teacher Supervision: A Case-Based Approach*. Cambridge University Press.
- Bell, M. (2001). Supported reflective practice: A programme of peer observation and feedback for academic teaching development. *International Journal for Academic Development*, 6(1), 29-39. <https://doi.org/10.1080/13601440110033643>
- Brancato, V. C. (2003). Professional development in higher education. *New Directions for Adult and Continuing Education*, 98, 59-65. <https://doi.org/10.1002/ace.100>
- Burk, K. S., Zhao, A. H., Glazer, D. I., Giess, C. S., Boland, G. W., & Khorasani, R. (2022). Assessment of a large-scale peer learning program's value by manual review of case submissions. *Journal of the American College of Radiology*, 19(10), 1138-1150. <https://doi.org/10.1016/j.jacr.2022.04.012>
- Byrne, J., Brown, H., & Challen, D. (2010). Peer development as an alternative to peer observation: A tool to enhance professional development. *International Journal for Academic Development*, 15(3), 215-228. <https://doi.org/10.1080/1360144X.2010.497685>
- Carragher, J., & McGaughey, J. (2016). The effectiveness of peer mentoring in promoting a positive transition to higher education for first-year undergraduate students: A mixed methods systematic review protocol. *Systematic Reviews*, 5(1), Article 68. <https://doi.org/10.1186/s13643-016-0245-1>
- Chao, T., Saj, T., & Tessier, F. (2006). Establishing a quality review for online courses. *Educause Quarterly*, 29(3), 32-39.
- Chism, N. (2007). *Peer review of teaching: A sourcebook*. Anker.
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and psychological measurement*, 20(1), 37-46. <https://doi.org/10.1177/001316446002000104>
- Corcelles-Seuba M., Ortiz M., O'Leary M., & Duran D. (2025). A systematic review of reciprocal peer observation in higher education. *Professional Development in Education*, 1-18. <https://doi.org/10.1080/19415257.2025.2480768>
- Corcelles-Seuba, M., Duran, D., Flores, M., Miquel, E., y Ribosa, J. (2023). Percepciones docentes sobre observación entre iguales: resistencias, agencia, procedimiento y objetivos de mejora. *Estudios sobre Educación*, 44, 35-58. <https://doi.org/10.15581/004.44.002>
- Crabtree, J. L., Scott, P. J., & Kuo, F. (2016). Peer Observation and Evaluation Tool (POET): A Formative Peer Review Supporting Scholarly Teaching. *The Open Journal of Occupational Therapy*, 4(3). <https://doi.org/10.15453/2168-6408.1273>
- Cronbach, L. J. (1972). *The dependability of behavioral measurements: Theory of generalizability for scores and profiles*. Wiley.
- De la Iglesia, B., Duma, L., Forteza, D., Llabrés, J., Moreno-Tallón, F. y Ribosa, J. (2025). El feedback entre docentes en un ciclo de observación entre iguales. *REICE. Revista Iberoamericana sobre Calidad, Eficacia y Cambio en Educación*, 23(3). <https://doi.org/10.15366/reice2025.23.3.007>



- Domínguez-Montes, J.A., Cenizo-Benjumea, J.M., Gálvez-González, J. & Otero-Saborido, F. (2024). Design and validation of a university education peer observation tool. *Revista de Gestão Social e Ambiental*, 18(10), 1-19. <https://doi.org/10.24857/rgsa.v18n10-301>
- Eddy, S., Converse, M., & Wenderoth, M. (2015). PORTAAL: A classroom observation tool assessing evidence-based teaching practices for active learning in large science, technology, engineering, and mathematics classes. *CBE Life Sciences Education*, 14(2), es2. <https://doi.org/10.1187/cbe.14-06-0095>
- Escartí, A., Gutiérrez, M., Pascual, C., & Wright, P. (2013). Observación de las estrategias que emplean los profesores de educación física para enseñar responsabilidad personal y social. *Revista de Psicología del Deporte*, 22(1), 159-166.
- Fletcher, J. A. (2018). Peer observation of teaching: A practical tool in higher education. *The Journal of Faculty Development*, 32(1), 51-64. <https://doi.org/10.13140/RG.2.2.19455.82084>
- Flores, M., Miquel, E., & Duran, D. (2025). Teacher collaborative knowledge building in Reciprocal Peer Observation. *European Journal of Psychology of Education*, 40(1), 45. <https://doi.org/10.1007/s10212-025-00945-7>
- Flores-Ferro, E., Maureira Cid, F., Castillo-Retamal, F., Matus-Castillo, C., Gamboa Jimenez, R., Moral Moreno, L., Muñoz Lara, M., Portugal Villar, J., & Saavedra Esteban, E. (2025). Construcción y validación la Escala de Percepción de la Investigación Universitaria (EPIU) en estudiantes de Educación Física de Chile. *Retos*, 73, 66-77. <https://doi.org/10.47197/retos.v73.116575>
- Gabín, B., Camerino, O., Anguera, M.T. & Castañer, M. (2012). Lince: Multiplatform sport analysis software. *Procedia Social and Behavioral Sciences*, 46, 4692-4694. <https://doi.org/10.1016/j.sbspro.2012.06.320>
- Gallardo-Fuentes, F., Carter Thuillier, B., Martínez-Angulo, C., Gallardo-Fuentes, J., & Peña Troncoso, S. (2025). Percepciones del alumnado en dos cohortes de la formación inicial en Educación Física: consistencia, compromiso y retos en la evaluación. *Retos*, 70, 39-50. <https://doi.org/10.47197/retos.v70.113463>
- Garcia, I., James, R. W., Bischof, P., & Baroffio, A. (2017). Self-Observation and Peer Feedback as a Faculty Development Approach for Problem-Based Learning Tutors: A Program Evaluation. *Teaching and Learning in Medicine*, 29(3), 313-325. <https://doi.org/10.1080/10401334.2017.1279056>
- Georgiou, H., Sharma, M., & Ling, A. (2018). Peer review of teaching: What features matter? A case study within STEM faculties. *Innovations in Education and Teaching International*, 55(2), 190-200. <https://doi.org/10.1080/14703297.2017.1342557>
- Gosling, D. (2002). *Models of Peer Review of Teaching*. Generic Centre: Learning and Teaching Support Network, University of East London.
- Hammersley-Fletcher, L., y Orsmond, P. (2005). Reflecting on reflective practices within peer observation. *Studies in Higher Education*, 30(2), 213-224. <https://doi.org/10.1080/03075070500043358>
- Hashim, N., & Ab Wahid, H. (2024). Entrepreneurial Mentoring Synergy: An In-Depth Analysis of Polytechnic and Community College (PolyCC) Lecturers as Mentors to Student Start-Up Entrepreneurs. *PaperASIA*, 40(4b), 239-247. <https://doi.org/10.59953/paperasia.v40i4b.206>
- Hernández-Mendo, A., Blanco-Villaseñor, Á., Pastrana, J. L., Morales- Sánchez, V., & Francisco Jesús Ramos-Pérez. (2016). SAGT: Aplicación Informática para Análisis de Generalizabilidad. *Revista Iberoamericana de Psicología del Ejercicio y el Deporte*, 11(1), 77-89.
- Huston, T., & Weaver, C.L. (2008). Peer coaching: Professional development for experienced faculty. *Innovative Higher Education*, 33(1), 5-20. <https://doi.org/10.1007/s10755-007-9061-9>
- Ibarra-Sáiz, M. S., Rodríguez-Gómez, G., Boud, D., Rotsaert, T., Brown, S., Salinas Salazar, M. L., & Rodríguez Gómez, H. M. (2020). El futuro de la evaluación en la educación superior. *RELIEVE - Revista Electrónica De Investigación Y Evaluación Educativa*, 26(1). <https://doi.org/10.7203/relieve.26.1.17323>
- Jia, Y., Spagnolo, A. B., Barrett, N., Murphy, A. A., Basto, P. M., Rothpletz-Puglia, P., & Luther, S. (2025). Validation of a peer observation and evaluation tool for online teaching in the US. *Educational technology research and development*, 73(1), 615-639. <https://doi.org/10.1007/s11423-024-10428-z>
- Johnston, A. L., Baik, C., & Chester, A. (2022). Peer review of teaching in Australian higher education: a systematic review. *Higher Education Research and Development*, 41(2), 390-404. <https://doi.org/10.1080/07294360.2020.1845124>



- Kojima, M., Kinomura, Y., & Kuzuhara, K. (2021). Development of observational indicators for evaluating handstand posture in the mat exercise in physical education class: Validity and reliability. *Journal of Physical Education and Sport*, 21(3), 2087–2096. <https://doi.org/10.7752/jpes.2021.s3266>
- Lachowsky NJ, Card KG, Cui Z, Sereda P, Roth EA, Hogg RS, Moore DM. (2019) Agreement between gay, bisexual and other men who have sex with men's period prevalence and event-level recall of sexual behaviour: an observational respondent-driven sampling study. *Sexual Health* 16, 84–87. <https://doi.org/10.1071/SH17223>
- Landis, J. R., y Koch, G. G. (1977). A one-way components of variance model for categorical data. *Biometrics*, 33, 671–679. <https://doi.org/10.2307/2529465>
- Ley Orgánica 2/2023, de 22 de marzo, del Sistema Universitario. *Boletín Oficial del Estado*, núm. 70, de 23 de marzo de 2023, pp. 43267 a 43339. <https://www.boe.es/eli/es/lo/2023/03/22/2/con>
- Little, B. B. (2009). Quality assurance for online nursing courses. *Journal of Nursing Education*, 48(7), 381–387. <https://doi.org/10.3928/01484834-20090615-05>
- Matarranz, M. (2021). El Espacio Europeo de Educación Superior y su sello de calidad. *Revista española de educación comparada*, 37, 153–1573. <https://doi.org/10.5944/reec.37.2021.27728>
- McGahan, S. J., Jackson, C. M., & Premer, K. (2015). Online course quality assurance: Development of a quality checklist. *InSight: A Journal of Scholarly Teaching*, 10, 126–140. <https://doi.org/10.46504/10201510mc>
- McNulty, C. A., Syeda, R. B., Brown, C. L., Bennett, C. V., Schofield, B., Allison, D. G., ... & Francis, N. (2020). Peer-education as a tool to educate on antibiotics, resistance and use in 16–18-year-olds: a feasibility study. *Antibiotics*, 9(4), 146. <https://doi.org/10.3390/antibiotics9040146>
- Menges, R. J. (1987). Colleagues as catalysts for change in teaching. In Kurfiss, J. (Ed.), *To improve the academy: Resources for faculty, instructional and organizational development*, 6 (pp. 83–93). Stillwater, OK: New Forums Press.
- OECD (2019). *TALIS 2018 Results (Volume I): Teachers and school leaders as lifelong learners*. OECD Publishing. <https://www.oecd.org/education/talis-2018-re-sults-volume-i-1d0bc92a-en.htm>
- Otero-Saborido, F., Domínguez-Montes, J.A., Cenizo-Benjumea, J.M. & González-Calvo, G. (2024). Peer observation of teaching in higher education: Systematic review of observation tools. *Educational Process: International Journal*, 13(1), 84–101. <https://doi.org/10.22521/edupij.2024.131.6>
- Peel, D. (2005). Peer observation as a transformatory tool? *Teaching in Higher Education*, 10(4), 489–504. <https://doi.org/10.1080/13562510500239125>
- Portell, M., Anguera, M. T., Chacón-Moscoso, S., & Sanduvete-Chaves, S. (2015). Guidelines for reporting evaluations based on observational methodology. *Psicothema*, 27(3), 283–289. <https://doi.org/10.7334/psicothema2014.276>
- Rabada-Rice, F., & Scott, R. S. (1986). A peer evaluation for measuring team teaching effectiveness. *The Journal of Nursing Education*, 25(6), 255–258. <https://doi.org/10.3928/0148-4834-19860601-10>
- Richards, J. C., & Farrell, T. S. C. (2005). *Professional development for language teachers: Strategies for teacher learning*. Cambridge University Press.
- Ridge, B. L., & Lavigne, A. L. (2020). Improving instructional practice through peer observation and feedback. *Education Policy Analysis Archives*, 28(61). <https://doi.org/10.14507/EPAA.28.5023>
- Sachs, J., & Parsell, M. (2014). *Peer review of learning and teaching in higher education: International perspectives*. Springer.
- Shortland, S. (2010). Feedback within peer observation: continuing professional development and unexpected consequences. *Innovations in Education and Teaching International*, 47(3), 295–304. <https://doi.org/10.1080/14703297.2010.498181>
- Skinner, M. E., & Welch, F. C. (1996). Peer coaching for better teaching. *College Teaching*, 44, 153–156.
- Smith, M. K., Jones, F. H., Gilbert, S. L., & Wieman, C. E. (2013). The classroom observation protocol for undergraduate STEM (COPUS): A new instrument to characterize university STEM classroom practices. *CBE—Life Sciences Education*, 12, 618–627. <https://doi.org/10.1187/cbe.13-08-0154>
- Soto, A., Camerino, O., Iglesias, X., Anguera, M.T. y Castañer, M. (2019). LINC PLUS: Software de investigación para análisis de vídeo del comportamiento. *Apunts. Educación Física y Deportes*, 137, 149–153. [https://dx.doi.org/10.5672/apunts.2014-0983.es.\(2019/3\).137.11](https://dx.doi.org/10.5672/apunts.2014-0983.es.(2019/3).137.11)

- Soto-Fernández, A., Camerino, O., Iglesias, X., Anguera, M. T., & Castañer, M. (2022). LINC PLUS software for systematic observational studies in sports and health. *Behavior Research Methods*, 54, 1263–1271. <https://doi.org/10.3758/s13428-021-01642-1>
- Stone, A. A., & Shiffman, S. (2002). Capturing momentary, self-report data: A proposal for reporting guidelines. *Annals of Behavioral Medicine*, 24(3), 236-243. https://doi.org/10.1207/S15324796ABM2403_09
- Strahl A, Gerlich C, Alpers GW, Ehrmann K, Gehrke J, Müller-Garnn A, & Vogel H. (2018). Development and evaluation of a standardized peer-training in the context of peer review for quality assurance in work capacity evaluation. *BMC Med Educ*, 18(1), 135. <https://doi.org/10.1186/s12909-018-1233-z>
- Sullivan, P. B., Buckle, A., Nicky, G., & Atkinson, S. H. (2012). Peer observation of teaching as a faculty development tool. *BMC Medical Education*, 12, 26. <https://doi.org/10.1186/1472-6920-12-26>
- Thomas, S., Chie, Q. T., Abraham, M., Jalarajan Raj, S., & Beh, L. S. (2014). A qualitative review of literature on peer review of teaching in higher education: An application of the SWOT framework. *Review of Educational Research*, 84(1), 112–159. <https://doi.org/10.3102/0034654313499617>
- Torres, A. C., Lopes, A., Valente, J. M. S., & Mouraz, A. (2017). What catches the eye in class observation? Observers' perspectives in a multidisciplinary peer observation of teaching program. *Teaching in Higher Education*, 22(7), 822-838. <https://doi.org/10.1080/13562517.2017.1301907>
- Trujillo, J. M., DiVall, M. V., Barr, J., Gonyeau, M., Van Amburgh, J. A., Matthews, S. J., & Qualters, D. (2008). Development of a peer teaching-assessment program and a peer observation and evaluation tool. *American Journal of Pharmaceutical Education*. [https://doi.org/10.1016/S0002-9459\(24\)00511-4](https://doi.org/10.1016/S0002-9459(24)00511-4)
- Var, E., Cam, Y., Pektaş, E., & Kuden, C. (2025). Does preoperative anxiety influence endodontic skill performance and students' subjective perceptions in an observational simulation-based study? *BMC Medical Education*, 25(1), 1382. <https://doi.org/10.1186/s12909-025-07957-3>
- Yiend, J., Weller, S., & Kinchin, I. (2014). Peer observation of teaching: The interaction between peer review and developmental models of practice. *Journal of Further and Higher Education*, 38(4), 465–484. <https://doi.org/10.1080/0309877X.2012.726967>

Datos de los/as autores/as y traductor/a:

José Antonio Domínguez Montes
Javier Gálvez González
Roos Van Gasse
Fernando Manuel Otero Saborido

jadommon@alumno.upo.es
jgalgon@upo.es
roos.vangasse@kuleuven.be
fmotero@upo.es

Autor
Autor
Autora
Autor

