

AULAS INCLUSIVAS CONSTRUIDAS DESDE EL DIAGNÓSTICO

Inclusive classrooms built from the diagnosis

María JIMÉNEZ RUIZ

*Universidad de Valladolid. Facultad de Educación y Trabajo Social. Departamento de Pedagogía.
Campus Universitario Miguel Delibes. Paseo de Belén, 1. 47011 Valladolid
majiruiz@pdg.uva.es*

Benito ARIAS MARTÍNEZ

Universidad de Valladolid. Departamento de Psicología

Henar RODRÍGUEZ NAVARRO

Universidad de Valladolid. Departamento de Pedagogía

Jairo RODRÍGUEZ MEDINA

Universidad de Valladolid. Departamento de Pedagogía

Recepción: 22 de noviembre de 2017

Aceptación definitiva: 6 de febrero de 2018

RESUMEN: Hoy en día es común hablar de aulas inclusivas sin aportar evidencias del auténtico significado del término. En este trabajo nos proponemos presentar el proceso de desarrollo y validación de la escala ACOGE cuyo objetivo es evaluar la calidad de las aulas escolares para atender la diversidad desde un enfoque inclusivo. En el proceso de desarrollo y validación hemos utilizado procedimientos de análisis cuantitativo y cualitativo, en las siguientes fases: (1) construcción de un *pool* de ítems basados en el constructo teórico y en investigaciones previas; (2) consulta a jueces expertos en educación inclusiva; (3) organización de grupos de discusión con profesionales de la educación: directores de centro, maestros de educación infantil-primaria, de educación compensatoria y pedagogía terapéutica; (4) escala piloto; y (5) evidencias de validez y fiabilidad. Los resultados permitieron establecer un marco teórico de referencia desde el que construir la escala de acuerdo con los postulados ofrecidos por la literatura y la experiencia aportada por los profesionales de la educación. Proporcionamos evidencias de validez de contenido basadas en el constructo Inclusión Educativa mediante indicadores

observables asociados a inclusión y calidad y se aportan orientaciones sobre la utilidad de la escala para los docentes.

PALABRAS CLAVE: calidad educativa; atención a la diversidad; evaluación; inclusión educativa; escala.

ABSTRACT: Nowadays it is common to speak of inclusive classrooms without providing evidence of the true meaning of the term. In this paper we propose to present the process of development and validation of the ACOGE scale whose objective is to evaluate the quality of school classrooms to address diversity from an inclusive approach. In the development process we have used quantitative and qualitative analysis procedures in the following phases: (1) construction of a pool of items based on the theoretical construct and previous research; (2) consultation with expert judges in inclusive education; (3) organization of discussion groups with education professionals: center directors, teachers of early childhood education, compensatory education and therapeutic pedagogy; and (4) pilot scale. The results allowed establishing a theoretical frame of reference from which to build the scale according to the postulates offered by the literature and the experience contributed by the professionals of the education. We provide evidence of content validity based on the Educational Inclusion construct through observable indicators associated with quality inclusion and provide guidance on the usefulness of the scale for teachers.

KEY WORDS: educational quality; attention to diversity; evaluation; educational inclusion; scale.

1. Introducción

LA EDUCACIÓN INCLUSIVA SE CONCEPTUALIZA dentro de un marco teórico en el que se proponen cuatro esferas teóricas que interactúan entre sí: *políticas, conceptos, estructuras y prácticas* (Ainscow, Booth y Dyson, 2006). La interacción entre dichas esferas determina la eficacia de la inclusión educativa. Ainscow *et al.* (2006) entienden este constructo como un proceso sistémico de innovación educativa que permita la presencia, el rendimiento y la participación de todo el alumnado escolarizado en cualquier centro. Para ello plantean la necesidad de detectar y eliminar las barreras que limitan dicho proceso. Las esferas teóricas del modelo de Ainscow *et al.* (2006) se operativizan en tres variables escolares relevantes: *presencia, progreso y participación* (ver Figura 1). La *presencia* se refiere a los lugares en los cuales están escolarizados los estudiantes. Da respuesta a la disyuntiva entre los centros normalizados, compartidos entre iguales y las escuelas específicas o segregadas para alumnado con necesidades educativas especiales; desde este punto de vista la inclusión supone una palanca para que los centros ordinarios puedan mejorar sus prácticas educativas desde el reconocimiento de la diversidad humana (Echeita, 2013). La variable *progreso* comprende las medidas que el centro adopta para proporcionar experiencias de calidad a todos los estudiantes y optimizar su rendimiento escolar en función de sus

características y necesidades, así como para garantizar el mayor nivel de aprendizaje significativo en todas las competencias establecidas en el currículo (Miras, 2001). Por último, la *participación* implica aprender con otros y colaborar con ellos en el transcurso de las clases y las lecciones. Supone una implicación activa con quien se está aprendiendo y con quien enseña. Requiere ser reconocido y aceptado por lo que uno es, considerando su bienestar personal: autoestima, compañerismo y ausencia de exclusión (Ainscow *et al.*, 2006; Echeita, Verdugo, Sandoval, Calvo y González-Gil, 2009). La participación implica las experiencias de la persona en la escuela e incluye los sentimientos y la valoración del bienestar personal y social. Tal y como plantean Florian, Rouse y Black-Hawkins (2011), tienen que ver con el sentimiento activo de “unirse a” que incluye responsabilidad recíproca y reconocimiento como derecho compartido por todos. Se refiere también a todos los aspectos escolares relacionados con la participación familiar, agentes sociales, antiguos estudiantes, entidades religiosas y barrio.

Figura 1. Variables escolares relevantes dentro del marco teórico de Inclusión Educativa (Booth y Ainscow, 2002, 2011)



Echeita y Ainscow (2011) defienden la necesidad de ofrecer un concepto que pueda ser empleado por toda la comunidad educativa, que supere la confusión existente sobre el término. Un aspecto fundamental que necesita de una mayor indagación (Echeita, Muñoz, Sandoval y Simón, 2014) es cómo promover dentro del profesorado los valores y principios éticos que sostienen, motivan y fortalecen la transformación hacia la inclusión, lo que Booth llama *alfabetización ética*. Para todos ellos, es imprescindible la reflexión sobre sus propias culturas y prácticas escolares. De ahí la necesidad de contar con instrumentos que faciliten esta reflexión y aporten pistas eficaces para mejorar la calidad educativa.

Cuando se ven frustrados nuestros esfuerzos por realizar cambio de acuerdo con nuestros valores, la acción con principios es nuestra propia recompensa y, de esa forma,

la laboriosa tarea de poner en relación los valores de la inclusión con las acciones, mantiene vivo un recurso para actuar de forma diferente (Booth, 2006: 217).

Consideramos en este sentido que una evaluación de carácter formativo y funcional nos permite promover una mejora continuada que repercuta en la calidad educativa (Casanova, 2004, 2011). El Informe TALIS (OCDE, 2009) revela que el papel de la evaluación ha cambiado. Tradicionalmente se centraba en el control para garantizar las políticas y las leyes, por lo que se ocupaba exclusivamente de aspectos administrativos. En la actualidad este mismo informe recoge la importancia y el valor que tienen los sistemas de autoevaluación del profesorado y los centros. Este nuevo enfoque nos permite establecer la bidireccionalidad entre *calidad* y *evaluación*: para la calidad, la evaluación es un elemento imprescindible y, a la vez, justificamos la evaluación como garantía de calidad (De la Orden, 2009). Sin duda, debemos enfocar la evaluación hacia el establecimiento de programas de mejora escolar que propicien reformas específicas en los centros centradas en aquellas variables que tengan una relación sistemática con la eficacia. Queremos destacar con esta idea la importancia de la evaluación “como medio para mejorar los centros, incrementar su eficacia y, en último término, lograr una educación de calidad” (Fernández y González, 1997: 3).

Como resumen de este pensamiento, adoptamos la frase de Mel Ainscow “lo que se puede medir y evaluar se puede conseguir” (Simón y Echeita, 2013: 24).

En este sentido, la literatura ofrece varios instrumentos creados para tal propósito: (a) Desarrollo de indicadores sobre Educación Inclusiva (Agencia Europea para el Desarrollo de la Educación del Alumnado con Necesidades Educativas Especiales, 2009); (b) Proyecto IQEA. Improving the Quality of Education for All – Mejorar la Calidad de la Educación para Todos (Ainscow, Hopkins, Southworth y West, 2001); (c) Guía INTER para la aplicación de la educación intercultural en la escuela (Interproject, 2005); (d) Cuestionarios “Calidad educativa y atención a la diversidad” (Casar, 2007); (e) Modelo para evaluar la calidad de la respuesta de la escuela a la diversidad (Duk, 2007); (f) Guía para la reflexión y valoración de prácticas inclusivas OEI (Marchesi, Durán, Giné y Hernández, 2009); (g) AVACO-EVADIE: Cuestionario para la evaluación de la atención a la diversidad como dimensión educativa en las instituciones escolares (Biencinto, González, García, Sánchez y Madrid, 2009); (h) IAQV: Instrumento de evaluación de calidad de vida; Grupo de investigación GREID (Muntaner, Forteza, Roselló, Verger e Iglesia, 2010); (i) ACADI. Autoevaluación de Centros Educativos para la Atención a la Diversidad desde una Perspectiva Inclusiva (Arnáiz y Guirao, 2015); (j) Index for Inclusion (Booth y Ainscow, 2002, 2011).

Tras el análisis de dichos instrumentos hemos llegado a una serie de razones que justifican la construcción de la Escala ACOGE: (a) los ítems de la mayoría de las escalas citadas recogen aspectos teóricos de difícil extrapolación a un contexto aplicado; es por esto que su utilidad en la práctica profesional es cuestionable; (b) consideramos que ciertos ítems tienen un alto índice de deseabilidad social en su formulación (p. e., *Todo el mundo merece sentirse acogido; Los niños son valorados por igual; Las opiniones del alumnado con NEE son consideradas del mismo modo que las de sus compañeros/as; El centro se esfuerza por disminuir las prácticas discriminatorias*); (c) existe un desequilibrio en cuanto a la carga conceptual de algunas dimensiones respecto a

otras; y (d) en la ciencia es necesario verificar los supuestos auxiliares que acompañan a cualquier modelo explicativo de la realidad, especialmente a aquellos relacionados con la validez y precisión de los instrumentos de medida. Por todo ello, constatamos que no existe una herramienta que permita llevar a cabo un proceso de evaluación formativa y funcional, por lo que consideramos necesario proponer la construcción de un instrumento que permita evaluar la calidad y eficacia educativa de la atención a la diversidad del alumnado desde un enfoque inclusivo.

2. Método

2.1. Participantes

En la primera consulta realizada participaron siete jueces: cinco profesores de universidad y dos expertos en Inclusión Educativa. El estudio piloto se realizó con 78 estudiantes universitarios de 2.º curso del Grado de Educación Infantil y Primaria de la Facultad de Educación y Trabajo Social de la Universidad de Valladolid, cuyos contenidos del Plan de Estudios hacen referencia a la Educación Inclusiva en al menos tres asignaturas cursadas. En los grupos de discusión participaron 24 profesionales españoles con un mínimo de cinco años de experiencia profesional de centros escolares de educación infantil y primaria. De ellos, 12 pertenecían a los equipos directivos y 12 a profesorado (tutores, profesores de Educación Compensatoria, Pedagogía Terapéutica y Audición y Lenguaje, profesionales de Equipos Psicopedagógicos y Departamentos de Orientación).

2.2. Instrumento

La Escala ACOGE evalúa aspectos observables relacionados con las tres variables relevantes que conforman el constructo Educación Inclusiva: *presencia, participación y progreso* (Ainscow *et al.*, 2006).

El cuestionario consta de 54 ítems que definen aspectos observables, concretos y fáciles de entender relacionados con la inclusión educativa. Están enunciados con formato declarativo en tercera persona y ordenados en las tres dimensiones que sustentan el constructo teórico de la Educación Inclusiva: (a) *Presencia*, se refiere a la localización en la que son educados los alumnos, centros, aulas, espacios comunes, compartidos, iguales para todos o centros, aulas, espacios específicos, segregados/separados; (b) *Participación*, consiste en aprender con otros y colaborar con ellos en el transcurso de las clases y las lecciones, supone la implicación activa con lo que se está aprendiendo (y enseñando), así como la relación y contribución de toda la comunidad educativa en el proceso de enseñanza-aprendizaje; y (c) *Progreso*, hace referencia a la preocupación por garantizar el mayor nivel de aprendizaje significativo en todas las competencias curriculares que asegure un avance constante (rendimiento) de cada alumno en función de sus necesidades.

El formato de respuesta elegido es doble, primero, se solicita al respondiente que valore el *grado de importancia* que tiene cada afirmación en la conceptualización de Educación Inclusiva mediante una *escala Likert* de cuatro puntos: *poco importante, algo importante, bastante importante, muy importante*; en segundo lugar, se pide que evalúe el grado de cumplimiento real del contenido del ítem en su práctica docente, con otra escala Likert de cuatro puntos: *poco cumplimiento, algo de cumplimiento, bastante cumplimiento, mucho cumplimiento*.

Previo al cuestionario, la escala incluye unas breves explicaciones y una ficha con datos sociodemográficos, donde se recogen una serie de datos referidos a la persona y el contexto educativo en el que desarrolla su labor docente: nombre del centro y la ciudad; tipo de centro (público, concertado, privado); ubicación del centro (rural, rural agrupado, urbano); línea del centro (línea 1, 2, 3); tipo de profesor (interino, funcionario de carrera); perfil profesional (equipo directivo, tutora de aula, pedagogía terapéutica, audición y lenguaje, compensatoria, especialistas, otro); entidades sociales con las que colabora el centro; género de la persona que rellena la encuesta; edad; años de experiencia; etapa donde imparte docencia (infantil, primaria); cuántas actividades formativas realiza por término medio en un año; actividades formativas a las que ha acudido en los tres últimos años; tipo de discapacidades presentes en el alumnado de su aula; número de alumnado con NEAE escolarizados en su aula; número de alumnado con desfase curricular.

2.3. Procedimiento

Previa solicitud a la Dirección Provincial de Valladolid y a la Consejería de Educación de la Junta de Castilla y León, y concesión del permiso necesario para realizar la investigación, el equipo investigador se puso en contacto, en primer lugar con los centros que habían participado en los grupos de discusión. Mediante contacto telefónico se les recordó el compromiso para colaborar en la validación de la escala; se les informó del proceso que se estaba llevando a cabo y se concretó el papel que debían desempeñar en el centro que dirigían para lograr la máxima participación del claustro de profesores. Una vez confirmada la disponibilidad para seguir participando en la investigación, se les remitió, mediante correo electrónico, el enlace de acceso de la escala ACOGE y un documento con información sencilla y relevante sobre la investigación para que pudiesen transmitir a su equipo una explicación convincente que implicase a los docentes a la participación. Realizamos una segunda ronda en la que contactamos con directores de diferentes centros de educación de infantil y primaria de la provincia de Valladolid, dando a conocer el proyecto y solicitando su participación en él. Les ofrecimos acudir a los centros a informar sobre la necesidad de llevar a cabo procesos inclusivos en las aulas, aportándoles materiales y prácticas basadas en evidencias contrastadas científicamente, recogidas en nuestro estudio.

3. Resultados

La construcción de la escala ACOGE se desarrolló a lo largo de las siguientes fases: 1) recopilación de un *pool* de ítems a partir de la literatura previa sobre inclusión educativa; 2) consulta a jueces expertos en educación inclusiva; 3) grupos de discusión con profesionales pertenecientes a equipos directivos, educación compensatoria, audición y lenguaje, pedagogía terapéutica y profesores tutores de las aulas; 4) escala piloto; y 5) análisis psicométrico de los datos para demostrar la fiabilidad y validez de la medida de los ítems constitutivos de la escala.

3.1. Fase 1. Construcción de un *pool* de indicadores e ítems observables

Una vez definido el objetivo y propósito de la escala, se estudió el contenido y propiedades de los instrumentos que hacen referencia al constructo de *Educación Inclusiva* (ver Sección de introducción). A partir de la información extraída, se formularon indicadores observables en forma de ítems para representar el constructo de la forma más precisa posible. Centramos nuestra atención en el contexto concreto del *aula escolar*, espacio donde el profesorado realiza su labor diaria. Para el desarrollo del primer *pool* de ítems, se partió del modelo de Booth y Ainscow (2006), que consta de cuatro esferas teóricas de evaluación (*Culturas, Estructuras, Políticas, Prácticas*) operacionalizadas en tres dimensiones (*Presencia, Progreso y Participación*). Se formularon 223 ítems extraídos de una exhaustiva revisión de fuentes secundarias donde se profundizó en el tema de la Inclusión Educativa, así como los instrumentos utilizados para su evaluación. Dicha revisión se ha concretado a través de artículos indexados en la base de datos ISI Web of Knowledge, priorizando estudios publicados en JCR, bajo los descriptores *inclusion education, assesment, mainstream school, educational quality, attention to diversity, educational inclusion, scale*.

3.2. Fase 2. Consulta a los jueces expertos

Para obtener evidencias de validez de contenido de los ítems, se realizaron consultas en dos ocasiones a siete jueces expertos. Se le encomendó la tarea de calificar los 223 ítems del *pool* inicial en cuanto (a) la dimensión teórica de pertenencia –*presencia, participación, progreso*–; (b) la *dificultad* que suponía cada ítem para su implementación en contextos de *Educación Inclusiva*; (c) el grado de *importancia* o relevancia de cada ítem para favorecer la inclusión; (d) el grado de *observabilidad* del ítem por un evaluador externo; y (e) la *sensibilidad* de cada ítem a los previsibles cambios en el caso de implementar programas de mejora de la inclusión. De forma adicional, se pidió a los jueces que propusieran nuevos ítems representativos que evaluaran la calidad de los contextos educativos para atender a la diversidad.

Una vez codificadas las respuestas de los jueces, se estimó el coeficiente de concordancia entre jueces y la calificación promedio. La concordancia se calculó mediante

los coeficientes alfa de Krippendorff (1970), Kappa de Cohen (1960), Índice de Perreault (Perreault y Leigh, 1989) y S de Bennett (Bennett, Alpert y Goldstein, 1954). Se calculó también el coeficiente de concordancia B_N de Bangdiwala para datos nominales (*dimensión*) y los coeficientes ponderados de B_N^W de Bangdiwala (1987) para datos ordinales (*dificultad, importancia, observabilidad, sensibilidad*). El coeficiente B_N informa únicamente de los acuerdos estrictos (p. e., los jueces asignan la misma puntuación al ítem). Con la versión ponderada, se tienen en cuenta los acuerdos parciales (p. e., en qué medida los jueces proporcionan puntuaciones más o menos alejadas del acuerdo estricto). Estos coeficientes oscilan entre “0” (acuerdo nulo) y “1” (acuerdo total). Para interpretar los coeficientes de concordancia se siguió el siguiente criterio (Bangdiwala, 1987): valores comprendidos entre .000 y .200 son indicadores de pobre nivel de acuerdo; entre .201 y .400, débil; entre .401 y .600, moderado; entre .601 y .800, bueno; superior a .801, excelente. La mayoría de las investigaciones consideran aceptables los coeficientes superiores a .400.

En el caso de la *dimensión*, se pidió a los jueces que clasificaran cada ítem en una de tres dimensiones (p. e., *Presencia, Participación, Progreso*). Partimos de la hipótesis de que tales categorías constituyen un conjunto exclusivo (p. e., no existen solapamientos entre ellas) y exhaustivo (p. e., cubren todo el recorrido teórico de la variable latente que se somete a estudio). Se esperaba, en consecuencia, que los jueces ubicaran cada uno de los ítems en una u otra de las tres dimensiones, pero no en más de una. De hacerlo así, ello mostraría que el ítem es confuso y debería en consecuencia eliminarse. El acuerdo completo se consigue cuando los siete jueces clasifican el ítem en la misma dimensión. Habrá desacuerdo (de mayor o menor magnitud) cuando la clasificación del ítem por los jueces tenga más discrepancias. Para analizar el acuerdo entre los jueces en las dimensiones de adscripción, se calcularon los coeficientes de B_N (Bangdiwala, 1987). Por último, para reafirmar las evidencias de validez de contenido de la escala se priorizaron, en la selección final de los ítems, aquellos a los que los jueces hubieran otorgado puntuaciones medias más elevadas en importancia y observabilidad. Se eliminaron de la versión final los ítems con varianzas más elevadas, así como aquellos en cuya calificación los jueces mostraron mayores discrepancias.

Se presentan los resultados referidos al grado de acuerdo y desacuerdo (Tabla 1) manifestado por los jueces en torno a las tres dimensiones teóricas en las que se basa la escala y conforme a los cuatro parámetros citados en los análisis de los datos (*Dificultad, Importancia, Observabilidad y Sensibilidad*).

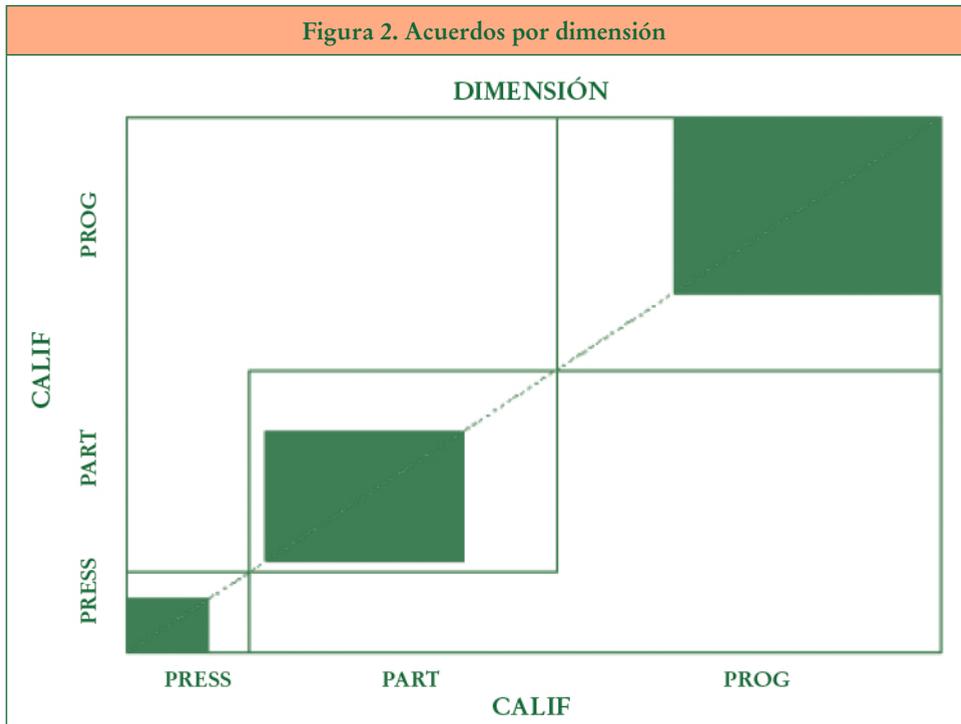
Tabla 1. Medias en B_N por dimensión, dificultad, importancia, observabilidad y sensibilidad					
	<i>Dimensión</i>	<i>Dificultad</i>	<i>Importancia</i>	<i>Observabilidad</i>	<i>Sensibilidad</i>
B_N	.457	—	—	—	—
B_N^W	—	.501	.826	.795	.842

El índice de acuerdo B_N de Bangdiwala proporciona una representación gráfica del grado de acuerdo de una tabla de contingencia y proporciona una medida de la fuerza de ese acuerdo (Tabla 2). Naturalmente, en la práctica se constatan discrepancias

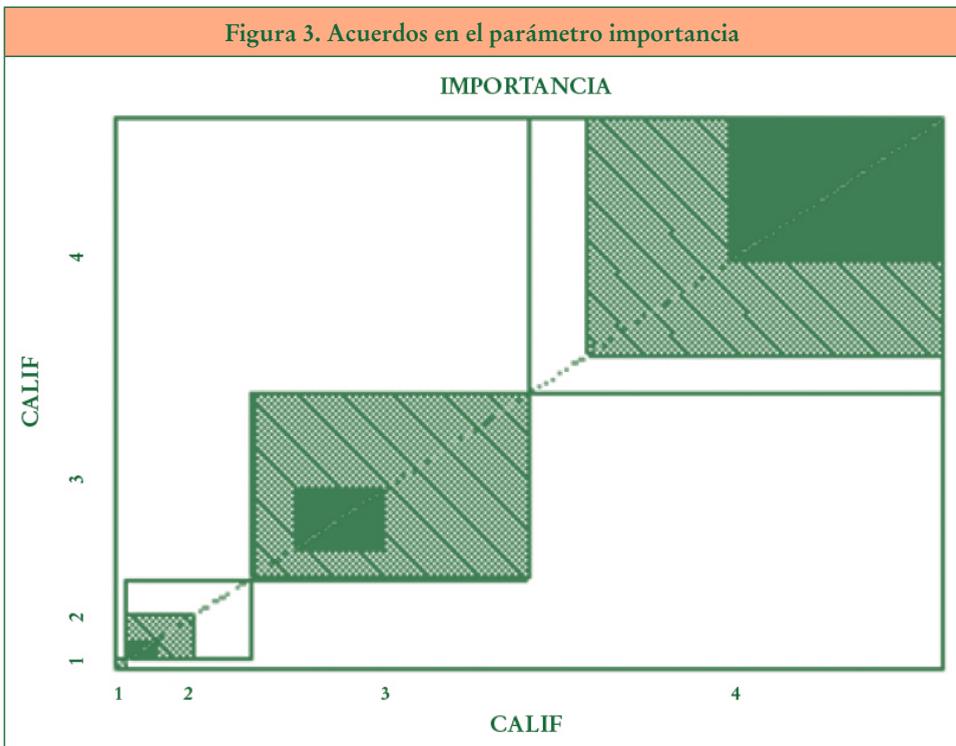
entre los jueces. El número de acuerdos queda representado en la diagonal. Se espera que estas cantidades sean superiores a las que están fuera de la diagonal.

Tabla 2. Matriz de acuerdos observada (Dimensión)				
	<i>Presencia</i>	<i>Participación</i>	<i>Progreso</i>	<i>Total</i>
<i>Presencia</i>	157			
<i>Participación</i>	32	381		
<i>Progreso</i>	46	178	512	
Total	235	591	736	1562

Como puede observarse, de las 1562 elecciones realizadas en conjunto por los 7 jueces en 157 ocasiones se ha coincidido en calificar a los ítems como pertenecientes a la dimensión *presencia*; en 32 ocasiones el ítem fue calificado simultáneamente en *presencia* y *participación*; en 381 ocasiones se ha coincidido en su clasificación en *participación*; en 56 ocasiones la coincidencia se ha producido en *presencia* y *progreso*; en 178 ocasiones, se han clasificado conjuntamente en *progreso* y *participación*. Finalmente, en 512 ocasiones se ha coincidido en clasificar los ítems en la categoría *progreso*, resultando el gráfico representado en la Figura 2.



En la Tabla 2 se advierte un desequilibrio en el número total de veces que se han elegido los diferentes criterios. Así, el dominio *presencia* cuenta con 235 apariciones, *participación* cuenta con 591 y *progreso* con 736. Si examinamos los diferentes rectángulos, vemos que el número de acuerdos en cada uno guarda una proporción similar con respecto al número de desacuerdos en esa categoría. En el caso de *presencia* ha sido del 67% (p. e., 157/235); en el de *participación*, del 64%; y en *progreso*, del 69%. En consecuencia, el conjunto de jueces ha logrado superar la mitad de la proporción de acuerdo posible. El coeficiente de acuerdo obtenido es $B_N = .45$. El resto de los análisis de concordancia se han llevado a cabo sobre las puntuaciones (ordinales) que los jueces han dado a los ítems atendiendo al grado en que se cumplían diferentes criterios (p. e., dificultad, importancia, observabilidad y sensibilidad). Se ha seguido el mismo procedimiento, pero, en este caso, al tratarse de una variable ordinal tomamos el coeficiente ponderado entre jueces (B_N^w) para la representación gráfica, y se han calculado los mismos coeficientes de acuerdo que en el caso anterior. Tomamos, para ilustrar este caso, las puntuaciones obtenidas en el parámetro *importancia* (Figura 3).



En las Figuras 2, 3 y 4 los cuadros totalmente coloreados muestran el acuerdo original (acuerdo observado). Estos están situados dentro de unos cuadros mayores

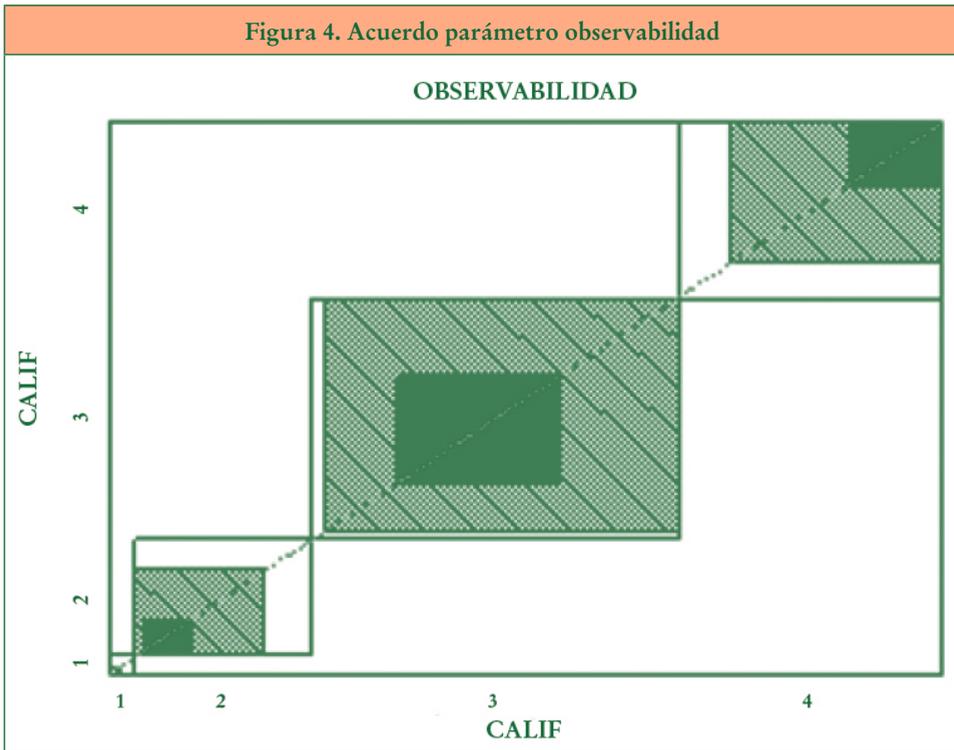
sombreados, que representan el acuerdo parcial. Ambos cuadros están inscritos en otro rectángulo de mayores dimensiones que nos muestra gráficamente el acuerdo máximo posible. La zona blanca de cada rectángulo constituiría la representación gráfica del desacuerdo.

Los resultados obtenidos (Tabla 3) revelan que la categoría 1 (“muy poco importante”) es elegida tan solo un 1% de las veces por los jueces (24 veces en un total de 1562 elecciones), la categoría 2 (“algo importante”) alcanzó el 14% de elecciones; aumenta considerablemente el porcentaje de elección de la categoría 3 (“bastante importante”) con un 34% y finalmente la categoría 4 (“muy importante”) ha sido seleccionada en el 50% de las ocasiones.

Tabla 3. Matriz de elecciones de cada categoría en el parámetro importancia					
	1	2	3	4	Total
1	1				
2	5	47			
3	9	75	171		
4	9	103	270	401	
Total	24	230	525	783	1562

En cuanto a discrepancias (Figura 3), la categoría en la que se advierten mayor desacuerdo es la 2 (“algo importante”). La zona blanca del rectángulo (desacuerdos) es mayor que la de acuerdos brutos y ponderados conjuntamente; en la categoría 3 (“bastante importante”), si se consideran conjuntamente los acuerdos brutos y los ponderados, podemos decir que el acuerdo es prácticamente unánime. El grado de discrepancia es algo más señalado en la categoría 4 (“muy importante”), donde se denota un relativo incremento de la zona blanca. En conjunto, la concordancia entre los jueces es muy aceptable si consideramos conjuntamente los acuerdos brutos y los ponderados ($B^w_N = .826$). En lo que atañe, pues, a la *importancia*, debemos considerar los resultados de los jueces plenamente satisfactorios. Pasamos a analizar estos mismos resultados en el parámetro de *observabilidad* de los ítems. Para ello haremos referencia a la Tabla 4 y a la Figura 4.

Tabla 4. Matriz de elecciones de los jueces por categoría en el parámetro de observabilidad					
	1	2	3	4	Total
1	2				
2	17	85			
3	21	139	304		
4	10	88	221	176	
Total	50	329	865	495	1562



El examen de los resultados en observabilidad indica que más del 50% de las elecciones corresponden a la categoría 3 (“bastante importante”), coincidiendo con menor número de discrepancias entre jueces dentro de la misma categoría. La categoría 4 (“muy importante”) fue seleccionada el 32% de las ocasiones, seguida por la categoría 2 (“algo importante”) con un 21% de elecciones y de forma casi insignificante la categoría 1 (“muy poco importante”) con un porcentaje del 3%. Tanto en el parámetro de importancia como en el de observabilidad los porcentajes de elección de la categoría 3 y 4 de forma conjunta superaron el 80%. Si analizamos las discrepancias por categorías, podemos decir que en la categoría 3, sumando los acuerdos brutos y ponderados, la diferencia fue insignificante, por lo que el acuerdo de los jueces fue prácticamente unánime. No fueron tan coincidentes en las categorías 2 y 4, donde advertimos unas diferencias mayores entre la suma de acuerdos brutos y ponderados y el acuerdo total, representado en la Figura 4 por las zonas en blanco en ambas categorías. Los datos aportados en la concordancia entre jueces fueron aún más modestos en el parámetro de observabilidad que en el de importancia, siendo aceptable el índice que considera conjuntamente los acuerdos brutos y ponderados ($B^w_N = .795$).

3.3. Fase 3. Grupos de discusión

En la fase 3 se realizó un Grupo de discusión (Merton y Kendall, 1946; Krueger, 1991; Krueger y Casey, 2000), donde participaron 24 profesionales españoles. Se estableció un contacto previo telefónico y escrito con los representantes de los profesionales de los centros escolares implicados en la investigación. Una vez aceptaron colaborar, se pidió a los participantes que: (a) realizaran una lectura de todos los ítems; (b) remitieran los ítems seleccionados como más importantes para evaluar la calidad de la inclusión en las aulas; y (c) invitaran a dos personas de sus centros a acudir al grupo de discusión en la Universidad.

Los participantes se dividieron en dos subgrupos. Se les pidió que realizaran una definición de Educación Inclusiva y un análisis cualitativo de los ítems a partir de la pregunta ¿en qué medida considera que se ajustan los ítems a la realidad educativa que conoce? Participaron 12 personas en cada subgrupo, guiadas por un moderador experto, encargado de encauzar las discusiones y opiniones para conocer cómo pensaban los participantes respecto a los ítems de la escala de Inclusión Educativa. Las respuestas de los participantes guiaron la selección de los ítems entre aquellos que fueron calificados como más importantes y menos importantes. Estos ítems sirvieron de hilo conductor en la discusión llevada a cabo en los grupos.

Los análisis cualitativos se realizaron conforme a las aportaciones ofrecidas de cada uno de los ítems considerados más importantes en el análisis, es decir, aquellos que obtuvieron una puntuación por encima de .2 (en una escala teórica de -1 a +1) según la valoración conjunta realizada. Cada uno de estos ítems seleccionados fue sometido a las siguientes preguntas para su discusión: (a) ¿Por qué cree que este ítem refleja su definición de Educación Inclusiva?; (b) ¿Cómo cree que se formularía mejor para que fuese observable y evaluable?; y (c) ¿Piensa que falta algún ítem para medir esa dimensión? Para finalizar la sesión se pidió a los participantes que sugiriesen pautas de intervención adecuadas desde el punto de vista de la Educación Inclusiva.

Los resultados cualitativos obtenidos en el grupo de discusión han contribuido a definir el constructo teórico medido por la escala ACOGE.

3.3.1. Análisis de las definiciones aportadas por los profesionales de la educación

Todas las definiciones que los profesionales realizaron sobre Educación Inclusiva reflejaron el significado de la variable *presencia*. De los 24 profesionales, 15 incluyeron también la variable *participación*, siendo los directores de centros los que dieron más importancia a esta idea participativa y comunitaria (11 de 12). Sin embargo, 8 de los 12 profesores tutores resaltaron en sus construcciones teóricas la idea de *progreso* y menos la de *participación*. De esto se infiere que cada rol profesional priorizó una variable u otra de la Inclusión Educativa en función (en parte) de la acción profesional con la que se encuentran en el día a día. Estos resultados se resumen en la Tabla 5.

Tabla 5. Resultados de las definiciones aportadas por los profesionales				
	<i>Presencia</i>	<i>Progreso</i>	<i>Participación</i>	<i>Total</i>
Directores	12	1	11	12
Profesores	12	8	4	12
Total	24	9	15	24

3.3.2. Nuevos ítems aportados por los profesionales de la educación

En la Tabla 6 se muestran los ítems propuestos y consensuados por los participantes en los grupos de discusión que fueron incluidos en la escala. Especificamos aquellos que han surgido como nuevas aportaciones y los que se modificaron con la finalidad de resultar más comprensibles o coincidentes con la realidad de la práctica educativa.

Tabla 6. Nuevos ítems propuestos por el grupo de discusión		
<i>Dimensión</i>	<i>Ítems</i>	<i>Cambio</i>
Presencia		
Participación	Coordinarse con profesorado de ciclo, equipo de orientación y jefatura de estudios	Nuevo
	Las metodologías puestas en práctica en el aula son acordes con el Proyecto Educativo de Centro	Nuevo
	Realizar análisis sobre el clima social del aula, implicando al alumnado, con el objetivo de mejorarlo	Fusión de dos ítems
	Parte de tu rol como maestro es facilitar que TODOS, a pesar de sus limitaciones, participen en las actividades programadas	Cambio: Conseguir por facilitar
	Aceptar asesoramiento realizado entre colegas por mejorar aspectos relacionados con la práctica docente	Cambio: Valorar la crítica constructiva por aceptar asesoramiento
Progreso	Valorar los logros individuales de cada estudiante mediante la evaluación de los diferentes resultados de aprendizaje	Fusión de dos ítems
	Las herramientas de evaluación del profesorado ayudan a mejorar su labor docente	Nuevo
	Mejorar la práctica docente en base a evidencias científicas	Nuevo
	El tutor es responsable del progreso de todo el alumnado que está en su aula	Nuevo

3.4. Fase 4. Escala piloto

Con los resultados obtenidos de los grupos de discusión se construyó una versión piloto inicial del instrumento, compuesta por 105 ítems. Para obtener una primera información sobre el funcionamiento de los ítems, se realizó un estudio piloto con 78 estudiantes universitarios. Se estimaron los índices de homogeneidad corregida (IHc) y se suprimieron los ítems con valores inferiores a .200. Se analizaron conforme a los parámetros *Importancia*, *Sensibilidad*, *Observabilidad* y *Dificultad*. Este análisis detectó nueve ítems con un funcionamiento inadecuado en la dimensión *presencia*, uno en la dimensión *participación* y uno en la dimensión de *progreso*. Se valoraron estos resultados en función de la pertinencia del contenido. A su vez, se utilizaron procedimientos analíticos cualitativos donde cada participante expuso las razones de tales valoraciones y los principios de actuación relevantes que conformaban cada bloque de ítems.

Para tomar una decisión final sobre la idoneidad o no de cada ítem, se organizó un nuevo grupo de discusión con profesionales especializados en Educación Compensatoria e Inclusión, en el que se indagó acerca de los motivos subyacentes a posibles defectos en los procesos de inclusión educativa. La sesión se inició con un trabajo individual de análisis de los ítems presentados. En esta ocasión, teniendo en cuenta la formación específica que poseía el grupo de profesionales, se les pidió que valorasen, en una escala de 1 a 4 (siendo 1 *nada* y 4 *mucho*), los parámetros de *importancia*, *sensibilidad*, *observabilidad* y *dificultad*. A su vez, también se les pidió su opinión acerca de la correcta o inadecuada ubicación de los ítems dentro de su dimensión teórica. Se modificaron y añadieron ítems según las propuestas generadas por el grupo de discusión.

3.4.1. Selección y modificación de los ítems de la escala piloto ACOGE (revisión cualitativa)

Para llegar a un equilibrio en cuanto al número de ítems en cada una de las tres dimensiones se siguieron los siguientes criterios de selección a partir de los resultados de los análisis previos: a) respuestas dadas por los jueces expertos; b) acuerdos generados en los grupos de discusión; y c) consenso adquirido en el grupo de investigadores.

En la dimensión *presencia*, el número de ítems considerados válidos por los jueces era inferior a 30, por lo que fue necesario rescatar ítems que habían sido seleccionados como *importantes* por los grupos de discusión. En las dimensiones *participación* y *progreso* el número de ítems que cumplieron criterios de validez en los análisis realizados de las respuestas de los jueces expertos fue superior a 30.

Para analizar el poder discriminativo de los ítems, utilizamos la prueba no paramétrica de Kruskal-Wallis para determinar si los ítems permitían discriminar entre grupos establecidos en función de la puntuación otorgada: *baja* (puntuación inferior al centil 25), *media baja* (puntuación inferior al centil 25), *media* (puntuaciones entre centiles 25 y 75) y *alta* (puntuaciones superiores al centil 75). Según los datos que obtuvimos pudimos afirmar que todos los elementos de la escala tenían poder discriminativo revelando, de esta manera, evidencias de validez de constructo de la escala ACOGE.

3.5. Fase 5. Evidencias de validez y fiabilidad

Para conocer el grado en que las relaciones de los ítems se correspondían con el constructo y mostrar así *evidencias de validez de constructo basada en la estructura interna de la escala*, sometimos a confirmación cinco modelos: *modelo unifactorial*, considerando la Inclusión como variable única; *modelo de tres factores correlacionados*; *modelo de tres factores ortogonales*; *modelo unifactorial* para cada una de las dimensiones (Presencia, Participación y Progreso); y *modelo de Análisis Factorial Confirmatorio Bifactor*. Tras los análisis realizados, en los que utilizamos como *método de estimación el de mínimos cuadrados ponderados robustos (WLSMW)* sobre la matriz de varianzas asintóticas y sobre la matriz de correlaciones policóricas, confirmamos el *modelo Bifactor*, que planteaba como hipótesis que la Inclusión consiste en una variable latente que explica la varianza de los cincuenta y cuatro ítems y tres variables no correlacionadas que explica cada una la variabilidad de sus correspondientes ítems. Justificamos su elección basándonos en el menor valor de *Ji-Cuadrado de Satorra-Bentler*, respecto al resto de modelos, y en el alcance de índices de ajuste complementarios satisfactorios: RMSEA = .022; CFI = .980; TLI = .979.

Los análisis de *fiabilidad de la escala* en términos de consistencia interna y error típico de medida mostraron que, tanto en términos globales basados en la covarianza de los ítems ($\alpha = .949$; $\lambda = .879$), como en términos basados en la correlación de las dos mitades ($r = .888$), existía una elevada fiabilidad. Confirmamos dicha fiabilidad mediante el análisis del *error típico de medida (ETM)* y su comparación con la *desviación típica (DT)*, encontrando una diferencia de más del doble en todas las dimensiones, lo que nos permitió reafirmar la precisión del instrumento.

4. Discusión y conclusiones

Este artículo ha presentado el desarrollo y validación de la escala ACOGE. Hemos secuenciado diferentes fases en las que se han alternado procedimientos psicométricos y análisis cualitativos referidos al análisis del contenido, todo ello con la finalidad de proporcionar evidencias de validez respecto al constructo de la Inclusión Educativa. Los informantes que han participado en la construcción del contenido teórico han sido profesionales de la educación con diferentes roles, que han aportado diferentes puntos de vista acerca de la *Inclusión Educativa* (directores de centros, tutores, profesores de Educación Compensatoria, Pedagogía Terapéutica y Audición y Lenguaje, personal perteneciente a Equipos Psicopedagógicos y Departamentos de Orientación).

La originalidad de este instrumento radica en que la construcción del mismo parte de la confluencia de las aportaciones teóricas y las consideraciones que los docentes tienen del concepto de Inclusión Educativa. Podemos de esta manera decir que es un instrumento hecho con ellos y para ellos. Cada vez tenemos más evidencias de que los docentes desempeñan un papel fundamental en el logro de sistemas educativos de calidad.

Ninguno de los instrumentos encontrados, referenciados en el artículo, están centrados en la práctica docente del maestro en su aula, por ello consideramos preciso elaborar esta escala centrada única y exclusivamente en ese microsistema, que ayudase a evaluar la práctica educativa y dar respuesta a la pregunta ¿qué objetivo perseguimos con la evaluación de las prácticas educativas? Defendemos que los procesos inclusivos tienen que estar acompañados de procesos evaluativos basados en dos principios fundamentales: el rigor y la aplicabilidad. La construcción de unos indicadores rigurosos, fiables y comparables; con elevada relevancia política; que conjuguen el conocimiento científico con lo que ocurre en las aulas, abarcando el conjunto de aspectos cruciales de los sistemas educativos, como son: las cualidades del docente, las metodologías, las técnicas y las estrategias de enseñanza-aprendizaje, así como las competencias docentes relacionadas directamente con los resultados del alumnado; constituye, hoy en día, una necesidad sentida por todos los responsables de la política educativa y asumida por los organismos internacionales que se ocupan de la educación: UNESCO, OCDE, UE. Esto es algo imprescindible para que los procesos evaluativos tengan sentido y empiecen a ser eficaces.

La práctica educativa de todo docente debe estar basada en un proceso sistemático de reflexión compartida. El punto de inflexión que nos permite modificar nuestra práctica es la reflexión. Un maestro reflexivo es capaz de conocer y aceptar la realidad de su aula y buscar diferentes soluciones a los problemas que en ella surgen. Esta actitud le va a permitir optimizar sus respuestas docentes integrándolas en su contexto, logrando de esta manera adaptarse a las necesidades de cada uno de sus alumnos. Como afirman Ainscow, Beresford, Harris, Hopkins y West (2010), “en realidad, la disposición para reflexionar sobre el propio ejercicio docente y anotar esas reflexiones de forma sistemática es una característica de los buenos profesionales” (p. 94).

Para dar utilidad a la escala hemos elaborado una guía que ayuda al docente en la reflexión y puesta en práctica de experiencias de éxito recogidas por la literatura científica que pueden adecuarse a cada uno de los contextos educativos en los que el docente desarrolla su labor profesional. Esta guía está organizada en seis apartados: (1) Explora tu aula. Escala ACOGE; (2) Conoce el contexto: alumnado, familias y entorno; (3) Organiza y prioriza tus sueños; (4) Haz los cambios siguiendo las prioridades; (5) Crea tu seña de identidad; y (6) Cuenta, comparte e implica a la comunidad educativa en tu modelo.

Como en cualquier investigación, este estudio presenta algunas limitaciones, entre las que se encuentran el tamaño de la muestra conseguida, que, a pesar de ser suficiente para permitirnos proceder a los análisis que una investigación de este tipo precisa, los resultados probablemente hubiesen alcanzado todos los criterios deseables si esta hubiese sido mayor. Otra limitación, que por otra parte fundamenta una de las motivaciones que nos llevaron a emprender este trabajo, ha sido la imposibilidad de aportar evidencias de validez referidas al criterio por no contar con datos que correlacionasen las puntuaciones de la escala ACOGE con las que se hubiesen obtenido en otras pruebas desarrolladas para el mismo fin.

Consideramos que el avance de instrumentos de evaluación como este, orientados hacia la intervención, puede mejorar las redes de colaboración entre investigadores

y educadores, así como generar prácticas escolares inclusivas y de calidad. Esta es, por tanto, la finalidad de la escala ACOGE, contribuir a mejorar las prácticas escolares desde una perspectiva inclusiva, superando las teorías del déficit que se han centrado en aportar ajustes específicos al alumnado, contemplando una perspectiva global y cultural desde donde proponer tales actuaciones.

Referencias bibliográficas

- AGENCIA EUROPEA PARA EL DESARROLLO DE LA EDUCACIÓN DEL ALUMNADO CON NECESIDADES EDUCATIVAS ESPECIALES (2009). *Desarrollo de indicadores sobre inclusión educativa en Europa*. Recuperado de https://www.european-agency.org/sites/default/files/development-of-a-set-of-indicators-for-inclusive-education-in-europe_indicators-ES.pdf.
- AINSCOW, M., BERESFORD, J., HARRIS, A., HOPKINS, D. y WEST, M. (2010). *Crear condiciones para la mejora del trabajo en el aula*. Madrid: Narcea.
- AINSCOW, M., BOOTH, T. y DYSON, A. (2006). *Improving schools, developing inclusion*. London: Routledge.
- AINSCOW, M., HOPKINS, D., SOUTHWORTH, G. y WEST, M. (2001). *Hacia escuelas eficaces para todos. Manual para la formación de equipos docentes*. Madrid: Narcea.
- ARNÁIZ, P. y GUIRAO, J. M. (2015). La autoevaluación de centros en España para la atención a la diversidad desde una perspectiva inclusiva: ACADI. *Revista Electrónica Interuniversitaria de Formación del Profesorado*, 18 (1), 45-101.
- BANGDIWALA, K. (1987). Using SAS software graphical procedures for the observer agreement chart. *Proceedings of the SAS Users Group International Conference*, 12, 1083-1088.
- BENNETT, E. M., ALPERT, R. y GOLDSTEIN, A. C. (1954). Communications through limited response questioning. *Public Opinion Quarterly*, 18, 303-308.
- BIENCINTO, C., GONZÁLEZ, C., GARCÍA, M., SÁNCHEZ, P. y MADRID, D. (2009). Diseño y propiedades psicométricas del AVACO_EVADIE. Cuestionario para la evaluación de la atención a la diversidad como dimensión educativa en las instituciones escolares. *Revista Electrónica de Investigación y Evaluación Educativa*, 15 (1), 1-36. Recuperado de: http://www.uv.es/RELIEVE/v15n1/RELIEVEv15n1_4.htm.
- BOOTH, T. (2006). Manteniendo el futuro con vida; convirtiendo los valores de la inclusión en acciones. En M. Á. VERDUGO y F. B. JORDÁN DE URRÍES (Coords.), *Rompiendo inercias. Claves para avanzar. VI jornadas científicas de investigación sobre personas con discapacidad* (pp. 211-217). Salamanca: Amarú.
- BOOTH, T. y AINSCOW, M. (2002). *The Index for Inclusion* (2nd ed.). United Kingdom: CSIE.
- BOOTH, T. y AINSCOW, M. (2011). *Index for inclusion: developing learning and participation in schools* (3rd ed.). United Kingdom: CSIE.
- CASANOVA, M. A. (2004). *Evaluación y calidad de centros educativos*. Madrid: La Muralla.
- CASANOVA, M. A. (2011). De la educación especial a la inclusión educativa. Estado de la cuestión y retos pendientes. *Participación Educativa*, 18, 8-24.
- CASAR, L. S. (2007). *Atención a la diversidad. Un estudio de su calidad en centros de la provincia de A Coruña* (Tesis doctoral no publicada). Universidad de A Coruña, A Coruña.
- COHEN, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20 (1), 37-46.
- DE LA ORDEN, A. (2009). Evaluación y calidad: análisis de un modelo. *ESE. Estudios sobre Educación*, 16, 17-36.

- DUK, C. (2007). "Inclusiva". Modelo para evaluar la respuesta a la escuela a la diversidad de necesidades educativas de los estudiantes. Proyecto FONDEF/CONICYT D04I1313. *REICE. Revista Electrónica Iberoamericana sobre Calidad, Eficacia y Cambio en Educación*, 5 (5e), 188-199. Recuperado de: <http://redalyc.uaemex.mx/redalyc/pdf/551/55121025027.pdf>.
- ECHEITA, G. (2013). Inclusión y exclusión educativa: de nuevo "Voz y Quebranto". *REICE. Revista Electrónica Iberoamericana sobre Calidad, Eficacia y Cambio en Educación*, 11 (2), 100-118.
- ECHEITA, G. y AINSCOW, M. (2011). La educación inclusiva como derecho marco de referencia y pautas de acción para el desarrollo de una revolución pendiente. *Tejuelo: Didáctica de la Lengua y la Literatura. Educación*, 12, 26-46.
- ECHEITA, G., MUÑOZ, Y., SANDOVAL, M. y SIMÓN, C. (2014). Reflexionando en voz alta sobre el sentido y algunos saberes proporcionados por la investigación en el ámbito de la educación inclusiva. *Revista Latinoamericana de Educación Inclusiva*, 8 (2), 25-48.
- ECHEITA, G., VERDUGO, M. Á., SANDOVAL, M., CALVO, I. y GONZÁLEZ-GIL, F. (2009). Paradojas y dilemas en el proceso de inclusión educativa en España. *Revista de Educación*, 349, 153-178.
- FERNÁNDEZ, M. J. y GONZÁLEZ, A. (1997). Desarrollo y situación actual de los estudios de eficacia escolar. *RELIEVE*, 3 (1), n. 1. Recuperado de http://www.uv.es/RELIEVE/v3n1/RELIEVEv3n1_3.htm.
- FLORIAN, L., ROUSE, M. y BLACK-HAWKINS, K. (2011). Researching achievement and inclusion to improve the educational experiences and outcomes of all learners. *Aula*, 17, 57-72.
- INTERPROJECT (2005). *Guía INTER. Una guía práctica para aplicar la educación intercultural en la escuela*. Recuperado de http://inter.up.pt/docs/spanish_guide.pdf.
- KRIPPENDORFF, K. (1970). La estimación de la fiabilidad, error sistemático y error aleatorio de datos de intervalo. *Educativa y Psicológica Medición*, 30 (1), 61-70.
- KRUEGER, R. A. (1991). *Focus groups: A practical guide for applied research*. Beverly Hills, CA: Sage.
- KRUEGER R. A. y CASEY, M. A. (2000). *Focus Group a practical guide for applied research* (3th Ed.). Thousand Oaks, CA: Sage.
- MARCHESI, A., DURÁN, D., GINÉ, C. y HERNÁNDEZ, L. (2009). *Guía para la reflexión y valoración de prácticas inclusivas*. Madrid: Organización de Estados Iberoamericanos para la Educación, la Ciencia y la Cultura (OEI).
- MERTON, R. y KENDALL, P. (1946). The Focused Interview. *The American Journal of Sociology*, 51 (6), 541-557.
- MIRAS, M. (2001). Afectos, emociones, atribuciones y expectativas: el sentido del aprendizaje escolar. En C. COLL, J. PALACIOS y A. MARCHESI (Comps.), *Desarrollo psicológico y educación. II Psicología de la Educación escolar*. 2.ª ed. (pp. 309-329). Madrid: Alianza.
- MUNTANER, J., FORTEZA, D., ROSELLÓ, M. R., VERGER, S. e IGLESIA, B. (2010). *Estándares e indicadores para analizar la calidad de vida del alumnado con discapacidad en su proceso educativo*. Palma de Mallorca: Edicions UIB.
- OCDE (2009). *Informe TALIS. La creación de entornos eficaces de enseñanza y aprendizaje. Síntesis de los primeros resultados*. Madrid: Santillana Educación.
- OCDE (2011). *PISA in Focus 2. Improving Performance: Leading from the Bottom*. París: OCDE.
- PERREAULT, W. D. y LEIGH, L. E. (1989). Reliability of nominal data based on qualitative judgments. *Journal of Marketing Research*, 26 (May), 135-148.
- SIMÓN, C. y ECHEITA, G. (2013). Comprender la inclusión educativa para intentar llevarla a la práctica. En H. RODRÍGUEZ y L. TORREGO (Coords.), *Educación inclusiva, equidad y derecho a la diferencia. Transformando la escuela* (pp. 33-65). Madrid: Wolters Kluwer España.

