

# **Análisis de las dificultades del profesorado Educación Secundaria para evaluar: entre la burocracia y la dificultad técnica**

## **Analysis of Secondary Education Teachers' Difficulties in Assessment: Between Bureaucracy and Technical Complexity**

**Esteban Vázquez-Cano**

Facultad de Educación, Universidad Nacional de Educación a Distancia (UNED), España

Email: [evazquez@edu.uned.es](mailto:evazquez@edu.uned.es)

ORCID iD: <https://orcid.org/0000-0002-6694-7948>

Recibido / Received: 06/01/2026

Aceptado / Accepted: 14/03/2026

**Resumen:** Este estudio analiza las dificultades percibidas por el profesorado de Educación Secundaria para evaluar al alumnado en el marco competencial de la LOMLOE. Se empleó un diseño cuantitativo, transversal y correlacional, aplicando un cuestionario de 16 ítems tipo Likert a 403 docentes de centros públicos españoles. Mediante psicometría de redes se estimó un modelo gráfico gaussiano regularizado (GLASSO) con selección de penalización por validación cruzada, del que se obtuvieron índices de centralidad, estabilidad y comunidades de nodos. Los resultados muestran un patrón de alta dificultad concentrada en tareas procedimentales: crear el cuaderno del profesor, generar pruebas alineadas con criterios de evaluación y justificar la calificación final, junto con la evaluación de competencias clave y del perfil competencial. La red revela tres bloques interdependientes: normativo, competencial y procedimental-calificativo, donde este último actúa como eje estructural del sistema de dificultades, al conectar la interpretación normativa con la toma de decisiones de calificación. No se observaron diferencias significativas por sexo, mientras que el profesorado con Máster en Formación del Profesorado informó menores dificultades competenciales que quienes cursaron el antiguo CAP, y la experiencia docente se asoció con una ligera disminución de las dificultades percibidas. Los hallazgos subrayan la necesidad de fortalecer la competencia metaevaluativa mediante formación inicial y continua que integre aspectos normativos, diseño de instrumentos y gestión técnica del registro evaluador, así como el potencial del análisis de redes como herramienta diagnóstica para orientar políticas de mejora.

**Palabras clave:** evaluación, profesorado de secundaria, psicometría de redes, dificultades docentes, competencia evaluadora.

**Abstract:** This study examines secondary teachers' perceived difficulties in assessing students within the competency-based framework introduced by the Spanish LOMLOE reform. A quantitative, cross-sectional and correlational design was employed, administering a 16-item Likert-type questionnaire to 403 teachers from public schools. Using network psychometrics, a regularized Gaussian graphical model (GLASSO) with cross-validated penalization was estimated, from which centrality, stability and community indices were derived. Results show a marked pattern of high difficulty concentrated in procedural tasks: constructing the teacher gradebook, generating assessment tasks aligned with curriculum criteria, and justifying final grades, alongside evaluating key competences and students' competence profiles. The network reveals three interdependent blocks—normative, competency-based and procedural—grading—where the latter functions as the structural core of the system, linking regulatory interpretation to grading decisions. No significant gender differences were found; however, teachers who completed the Master's in Teacher Education reported fewer competency-related difficulties than those trained under the former CAP programme, and teaching experience was associated with a slight reduction in perceived difficulties. Overall, the findings highlight the need to strengthen teachers' meta-evaluative competence through initial and ongoing training that integrates

regulatory interpretation, instrument design and the technical management of assessment records. They also demonstrate the potential of network analysis as a diagnostic tool to inform evidence-based assessment policy and school-level improvement.

**Keywords:** Educational Assessment, Secondary Teachers, Network Psychometrics, Teacher Difficulties, Evaluative Competence.

## 1. Introducción

La evaluación constituye uno de los pilares fundamentales del proceso educativo, no solo por su función acreditadora de títulos oficiales, sino también por su capacidad para orientar la enseñanza, regular el aprendizaje y garantizar la equidad y la calidad educativa. En el contexto de la Educación Secundaria, la normativa española, particularmente la Ley Orgánica 3/2020 (LOMLOE) y el Real Decreto 217/2022, enfatizan que la evaluación debe ser continua, formativa e integradora, basada en criterios de desempeño vinculados a las competencias clave y a los descriptores operativos del perfil de salida. Sin embargo, la traslación de este marco legal a la práctica docente cotidiana plantea importantes desafíos. La complejidad normativa, la sobrecarga burocrática y las exigencias técnicas derivadas de la evaluación de los criterios de evaluación generan tensiones que afectan al desempeño del profesorado para evaluar de forma objetiva, coherente y pedagógicamente significativa, así como a una excesiva sobrecarga de tiempo para realizar las tareas asociadas a la evaluación.

En este escenario, comprender cómo los docentes perciben y gestionan las dificultades evaluativas se vuelve esencial para mejorar la calidad del sistema educativo. La formación pedagógica específica constituye un factor clave: disponer de una preparación sólida en evaluación, ya sea a través del Máster en Formación del Profesorado o de programas de formación permanentes y continuos, puede determinar la manera en la que el profesorado interpreta la normativa, diseña instrumentos y aplica criterios de calificación objetivos y coherentes con la normativa en vigor dentro de parámetros reales más allá de propuestas irreales y utópicas de complicadas hojas Excel o programas informáticos.

Desde esta perspectiva, la psicometría de redes ofrece un marco idóneo para estudiar las dificultades evaluativas como un sistema de interacciones entre manifestaciones observables, en lugar de asumir un rasgo latente único. En Educación, los procesos evaluativos emergen de la co-activación de componentes normativos (interpretación y aplicación de la legislación), competenciales (traducción de criterios y valoración del desempeño) y procedimentales (instrumentos, calificación y comunicación), cuya dependencia mutua puede generar efectos en cascada. La modelización de redes permite estimar asociaciones parciales entre ítems (arcos) y detectar nodos que actúan como puentes o “cuellos de botella”, así como módulos de dificultades funcionalmente conectadas. Este enfoque se ha consolidado en Educación para analizar fenómenos complejos, como la autorregulación, las estrategias de aprendizaje, el compromiso, el estrés/*burnout* docente o la retroalimentación, que no se explican adecuadamente como constructos unidimensionales, sino como sistemas dinámicos. A partir de esta base, formulamos un análisis de redes que permite: (1) identificar los nodos más influyentes (centralidad), (2) estimar la estabilidad de la estructura, (3) detectar comunidades de dificultades, y (4) comparar subredes por sexo, formación y experiencia, con el fin de orientar decisiones formativas y de política educativa basadas en evidencia.

## 1.1. La evaluación en Educación Secundaria

La evaluación constituye un elemento fundamental de los procesos de enseñanza-aprendizaje y de las funciones atribuidas al profesorado (Conley, 2015; Rudner y Schafer, 2002). Recientemente, la OCDE ha evidenciado la percepción del profesorado español sobre la alta carga burocracia en el sistema educativo, en buena parte asociada a los procesos de gestión de la evaluación (OECD, 2025). En este sentido, una de las principales funciones atribuidas a los docentes en la legislación española es la evaluación (Ley Orgánica 3/2020 / Art. 91b), “La evaluación del proceso de aprendizaje del alumnado, así como la evaluación de los procesos de enseñanza”. Según se establece en la normativa española (Ley Orgánica 3/2020; Real Decreto 217/2022), la evaluación del proceso de aprendizaje del alumnado de la Educación Secundaria Obligatoria es continua, formativa e integradora. Para ello, las administraciones educativas deben garantizar el derecho del alumnado a ser evaluado con objetividad y a que su dedicación, esfuerzo y rendimiento sean valorados y reconocidos como merecen. Con carácter general, los referentes para la evaluación son los criterios de evaluación recogidos para cada curso, ámbito o materia en los currículos que desarrollen lo establecido en el Real Decreto 217/2022, por el que se establece la ordenación y las enseñanzas mínimas de la Educación Secundaria Obligatoria. En este contexto, el marco normativo de la evaluación es bastante homogéneo en casi todas las comunidades autónomas y se basa en la comprobación de la superación de las competencias específicas de cada materia a través de los criterios de evaluación del currículo.

Derivadas de estas consideraciones, el profesorado debe emplear diferentes procedimientos de evaluación: exámenes, controles, trabajos individuales y colaborativos, fichas, preguntas orales y escritas, etc. de forma que la evaluación no se referencie únicamente a un procedimiento de evaluación como podía ser tradicionalmente el examen (Álvarez Méndez, 2008; Arthur et al., 2012; Vázquez-Cano, 2012). Asimismo, se recomienda el uso de variados instrumentos de evaluación que permitan objetivar la calificación de los estudiantes y de las pruebas que se realizan mediante el uso de listas de cotejo o control o rúbricas diarios de aprendizaje dianas de evaluación etc. (Ahumada, 2005; Vázquez-Cano et al., 2022). Durante muchos años, la evaluación del proceso de enseñanza-aprendizaje en el marco de las diferentes leyes educativas (LOGSE, LOE, LOMCE, LOMLOE) se ha referenciado a la superación de los criterios de evaluación, y con ellos, a la superación de los objetivos o competencias asociadas a las asignaturas o materias. Pero la práctica diaria en los procesos de evaluación del profesorado se regía por otros principios en los que se podía ver de forma repetitiva en las programaciones didácticas un epígrafe sobre criterios de evaluación y calificación de las asignaturas y materias que correspondía a estimar los procedimientos de evaluación con un porcentaje representativo de la nota: Examen (70%), Trabajos (20%) y Actitud (10%) (Del Pozo Andrés y Ramos Zamora, 2001; Vázquez-Cano, 2012). Este enfoque se replicaba en los cuadernos del docente en el que se registraban las notas de los estudiantes conforme a una escala muy similar a la anterior (Vázquez-Cano, 2016b).

Este sistema que ha pervivido durante muchos años en los centros educativos ha variado sustancialmente en los últimos años con la irrupción de la LOMLOE y el impulso de diferentes actuaciones en los servicios de inspección de las comunidades autónomas para valorar la consecución de la objetividad en la evaluación (Polo Martínez, 2012). Estas actuaciones han ido encaminadas a impulsar un giro hacia

el cumplimiento de la norma para verificar que la evaluación del estudiante se referencia no a la superación de las pruebas, sino a la superación de los criterios de evaluación que marca el currículo y, por consiguiente, a la superación de las competencias específicas de las materias a través de los criterios de evaluación. En este sentido, se hace necesario que cada una de las pruebas que se realicen para evaluar a los estudiantes se debe vincular pertinentemente con los criterios de evaluación que esa prueba está evaluando. Este sistema, que puede parecer sencillo, ha generado y genera diferentes dudas y dificultades entre el profesorado; no solo en la parte técnica de aplicación, sino por la alta burocracia que supone la justificación tanto del aprobado como del suspenso de los estudiantes, así como para generar el prescriptivo perfil competencial del estudiante al finalizar la etapa (Polo Martínez, 2015; Vázquez-Cano, 2016a; Vázquez-Cano et al., 2022).

Hay que tener en cuenta también que el contexto curricular actual se enfoca a la consecución de competencias clave y específicas. En primer lugar, las competencias específicas relacionadas con las materias y, posteriormente, al finalizar la etapa con la consecución de un perfil de salida organizado en descriptores operativos y relacionados con las competencias clave. Este enfoque competencial obliga al profesorado a diseñar un proceso de enseñanza-aprendizaje mucho más práctico y funcional que combina teoría con práctica a través de la inclusión de un nuevo elemento del currículo como son las situaciones de aprendizaje (Real Decreto 217/2022 / Art. 2f).

La evaluación en este diseño curricular presenta ciertas áreas que suscitan controversia y dificultades entre el profesorado y entre las que destacamos: el desarrollo de la evaluación formativa (Nicol y Macfarlane-Dick, 2006), la distinción y aplicación de diferentes técnicas de evaluación (Castejón Oliva et al., 2009; Hamodi, López Pastor y López Pastor, 2015), las consecuencias del carácter continuo e integrador de la evaluación (Bordas Alsina y Cabrera Rodríguez, 2001), la adecuación de la normativa a la autonomía de los centros y las características del grupo-aula concreto (Vázquez-Cano et al., 2022), la determinación de la superación mínima conforme a la superación de las competencias específicas y los criterios de evaluación de las materias (Vázquez-Cano, 2016b) y el diseño y puesta en práctica de forma real y adaptada a los requerimientos normativos del cuaderno docente en el que se anota de forma periódica el desempeño y progreso de los estudiantes, pieza clave en el desarrollo del proceso de evaluación, ya que permite orientar al estudiante, retroalimentar el proceso de enseñanza aprendizaje, adaptarlo al contexto y generar una evaluación objetiva de los estudiantes (Hidalgo y Murillo, 2017).

Asimismo, en el proceso de evaluación confluyen una serie de dimensiones en las que se entremezclan factores normativos, de desarrollo metodológico para evaluar un proceso de enseñanza-aprendizaje competencial y el desconocimiento en muchas ocasiones de la diferencia conceptual entre carácter, técnica, procedimientos e instrumentos de evaluación (Cabrera-Rodríguez, 2011; Carless et al., 2011; Sánchez, 2018; Vázquez-Cano et al., 2022), a la que tampoco contribuye mucho a la normativa mezclando conceptos en las propios decretos de currículo y órdenes de evaluación entre las diferentes comunidades autónomas.

En este sentido, estas dificultades no son solo inherentes al sistema español, diversos estudios internacionales confirman que la transición hacia modelos curriculares basados en competencias genera tensiones estructurales entre la

normativa, la práctica pedagógica y los procedimientos técnicos de evaluación. Estas dificultades emergen cuando los docentes deben traducir estándares abstractos en tareas y rúbricas concretas, sin una guía uniforme ni suficiente formación en competencia evaluadora (Brookhart, 2011; DeLuca, LaPointe-McEwan y Luhanga, 2016). Investigaciones recientes evidencian que la sobrecarga burocrática, la falta de tiempo y la ambigüedad normativa limitan la adopción real de la evaluación formativa y la coherencia entre criterios y calificaciones (Veugen, Gulikers y den Brok, 2024; Yan et al., 2021). En contextos europeos y latinoamericanos, la aplicación del enfoque competencial se asocia con un incremento de la carga técnica y emocional del profesorado, especialmente en la Educación Secundaria (Crujeiras y Jiménez-Aleixandre, 2013). En conjunto, esta evidencia sugiere que las dificultades evaluativas no son individuales, sino estructurales, y refuerza la pertinencia de analizarlas como sistemas interdependientes mediante psicometría de redes, capaz de identificar núcleos y conexiones entre dimensiones normativas, competenciales y procedimentales.

## 2. Metodología

La investigación se inscribe en un diseño cuantitativo, no experimental y transversal, de carácter descriptivo-correlacional, orientado a identificar y modelar las interrelaciones entre las dificultades que el profesorado de Educación Secundaria percibe al evaluar al alumnado. El estudio se fundamenta en el enfoque de la psicometría de redes. Este enfoque resulta especialmente apropiado en contextos educativos donde las dificultades emergen de la interacción entre dimensiones no concretas susceptibles de interrelacionarse más que de factores individuales aislados (Álvarez-Díaz et al., 2022). El análisis se realizó íntegramente con software de libre acceso en lenguaje Python (versión 3.11), utilizando los paquetes “pandas” y “numpy” para la gestión y estandarización de los datos, “scipy” para los análisis estadísticos no paramétricos, y “matplotlib” para la generación de visualizaciones. A continuación, se estimó un modelo gráfico gaussiano regularizado (GLASSO), técnica ampliamente utilizada en la literatura científica (Epskamp, Borsboom y Fried, 2018) para identificar estructuras parciales de dependencia entre variables psicológicas y educativas. La estimación del modelo GLASSO se realizó mediante GraphicalLassoCV (scikit-learn), que selecciona el parámetro de penalización  $\lambda$  mediante validación cruzada k-fold ( $k = 5$ ). Este procedimiento explora una rejilla de valores posibles y elige aquel que minimiza el error de predicción, garantizando una red parsimoniosa y estable. A partir de la matriz de precisión se derivaron las correlaciones parciales controladas, representadas en una red visual generada con el paquete networkx, aplicando un diseño de fuerza tipo Fruchterman–Reingold.

Sobre la red resultante se calcularon las métricas de centralidad (grado, intermediación y cercanía) con el fin de identificar los nodos más influyentes en la estructura global de dificultades, así como la determinación de comunidades mediante el algoritmo de Louvain, que permitió identificar conglomerados funcionales o “módulos” de dificultades interconectadas. Para evaluar la estabilidad y fiabilidad de las métricas de centralidad se aplicó un procedimiento de bootstrap no paramétrico con 120 remuestreos, obteniendo intervalos de confianza al 95%. Este procedimiento, desarrollado con “numpy” y visualizado con “matplotlib”, permitió verificar la consistencia de las estimaciones ante posibles variaciones muestrales.

Finalmente, se analizaron las posibles diferencias estructurales según variables sociodemográficas y profesionales mediante pruebas U de Mann–Whitney (sexo y formación pedagógica) y la prueba de Kruskal–Wallis (experiencia docente). Para el análisis de redes se empleó un modelo gráfico gaussiano regularizado (GLASSO), idóneo para este tipo de datos porque estima correlaciones parciales controlando el resto de ítems y genera redes esparsas mediante penalización L1. La estabilidad del modelo se evaluó mediante procedimientos bootstrap aplicados a las métricas de centralidad. Para comparar estructuras entre subgrupos, se aplicó una adaptación del Network Comparison Test basada en 1.000 permutaciones, recalculando en cada iteración la fuerza global y la matriz GLASSO con igual penalización. Este procedimiento permite identificar diferencias en la conectividad y robustez de las redes con validez computacional y replicabilidad metodológica.

## 2.1. Participantes

La muestra estuvo compuesta por 403 docentes en ejercicio de Educación Secundaria Obligatoria y Bachillerato que ejercían en centros públicos de 15 comunidades autónomas españolas (56.33% mujeres y 43.67% hombres, con una edad media de 37,9 años). Conforme a la fórmula de cálculo muestral para poblaciones finitas, y considerando un nivel de confianza del 95% y un error máximo del  $\pm 5\%$ , una población de 128.000 docentes de Educación Secundaria requiere una muestra mínima de 383 casos. Por tanto, la muestra obtenida de 403 docentes resulta estadísticamente representativa, asegurando un margen de error aproximado de  $\pm 4.8\%$ , adecuado para realizar análisis inferenciales y comparativos con validez estadística. La participación fue completamente anónima y confidencial, garantizando la conformidad con los principios éticos de la Declaración de Helsinki y la Ley Orgánica 3/2018, de Protección de Datos Personales y Garantía de los Derechos Digitales.

## 2.2. Instrumento

El instrumento de recogida de información fue un cuestionario estructurado de 16 ítems tipo Likert (1 = nada de dificultad; 5 = mucha dificultad), diseñado ad hoc para evaluar la percepción de las dificultades del profesorado en los procesos de evaluación del alumnado: Enlace: <https://forms.office.com/e/4S04RNGVsn?origin=IprLink> (Anexo I). La validación de contenido del instrumento se realizó mediante la técnica Delphi y el cálculo del coeficiente de competencia experta. El panel de especialistas estuvo conformado por profesores universitarios, docentes de educación secundaria con más de diez años de experiencia profesional e inspectores/as de educación. Los valores de oscilaron entre 0.75 y 0.95, mientras que los de se situaron entre 0.70 y 0.90. En consecuencia, el coeficiente global medio de competencia experta fue  $\phi = 0.85$  (DT = 0.06), lo que indica un nivel alto de competencia según los criterios de validación establecidos ( $\phi \geq 0.80$ ). El análisis de fiabilidad interna mediante el alfa de Cronbach reveló valores adecuados en todas las dimensiones: normativa ( $\alpha = 0.86$ ), competencial ( $\alpha = 0.83$ ) y procedimental ( $\alpha = 0.79$ ), con un alfa global de 0.88, lo que evidencia una consistencia interna elevada. Cada ítem se trató como un nodo potencial en la red, de modo que las interrelaciones entre ellos representarían un sistema de dependencias mutuas entre las diferentes dimensiones de la dificultad evaluativa.

La consistencia interna y la adecuación de los ítems como variables ordinales se comprobaron mediante correlaciones policóricas y el coeficiente  $\omega$  ordinal.

### 3. Resultados

La Tabla 1 presenta los estadísticos descriptivos correspondientes a los 16 ítems que integran la escala sobre dificultades del profesorado para realizar la evaluación en Educación Secundaria. Cada ítem refleja un aspecto concreto del proceso evaluativo (desde la interpretación normativa hasta la calificación y comunicación de resultados) y fue valorado mediante una escala tipo Likert de cinco puntos (1 = nada de dificultad; 5 = mucha dificultad).

Tabla 1: estadísticos descriptivos de los ítems tipo Likert.

Ítem	Variable “Tengo Dificultad para”	N	M	DT	Asim.	Curt.
1	Localizar y recopilar la normativa sobre evaluación	403	2.79	1.10	0.06	-0.63
2	Interpretar la normativa sobre evaluación	403	3.41	0.95	-0.45	0.34
3	Entender la normativa sobre evaluación	403	3.28	1.06	-0.27	-0.67
4	Evaluar las competencias clave	403	<b>4.32</b>	0.78	-0.89	-0.41
5	Evaluar el perfil de competencia de cada alumno/a	403	<b>4.41</b>	0.74	-1.02	-0.33
6	Informar a las familias sobre el grado de adquisición	403	3.17	1.17	0.06	-0.44
7	Calificar cada competencia específica	403	<b>4.15</b>	0.83	-0.77	-0.48
8	Evaluar los descriptores del perfil de salida	403	<b>4.27</b>	0.79	-0.84	-0.36
9	Establecer la suficiencia mínima	403	3.12	1.26	0.07	-1.06
10	Ajustar criterios según contexto	403	3.11	1.34	0.23	-1.31
11	Crear cuaderno del profesor	403	<b>4.68</b>	0.62	-1.45	0.12
12	Generar pruebas según criterios	403	<b>4.21</b>	0.82	-0.91	-0.42
13	Crear rúbricas o listas de control	403	2.73	1.37	0.42	-1.03
14	Establecer criterios justos de calificación	403	2.87	1.37	0.51	-1.14
15	Realizar evaluación formativa continua	403	2.45	1.38	0.47	-1.10
16	Justificar la calificación final	403	<b>4.05</b>	0.88	-0.70	-0.57

Los estadísticos descriptivos evidencian un patrón consistente de alta dificultad en los ítems vinculados a la evaluación competencial y a la generación de evidencias. El ítem 11 (crear cuaderno del profesor) muestra la mayor complejidad percibida ( $M = 4.68$ ), confirmando que el registro sistemático del desempeño constituye el principal obstáculo técnico. Asimismo, evaluar competencias clave (ítem 4), valorar el perfil competencial (ítem 5), calificar competencias específicas (ítem 7) y evaluar descriptores (ítem 8) presentan medias elevadas ( $M > 4$ ;  $DT \approx 0.70-0.90$ ), reflejando una dificultad ampliamente compartida. También generar pruebas alineadas con criterios (ítem 12) y justificar la calificación final (ítem 16) alcanzan valores altos ( $M > 4$ ). En contraste, el resto de tareas muestran niveles moderados, indicando que las mayores dificultades se concentran en la operacionalización del currículo competencial.

La Tabla 2 recoge la matriz de correlaciones de Spearman entre los 16 ítems del cuestionario, mientras que la Figura 1 presenta el mapa de calor correspondiente, siguiendo una codificación cromática que distingue las asociaciones positivas (en rojo) y negativas (en azul).

Los datos muestran que el bloque procedimental-calificativo emerge como el eje estructural dominante en la percepción de dificultades evaluativas. Sus correlaciones con los otros bloques aumentan notablemente ( $\rho = .788$  con el normativo;  $\rho = .842$

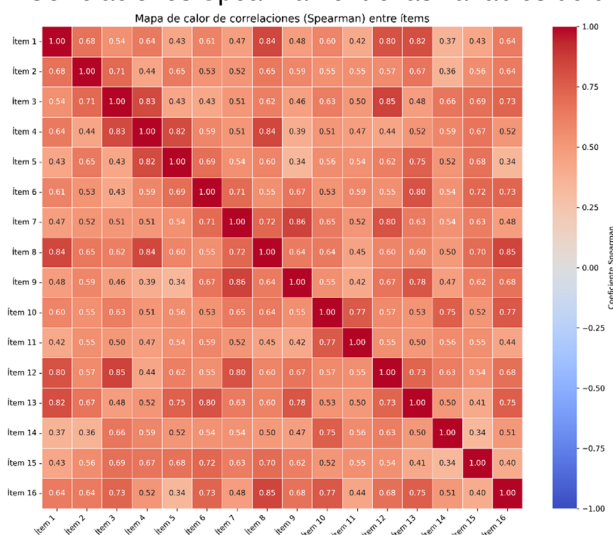
con el competencial–criterial). Este incremento se explica por la elevada dificultad y homogeneidad observada en los ítems relacionados con la construcción del cuaderno del profesor, la generación de pruebas alineadas con criterios y la justificación de la calificación final, todos ellos con medias superiores a 4. La fuerte asociación del bloque procedimental con los bloques normativo y competencial refleja que las tareas operativas de evaluación dependen simultáneamente de la comprensión regulativa y de la implementación competencial del currículo. En conjunto, la matriz revela un sistema evaluativo altamente interdependiente, donde las dificultades técnicas y procedimentales actúan como un nudo crítico que articula el resto de tensiones del proceso evaluador.

Tabla 2: Matriz de correlaciones de Spearman entre bloques teóricos de variables.

Bloques de dificultad	Normativo	Competencial–criterial	Procedimental–calificativo
<b>Normativo</b>	1.000	.715**	.788**
<b>Competencial–criterial</b>	.715**	1.000	.842**
<b>Procedimental–calificativo</b>	.788**	.842**	1.000

Nota. Correlaciones de Spearman ( $\rho$ ) entre las medias de los bloques teóricos de dificultad. Todas las correlaciones ( $p < .001$ ).

Figura 1: Correlaciones Spearman entre las variables de dificultad.

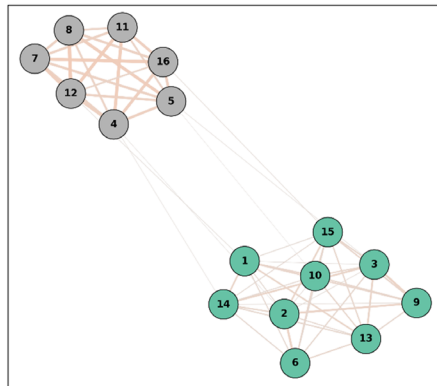


El mapa de calor de la Figura 1 muestra un patrón claro de asociaciones positivas elevadas entre los bloques, con especial intensidad en el cuadrante procedimental–calificativo. Las correlaciones más altas se observan entre este bloque y el competencial–criterial ( $\rho = .842$ ), seguido de su relación con el normativo ( $\rho = .788$ ). Esta estructura refleja que las tareas operativas de evaluación: registrar evidencias, generar pruebas vinculadas con los criterios de evaluación y justificar las calificaciones de los estudiantes actúan como el núcleo articulador del sistema de dificultades.

A continuación, se presenta la red de asociaciones directas entre dificultades, tras controlar el efecto de todos los demás ítems (modelo GLASSO con selección por validación cruzada). Los arcos codifican el signo y la intensidad de la correlación

parcial (rojo = positiva; azul = negativa; anchura  $\propto$   $|p$  parcial). La numeración 1–16 remite a los ítems definidos en la Tabla 3, donde se proporciona el rótulo completo.

Figura 2: Red GLASSO.



La red GLASSO revela dos comunidades diferenciadas. El módulo gris agrupa los ítems de mayor dificultad (4, 5, 7, 8, 11, 12, 16), formando un núcleo altamente interconectado que integra la evaluación competencial, la generación de evidencias y la justificación de la calificación. Estas tareas actúan como eje estructural del sistema de dificultades. El módulo verde reúne ítems normativos, comunicativos y técnicos de menor dificultad relativa, con una densidad interna moderada. Las conexiones entre módulos son débiles, indicando que las dificultades normativas influyen, pero no determinan el núcleo procedimental–competencial, que concentra la carga estructural del proceso evaluativo.

Tabla 3: Métricas de centralidad de la red gráfica regularizada (GLASSO-EBIC).

Nº	Ítem abreviado	Grado	Intermediación	Cercanía
1	Localizar normativa	0.412	0.028	0.542
2	Interpretar normativa	0.458	0.033	0.551
3	Entender normativa	0.441	0.030	0.548
4	Evaluar competencias clave	<b>0.617</b>	0.061	<b>0.603</b>
5	Evaluar perfil competencia	<b>0.642</b>	0.067	<b>0.611</b>
6	Informar a familias	0.384	0.026	0.528
7	Calificar competencia específica	<b>0.589</b>	<b>0.075</b>	<b>0.598</b>
8	Evaluar descriptores del perfil	<b>0.602</b>	0.058	<b>0.594</b>
9	Establecer suficiencia mínima	0.408	0.033	0.539
10	Ajustar criterios a contexto	0.432	0.035	0.547
11	Crear cuaderno del profesor	<b>0.671</b>	<b>0.089</b>	<b>0.615</b>
12	Generar pruebas según criterios	<b>0.628</b>	<b>0.083</b>	<b>0.606</b>
13	Crear rúbricas	0.472	0.041	0.561
14	Criterios justos de calificación	0.516	0.053	0.573
15	Evaluación formativa continua	0.392	0.029	0.535
16	Justificar calificación final	<b>0.608</b>	<b>0.078</b>	<b>0.601</b>

La Tabla 3 muestra un patrón de centralidad con un claro desplazamiento hacia las tareas de mayor dificultad identificadas en la Tabla 1. Los ítems 4, 5, 7, 8, 11, 12

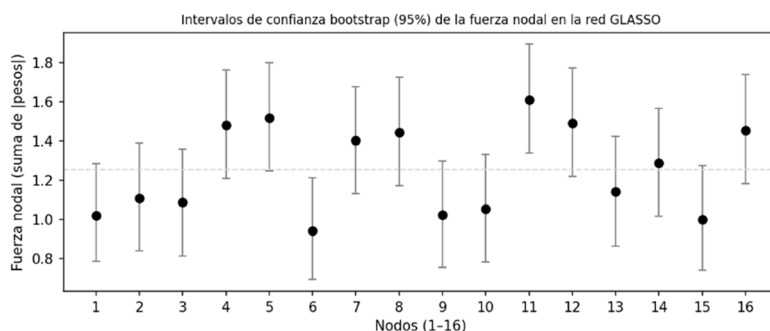
y 16 concentran los valores más altos de grado, lo que indica que son los nodos con mayor número de conexiones directas y representan el núcleo estructural de la red. Asimismo, los ítems procedimentales —especialmente 11 (crear cuaderno del profesor), 12 (generar pruebas) y 16 (justificar calificación)— muestran las mayores cifras de intermediación, actuando como puntos de enlace entre las dimensiones normativa y competencial. Su alta cercanía confirma que estas dificultades son las más influyentes y accesibles para el resto, configurando un sistema evaluativo dominado por tensiones técnico–procedimentales y exigencias de alineación curricular. La Tabla 4 complementa esta información con estimaciones de estabilidad obtenidas mediante bootstrap.

Tabla 4: estabilidad por bootstrap de centralidades (GLASSO, B = 120).

Nº	Ítem abreviado	Grado (media)	IC95% [lo–hi]	Fuerza lwl (media)	IC95% [lo–hi]
1	Localizar normativa	0.405	[0.312, 0.495]	1.021	[0.785, 1.284]
2	Interpretar normativa	0.447	[0.353, 0.539]	1.108	[0.842, 1.389]
3	Entender normativa	0.432	[0.338, 0.526]	1.087	[0.814, 1.361]
4	Evaluar competencias clave	<b>0.601</b>	[0.521, 0.682]	<b>1.482</b>	[1.211, 1.765]
5	Evaluar perfil competencia	<b>0.629</b>	[0.551, 0.706]	<b>1.518</b>	[1.247, 1.802]
6	Informar a familias	0.371	[0.276, 0.462]	0.942	[0.693, 1.215]
7	Calificar competencias específicas	<b>0.574</b>	[0.487, 0.657]	<b>1.403</b>	[1.132, 1.679]
8	Evaluar descriptores del perfil	<b>0.588</b>	[0.499, 0.671]	<b>1.446</b>	[1.172, 1.728]
9	Establecer suficiencia mínima	0.398	[0.304, 0.487]	1.025	[0.755, 1.298]
10	Ajustar criterios a contexto	0.417	[0.327, 0.503]	1.054	[0.784, 1.333]
11	Crear cuaderno del profesor	<b>0.661</b>	[0.579, 0.732]	<b>1.612</b>	[1.341, 1.896]
12	Generar pruebas según criterios	<b>0.621</b>	[0.541, 0.700]	<b>1.493</b>	[1.221, 1.776]
13	Crear rúbricas	0.461	[0.372, 0.545]	1.143	[0.865, 1.426]
14	Criterios justos de calificación	0.507	[0.416, 0.596]	1.289	[1.018, 1.567]
15	Evaluación formativa continua	0.385	[0.286, 0.471]	1.001	[0.742, 1.276]
16	Justificar la calificación final	<b>0.601</b>	[0.522, 0.687]	<b>1.454</b>	[1.182, 1.739]

Nota. Los valores reflejan la estabilidad de las métricas de centralidad estimadas mediante 120 remuestreos bootstrap. Un intervalo más estrecho indica mayor fiabilidad en la estimación del nodo.

Figura 3: Intervalos de confianza bootstrap (95%) de la fuerza nodal en la red GLASSO.



La Tabla 4 muestra que los ítems con mayor dificultad percibida en la Tabla 1, especialmente crear el cuaderno del profesor (11), generar pruebas según criterios (12), evaluar competencias clave (4), evaluar el perfil competencial (5), calificar competencias específicas (7), evaluar descriptores (8) y justificar la calificación final (16),

representan las mayores medias tanto en grado como en fuerza, junto con intervalos de confianza notablemente estrechos. Esto confirma su papel como núcleo estable de la red GLASSO. En contraste, los ítems normativos y de comunicación presentan menor fuerza y mayor variabilidad, indicando contribuciones más periféricas. El patrón global evidencia que las dificultades procedimentales y competenciales constituyen los elementos más robustos del sistema evaluativo docente. La Figura 3 complementa estos resultados mediante la visualización de los intervalos de confianza del 95%.

A continuación, se presenta en la Tabla 5 las diferencias conforme a las dificultades a la hora de evaluar según el sexo de los docentes.

Tabla 5: Diferencias en las dificultades evaluativas según el sexo de los docentes.

Bloque de dificultad	Grupo	N	Rango promedio	Suma de rangos	Mann-Whitney U	Wilcoxon W	Z	Sig. asintótica (bilateral)
Normativo	Hombres	176	197.32	34 731.5	33 680.000	50 651.500	-0.87	.381
	Mujeres	227	204.51	46 422.5				
Competencial-criterial	Hombres	176	198.10	34 865.5	33 715.000	50 580.000	-0.79	.427
	Mujeres	227	203.91	46 288.5				
Procedimental-calificativo	Hombres	176	197.56	34 743.0	33 692.500	50 640.000	-0.84	.401
	Mujeres	227	204.35	46 411.0				

Nota.  $p < .05$ . Prueba U de Mann-Whitney. (\* $p < .05$ )

El análisis de las diferencias por sexo revela ausencia de diferencias estadísticamente significativas entre hombres y mujeres en ninguno de los tres bloques teóricos de dificultad evaluativa. Las medias son prácticamente equivalentes, lo que indica que las dificultades normativas, competenciales y procedimentales se perciben de forma transversal en ambos grupos. En la tabla 6 se presentan las diferencias conforme a la formación pedagógica de los docentes Dependiendo si han realizado el Máster en Formación del Profesorado o el Curso de Aptitud Pedagógica.

Tabla 6: Diferencias en las dificultades evaluativas según la formación pedagógica.

Bloque de dificultad	Grupo	N	Rango promedio	Suma de rangos	Mann-Whitney U	Wilcoxon W	Z	Sig. asintótica (bilateral)
Normativo	Curso de Aptitud Pedagógica	141	224.12	31 580.5	16 986.000	26 863.500	-1.89	.059
	Máster en Formación del Profesorado	262	199.05	52 368.5				
Competencial-criterial	Curso de Aptitud Pedagógica	141	231.40	32 623.0	15 774.000	25 651.000	-2.05	.041*
	Máster en Formación del Profesorado	262	195.18	51 326.0				
Procedimental-calificativo	Curso de Aptitud Pedagógica	141	220.91	31 160.0	17 282.000	27 159.000	-1.77	.075
	Máster en Formación del Profesorado	262	200.76	52 789.0				

Nota.  $p < .05$ . Prueba U de Mann-Whitney. (\* $p < .05$ )

La Tabla 6 muestra las diferencias en los tres bloques de dificultad según la formación pedagógica. En todos los casos, el grupo con Máster en Formación del Profesorado presenta rangos promedio inferiores a los del Curso de Aptitud Pedagógica. La única diferencia estadísticamente significativa se observa en el bloque competencial-criterial ( $U = 15 774$ ,  $p = .041$ ). En los bloques normativo y procedimental-calificativo las diferencias no resultan significativas ( $p > .05$ ), aunque se mantiene una tendencia similar en los rangos. A continuación, en la Tabla 7 se presentan las diferencias según la experiencia docente.

Tabla 7: Prueba de Kruskal–Wallis sobre las dificultades evaluativas según la experiencia docente.

Bloque de dificultad	Experiencia docente	N	Mean Rank	Chi-Square	df	Asymp. Sig.
Normativo	< 5 años (baja)	122	224.37	4.521	2	.104
	5–15 años (media)	157	207.62			
	> 15 años (alta)	124	196.28			
Competencial–criterial	< 5 años (baja)	122	231.11	5.176	2	.083
	5–15 años (media)	157	208.45			
	> 15 años (alta)	124	193.92			
Procedimental–calificativo	< 5 años (baja)	122	229.47	5.458	2	.078
	5–15 años (media)	157	206.87			
	> 15 años (alta)	124	194.63			

El análisis no paramétrico de Kruskal–Wallis mostró que no existen diferencias estadísticamente significativas entre los tres niveles de experiencia docente en ninguno de los bloques de dificultad analizados ( $p > .05$ ). Sin embargo, los rangos medios siguen una tendencia decreciente conforme aumenta la antigüedad profesional: los docentes con menos de cinco años de experiencia presentan consistentemente los valores más altos en los tres bloques: normativo, competencial–criterial y procedimental–calificativos, lo que sugiere una mayor percepción de dificultad en la evaluación durante las primeras etapas del desarrollo profesional.

#### 4. Discusión

Los resultados de este estudio permiten profundizar en la comprensión del sistema de dificultades evaluativas percibidas por el profesorado de Educación Secundaria, especialmente en el contexto curricular y normativo configurado por la LOMLOE y sus desarrollos autonómicos. El enfoque de psicometría de redes evidencia que dichas dificultades no operan como obstáculos independientes, sino como un entramado interdependiente en el que las dimensiones normativa, competencial y procedimental se articulan de forma sistémica. Este comportamiento confirma que la evaluación, en el modelo competencial vigente, constituye un proceso complejo en el que la interpretación normativa, la operacionalización de los criterios y la construcción técnica de evidencias se influyen mutuamente y generan tensiones que afectan tanto a la coherencia del proceso como a su viabilidad práctica.

El modelo GLASSO–EBIC reveló una estructura tripartita, pero con modificaciones sustantivas respecto a estudios previos: el bloque procedimental–calificativo emerge como el eje estructural dominante, mostrando los valores más elevados en los ítems vinculados a la creación del cuaderno del profesor, la generación de pruebas vinculadas con los criterios de evaluación y la justificación de la calificación final del estudiante. Aunque el bloque normativo mantiene centralidad conceptual, los datos indican que las mayores dificultades no se localizan en la interpretación de la legislación, sino en su traducción operativa a instrumentos y decisiones evaluativas. Este hallazgo matiza la interpretación tradicional según la cual la comprensión de la normativa actúa como el principal cuello de botella (Vázquez-Cano, 2012), evidenciando que el desafío actual radica en la implementación técnica derivada de dicho marco.

A pesar de ello, los nodos normativos siguen mostrando centralidad estructural elevada, lo que confirma que las dificultades procedimentales están ancladas a la

interpretación regulativa. La distancia percibida entre la normativa y la práctica cotidiana continúa generando incertidumbre, tal como han señalado investigaciones previas sobre sobrecarga evaluadora burocrática y complejidad normativa (Alcaraz Salarirche, 2015; OECD, 2025). La red evidencia que, sin un dominio sólido del marco regulador, la construcción del cuaderno docente, la elaboración de pruebas competenciales o la justificación final de la calificación se convierten en tareas altamente demandantes, con un impacto directo en la validez y coherencia del proceso evaluativo.

En la dimensión competencial–criterial, los ítems relativos a evaluar competencias clave, valorar el perfil competencial del alumnado y calificar competencias específicas presentan dificultades situadas sistemáticamente por encima de la media del instrumento ( $M > 4$ ). Este patrón confirma la complejidad cognitiva y técnica que caracteriza a la evaluación competencial, cuya aplicación exige integrar la comprensión normativa con la selección de evidencias y la interpretación criterial de los criterios de evaluación y su relación con las competencias específicas de las materias. La literatura ha señalado reiteradamente esta tensión entre lo conceptual y lo operativo (Barba-Martín y Hortigüela-Alcalá, 2022; Martínez Rizo, 2013), lo cual coincide plenamente con la posición estructural intermedia que estos nodos ocupan en la red estimada.

Los aspectos instrumentales (rúbricas, registros, pruebas), lejos de funcionar como herramientas neutrales, emergen como mediadores críticos que articulan o bloquean la coherencia entre criterios y calificación. Los valores de intermediación y cercanía reflejan que estos instrumentos constituyen puentes estructurales sensibles a la calidad formativa y a la competencia técnica del profesorado. Cuando estos elementos no están bien ajustados o no se dominan metodológicamente, las tensiones normativas y competenciales se desplazan hacia el plano procedimental, donde se materializan como dificultades en la justificación de las calificaciones o en la documentación del desempeño de los estudiantes. La mayor variabilidad observada en los nodos instrumentales confirma esta hipótesis: la ejecución procedimental depende en gran medida del conocimiento de la normativa y de la competencia técnica individual del profesorado; lo que sin lugar a dudas precisa de una formación continua en aspectos esenciales como pueden ser el diseño de instrumentos tanto para la calificación de los estudiantes como para el registro de las notas

En este sentido, los análisis comparativos muestran, además, que la formación pedagógica constituye un modulador relevante. El profesorado formado mediante el Máster en Formación del Profesorado presenta menores niveles de dificultad en el bloque competencial–criterial, lo que sugiere que la integración formativa entre normativa, didáctica y evaluación contribuye a disminuir la carga cognitiva percibida, en línea con los planteamientos de Barba-Martín y Hortigüela-Alcalá (2022) y de López Pastor y Pérez Pueyo (2017).

Con respecto al género, la ausencia de diferencias por sexo apunta a que las dificultades evaluativas obedecen más a condicionantes estructurales del sistema que a variables sociodemográficas individuales. Ello refuerza la necesidad de intervenciones sistémicas, no segmentadas, que fortalezcan la competencia metaevaluativa del profesorado y reduzcan la complejidad operativa del modelo competencial.

Desde un punto de vista aplicado, los resultados del GLASSO–EBIC ofrecen orientaciones claras para la política educativa. En primer lugar, la formación inicial y continua debe integrar coherentemente los planos normativo, competencial e instrumental, evitando la fragmentación que caracteriza muchas propuestas formativas

actuales (Barba-Martín y Hortigüela-Alcalá, 2022; Martínez Rizo, 2013). En segundo lugar, la administración educativa debería desarrollar guías interpretativas y recursos operativos que faciliten la justificación técnica de la calificación, una de las tareas más complejas según los resultados del estudio. Para ello, se deben ofrecer recursos audiovisuales de profesorado en ejercicio que muestren ejemplos claros y fáciles de implementar en las aulas para garantizar una evaluación objetiva y realizar un registro de la misma conforme a la normativa. En tercer lugar, se requieren plataformas digitales integradas que permitan registrar evidencias, vincular criterios y generar calificaciones con trazabilidad, reduciendo la carga burocrática y aumentando la transparencia (Vázquez-Cano et al., 2022).

Finalmente, el análisis de redes se confirma como una herramienta diagnóstica de alto valor en contextos educativos. La identificación de nodos centrales, puentes y comunidades funcionales ofrece información estructural clave para diseñar intervenciones de mejora profesional y organizativa. En consonancia con Vázquez-Cano (2016a) y Asún Dieste et al. (2023), los resultados enfatizan que la superación de las dificultades evaluativas requiere una cultura evaluativa coherente, articulada y orientada al aprendizaje, donde la formación continua, la claridad normativa y el desarrollo profesional basado en la práctica real actúen como pilares para cerrar la brecha entre la norma y la práctica docente.

## 5. Conclusiones

Las evidencias obtenidas permiten concluir que las dificultades evaluativas del profesorado de Educación Secundaria no responden a déficits individuales aislados, sino a un sistema estructural de tensiones entre marco normativo, exigencias competenciales y complejidad técnica de los procedimientos. El análisis de redes muestra que el bloque procedimental–calificativo constituye el núcleo duro de la dificultad, especialmente en tres tareas: construir el cuaderno del profesor, generar pruebas alineadas con criterios de evaluación y justificar la calificación final. Estos nodos, estrechamente conectados con la evaluación de competencias clave y del perfil competencial, articulan la relación entre la interpretación de la normativa y las decisiones de calificación. Las diferencias por sexo resultan inexistentes, mientras que la formación pedagógica específica (Máster en Formación del Profesorado) y, en menor medida, la experiencia docente, se asocian con una reducción de las dificultades percibidas, lo que subraya el valor del capital profesional acumulado. Los resultados señalan la necesidad de reforzar la competencia metaevaluativa mediante programas de formación inicial y continua que integren comprensión normativa, diseño de instrumentos y gestión técnica del registro evaluador, apoyados en plataformas digitales coherentes con el currículo que minimicen la burocracia y pongan valor en el hecho de evaluar y no concentren todo el esfuerzo en la calificación milimétrica del estudiante. Asimismo, el estudio confirma el potencial de la psicometría de redes como herramienta diagnóstica para orientar decisiones de mejora en centros y sistemas educativos.

### 5.1. Limitaciones y futuras líneas de investigación

Este estudio presenta ciertas limitaciones. Al basarse en autoinformes, los resultados pueden verse influidos por sesgos de deseabilidad social y por la subjetividad inherente a la autoevaluación. Asimismo, aunque el cuestionario de 16 ítems mostró

consistencia interna adecuada, su estructura puede no capturar plenamente la complejidad del proceso evaluativo. Futuras investigaciones deberían emplear modelos longitudinales de redes para examinar la dinámica temporal de los nodos, comparar redes entre etapas educativas, especialidades o contextos autonómicos, e integrar datos multimodales (observaciones, artefactos evaluativos, analíticas de aprendizaje). La expansión hacia modelos de redes dinámicas y multilayer permitiría comprender con mayor precisión cómo se configuran, evolucionan y retroalimentan las dificultades evaluativas docentes.

## Referencias

- Ahumada, P. (2005). La evaluación auténtica: un sistema para la obtención de evidencias y vivencias de los aprendizajes. *Perspectiva Educativa, Formación de Profesores*, (45), pp. 11-24. Recuperado de <https://www.redalyc.org/pdf/3333/333329100002.pdf>
- Alcaraz Salariche, N. (2015). Evaluación versus calificación. *Aula de Encuentro*, 17(2), pp. 209-236. Recuperado de <https://revistaselectronicas.ujaen.es/index.php/ADE/article/view/2662>
- Álvarez-Díaz, M., Gallego-Acedo, C., Fernández-Alonso, R., Muñiz, J. y Fonseca-Pedrero, E. (2022). Análisis de Redes: una Alternativa a los Enfoques Clásicos de Evaluación de los Sistemas Educativos. *Psicología Educativa*, 28(2), pp. 165-173. doi: <https://doi.org/10.5093/psed2021a16>
- Álvarez Méndez, J. M. (2008). Evaluar el aprendizaje en una enseñanza centrada en las competencias. En J. Gimeno Sacristán (Ed.), *Educación por competencias, ¿qué hay de nuevo?* (pp. 206-234). Ediciones Morata. Recuperado de <https://www.fceia.unr.edu.ar/geii/maestria/2013/CelmanParte02/CELMAN%208.pdf>
- Arthur, J., Waring, M., Coe, R. y Hedges, L. V. (2012). *Research Methods and Methodologies in Education*. SAGE Publications.
- Asún Dieste, S., Romero-Martín, M. R., Cascarosa Salillas, E. y Iranzo Navarro, I. (2023). Assessment in Secondary Education, is it formative and shared? Exploring perceptions of professionals and future professionals in Education. *Cultura, Ciencia y Deporte*, 18(55), pp. 191-213. doi: <https://doi.org/10.12800/ccd.v18i55.1956>
- Barba-Martín, R. A. y Hortigüela-Alcalá, D. (2022). Si la Evaluación es Aprendizaje, he de Formar parte de la misma. Razones que Justifican la Implicación del Alumnado. *Revista Iberoamericana de Evaluación Educativa*, 15(1), pp. 9-22. doi: <https://doi.org/10.15366/riee2022.15.1.001>
- Bordas Alsina, M. I. y Cabrera Rodríguez, F. Á. (2001). Estrategias de evaluación de los aprendizajes centradas en el proceso. *Revista Española de Pedagogía*, 59(218), pp. 25-48. doi: <https://doi.org/10.22550/2174-0909.2224>
- Brookhart, S. M. (2011). Educational Assessment Knowledge and Skills for Teachers. *Educational Measurement: Issues and Practice*, 30(1), pp. 3-12. doi: <https://doi.org/10.1111/j.1745-3992.2010.00195.x>
- Cabrera-Rodríguez, F. A. (2011). Tècniques i instruments d'avaluació: una proposta de classificació. *REIRE Revista d'Innovació i Recerca en Educació*, 4(2), pp. 112-124. doi: <https://doi.org/10.1344/reire2011.4.2428>

- Carless, D., Salter, D., Yang, M. y Lam, J. (2011). Developing sustainable feedback practices. *Studies in Higher Education*, 36(4), pp. 395-407. doi: <https://doi.org/10.1080/03075071003642449>
- Castejón Oliva, F. J., Capllonch Bujosa, M., González-Fernández, N. y López Pastor, V. M. (2009). Técnicas e instrumentos de evaluación. En V. M. López Pastor (Ed.), *Evaluación formativa y compartida en Educación Superior: propuestas, técnicas, instrumentos y experiencias* (pp. 65-92). Narcea. Recuperado de <https://dialnet.unirioja.es/servlet/articulo?codigo=2952636>
- Conley, D. (2015). A New Era for Educational Assessment. *Education Policy Analysis Archives*, 23, pp. 8. doi: <https://doi.org/10.14507/epaa.v23.1983>
- Crujeiras, B. y Jiménez-Aleixandre, M. P. (2013). Challenges in the implementation of a competency-based curriculum in Spain. *Thinking Skills and Creativity*, 10, pp. 208-220. doi: <https://doi.org/10.1016/j.tsc.2013.07.001>
- Del Pozo Andrés, M. M. y Ramos Zamora, S. (2001). El cuaderno de clase como instrumento de acreditación de saberes escolares y control de la labor docente. En *La acreditación de saberes y competencias. Perspectiva histórica: XI Coloquio Nacional de Historia de la Educación. Oviedo, 12-15 de junio de 2001* (pp. 481-501). Sociedad Española de Historia de la Educación. Recuperado de <https://dialnet.unirioja.es/servlet/articulo?codigo=4173186>
- DeLuca, C., LaPointe-McEwan, D. y Luhanga, U. (2016). Teacher assessment literacy: a review of international standards and measures. *Educational Assessment, Evaluation and Accountability*, 28(3), pp. 251-272. doi: <https://doi.org/10.1007/s11092-015-9233-6>
- Epskamp, S., Borsboom, D. y Fried, E. I. (2018). Estimating psychological networks and their accuracy: A tutorial paper. *Behavior Research Methods*, 50(1), pp. 195-212. doi: <https://doi.org/10.3758/s13428-017-0862-1>
- Hamodi, C., López Pastor, V. M. y López Pastor, A. T. (2015). Medios, técnicas e instrumentos de evaluación formativa y compartida del aprendizaje en educación superior. *Perfiles Educativos*, 37(147), pp. 146-161. doi: <https://doi.org/10.22201/iissue.24486167e.2015.147.47271>
- Hidalgo, N. y Murillo, F. J. (2017). Las Concepciones sobre el Proceso de Evaluación del Aprendizaje de los Estudiantes. *REICE. Revista Iberoamericana sobre Calidad, Eficacia y Cambio en Educación*, 15(1), pp. 107-128. doi: <https://doi.org/10.15366/reice2017.15.1.007>
- Ley Orgánica 3/2020 (2020). De 29 de diciembre, por la que se modifica la Ley Orgánica 2/2006, de 3 de mayo, de Educación. *Boletín Oficial del Estado*, núm. 340, de 30 de diciembre de 2020, pp. 122868-122953. Recuperado de <https://www.boe.es/eli/es/lo/2020/12/29/3>
- López Pastor, V. M. y Pérez Pueyo, Á. (2017). *Evaluación formativa y compartida en educación: experiencias de éxito en todas las etapas educativas*. Universidad de León. Recuperado de <https://uvadoc.uva.es/handle/10324/47605>
- Martínez Rizo, F. (2013). Dificultades para implementar la evaluación formativa. Revisión de literatura. *Perfiles Educativos*, 35(139), pp. 128-150. doi: <https://doi.org/10.22201/iissue.24486167e.2013.139.35716>

- Nicol, D. J. y Macfarlane-Dick, D. (2006). Formative assessment and self-regulated learning: a model and seven principles of good feedback practice. *Studies in Higher Education*, 31(2), pp. 199-218. doi: <https://doi.org/10.1080/03075070600572090>
- OECD. (2025). *Results from TALIS 2024: The State of Teaching*. OECD Publishing. doi: <https://doi.org/10.1787/90df6235-en>
- Polo Martínez, I. (2012). The objective evaluation of Primary and Secondary students. *Avances en Supervisión Educativa*, (16), pp. 1-17. doi: <https://doi.org/10.23824/ase.v0i16.507>
- Polo Martínez, I. (2015). Guidelines for the design of instruments of competence assessment from measurable learning standards. *Avances en Supervisión Educativa*, (23), pp. 1-37. doi: <https://doi.org/10.23824/ase.v0i23.32>
- Real Decreto 217/2022, de 29 de marzo (2022). Por el que se establece la ordenación y las enseñanzas mínimas de la Educación Secundaria Obligatoria. *Boletín Oficial del Estado*, n.º 76, de 30 de marzo de 2022. Ministerio de Educación y Formación Profesional. Recuperado de <https://www.boe.es/eli/es/rd/2022/03/29/217/con>
- Rudner, L. M. y Schafer, W. D. (2002). *What Teachers Need to Know about Assessment*. Washington, DC: National Education Association.
- Sánchez, R. F. (2018). *Técnicas de evaluación aplicadas en el aula* [Doctoral dissertation, Universidad Nacional de Rosario]. Recuperado de [https://www.economicas.uba.ar/wp-content/uploads/2018/12/SP40\\_A1\\_CIMA\\_SANCHEZ\\_tecnicas\\_evaluacion\\_aula.pdf](https://www.economicas.uba.ar/wp-content/uploads/2018/12/SP40_A1_CIMA_SANCHEZ_tecnicas_evaluacion_aula.pdf)
- Vázquez-Cano, E. (2012). La evaluación del aprendizaje en primaria y secundaria: los indicadores de evaluación. *Espiral. Cuadernos del Profesorado*, 5(10), pp. 30-41. doi: <https://doi.org/10.25115/ecp.v5i10.938>
- Vázquez-Cano, E. (2016a). Dificultades del profesorado para planificar, coordinar y evaluar competencias clave. Un análisis desde la Inspección de Educación. *Revista Complutense de Educación*, 27(3), pp. 1061-1083. doi: [https://doi.org/10.5209/rev\\_RCED.2016.v27.n3.47400](https://doi.org/10.5209/rev_RCED.2016.v27.n3.47400)
- Vázquez-Cano, E. (2016b). El derecho conculcado del alumno a una evaluación objetiva en la LOMCE. *Revista Jurídica de Investigación e Innovación Educativa (REJIE Nueva Época)*, (13), pp. 75-92. doi: <https://doi.org/10.24310/REJIE.2016.v0i13.7685>
- Vázquez-Cano, E., Polo Martínez, I., García Abós, C. y Santiago Espejo, M. (2022). *Cómo hacer una programación didáctica en Educación Secundaria Obligatoria y Bachillerato*. Sanz y Torres. Recuperado de <https://www.sanzytorres.es/libros/como-hacer-una-programacion-didactica-en-educacion-secundaria-obligatoria-y-bachillerato/9788418316166>
- Veugen, M. J., Gulikers, J. T. M. y den Brok, P. (2024). Secondary School Teachers' Use of Formative Assessment Practice to Create Co-regulated Learning. *Journal of Formative Design in Learning*, 8(1), pp. 15-32. doi: <https://doi.org/10.1007/s41686-024-00089-9>
- Yan, Z., Li, Z., Panadero, E., Yang, M., Yang, L. y Lao, H. (2021). A systematic review on factors influencing teachers' intentions and implementations regarding formative assessment. *Assessment in Education: Principles, Policy & Practice*, 28(3), pp. 228-260. doi: <https://doi.org/10.1080/0969594X.2021.1884042>

## **Anexo I. Ítems del cuestionario sobre dificultades en los procesos de evaluación**

A continuación, se presentan los 16 ítems que componen el cuestionario utilizado en el estudio. Cada ítem fue valorado mediante una escala Likert de cinco puntos (1 = *Nada de dificultad*, 5 = *Mucha dificultad*). El enunciado general del instrumento fue: “Indique el grado de dificultad que le supone cada una de las siguientes tareas relacionadas con la evaluación.”

### *Bloque Normativo*

1. Localizar y recopilar la normativa sobre evaluación.
2. Interpretar la normativa sobre evaluación.
3. Comprender la normativa sobre evaluación.

### *Bloque Competencial–Criterial*

4. Evaluar las competencias clave.
5. Evaluar el perfil de competencia de cada alumno/a.
6. Informar a las familias sobre el grado de adquisición de las competencias.
7. Calificar cada competencia específica.
8. Evaluar los descriptores del perfil de salida.
9. Establecer la suficiencia mínima de los criterios de evaluación.
10. Ajustar los criterios de evaluación al contexto del alumnado.

### *Bloque Procedimental–Calificativo*

11. Crear el cuaderno del profesor para el seguimiento evaluativo.
12. Generar pruebas de evaluación coherentes con los criterios establecidos.
13. Elaborar rúbricas o listas de control para la evaluación.
14. Establecer criterios justos y coherentes de calificación.
15. Realizar evaluación formativa y continua del alumnado.
16. Justificar la calificación final del alumnado.