



Revista Electrónica Educare
ISSN: 1409-4258
Universidad Nacional. CIDE

Velandia-Mesa, Cristian; Serrano-Pastor, Francisca José; Martínez-Segura, María José
Evaluación de la investigación formativa: Diseño y validación de escala

Revista Electrónica Educare, vol. 25, núm. 1, 2021, Enero-Abril, pp. 35-54

Universidad Nacional. CIDE

DOI: 10.15359/ree.25-1.3

Disponible en: <http://www.redalyc.org/articulo.oa?id=194165972003>

- Cómo citar el artículo
- Número completo
- Más información del artículo
- Página de la revista en redalyc.org

UAEM
redalyc.org

Sistema de Información Científica Redalyc
Red de Revistas Científicas de América Latina y el Caribe, España y Portugal
Proyecto académico sin fines de lucro, desarrollado bajo la iniciativa de acceso abierto

[Cierre de edición el 01 de Enero del 2021]

<http://doi.org/10.15359/ree.25-1.3><http://www.una.ac.cr/educare>

educare@una.ac.cr

Evaluación de la investigación formativa: Diseño y validación de escala

Assessing Education Research Training: Scale Design and Validation

Avaliação da investigação formativa: Desenho e validação de escala

Cristian Velandia-Mesa

Universidad El Bosque
Bogotá, Colombia



velandiacristian@unbosque.edu.co
<https://orcid.org/0000-0002-7195-3365>



Francisca José Serrano-Pastor

Universidad de Murcia
Murcia, España
fjserran@um.es



<https://orcid.org/0000-0002-5193-209X>

María José Martínez-Segura

Universidad de Murcia
Murcia, España
mjmarti@um.es



<https://orcid.org/0000-0001-7022-5780>

Recibido • Received • Recebido: 07 / 03 / 2019

Corregido • Revised • Revisado: 05 / 08 / 2020

Aceptado • Accepted • Aprovado: 19 / 11 / 2020

Resumen: El artículo presenta los resultados de un estudio cuyo propósito se focalizó en el diseño y validación de una escala para la valoración de la formación en investigación educativa, basada en el modelo matemático en torno a las respuestas del ítem (IRT) y adaptada a la escala acumulativa de Guttman. Atendiendo a los niveles de investigación científica, el estudio fue desarrollado en cuatro fases: 1) exploratoria (valididad de contenido) por medio de muestreo teórico, entrevistas semiestructuradas y grupos focales; 2) descriptiva (valididad de constructo) a través del método Delphi; 3) analítica (consistencia) y 4) explicativa (análisis R&R) por medio la aplicación de la escala en mención. Intervinieron en el estudio 19 participantes del profesorado investigador, 5 egresadas del programa de licenciatura y 26 integrantes del alumnado en formación investigativa. La argumentación teórico-contextual, la clasificación de reactivos a través de conglomerados, las medidas de concordancia, los índices de consistencia interna y el análisis de estabilidad (R&R) permitieron fundamentar la validación de la escala. La ejecución de la escala a través de constructos, reactivos e indicadores binarios diferenciales proporciona una valiosa aproximación a la valoración de los procesos de formación investigativa al tener en cuenta las dimensiones del saber, el hacer

y el ser. Con los hallazgos del estudio se busca contribuir en el fortalecimiento de la investigación educativa por medio de la reflexión en torno a su evaluación.

Palabras claves: Investigación educativa; relación investigación-docencia; formación de investigadores e investigadoras; escala de medición; instrumentos de evaluación; evaluación de competencias.

Abstract: The article presents the results of a study whose purpose was the design and validation of a scale to evaluate education research training. The study was based on the Item Response Theory (IRT) mathematical model, with adaptation to Guttman's cumulative scale. The levels of scientific research were considered to conduct the study in the following phases: 1) exploratory phase (content validity) by means of theoretical sampling, semi-structured interviews, and focus groups; 2) descriptive stage (construct validity) through the Delphi method; 3) analytical phase (consistency); 4) explanatory stage (R & R analysis) through the application of the scale in mention. Nineteen professors-researchers, five graduates of the degree program, and 26 students of research training participated in the study. Theoretical-contextual argumentation, classification of reagents in hierarchical conglomerates, concordance measures, internal consistency indices, and stability analysis allowed to establish the validation of the scale. The execution of the scale through constructs, reagents, and binary differential indicators provides a valuable approach to evaluating research training processes, taking into account the dimensions of knowledge, doing, and being. In essence, the study's findings seek to contribute to strengthening education research around its assessment.

Keywords: Education research, research-teaching relationship, training of researchers, measurement scale, evaluation instruments, competency evaluation.

Resumo: O artigo apresenta os resultados de um estudo cujo fim focalizou-se no desenho e validação de uma escala de valoração da formação em pesquisa educacional, baseada no modelo matemático ao redor das respostas do item (IRT), e adaptado à escala cumulativa de Guttam. De acordo aos modelos de pesquisa científica, o estudo foi desenvolvido em quatro estágios: 1) exploratório (validade do conteúdo) por meio de amostragem teórica, entrevistas semiestruturadas e grupos focais; 2) descritivo (validade da construção) através do método delphi; 3) analítico (consistência) e 4) explicativo (análise R&R) por meio da aplicação da escala em menção. No estudo participaram 19 professores-pesquisadores, 5 formados no programa de licenciatura e 26 estudantes de formação em pesquisa. A argumentação teórico-contextual, a classificação de reativos por meio de conglomerados, as medidas de concordância, os índices de consistência interna e a análise de estabilidade (R&R) permitiram fundamentar a validação da escala. A execução da escala a través de construções, reativos e indicadores binários diferenciais deram uma valiosa aproximação à valoração dos processos de formação em pesquisa, levando em conta as dimensões do saber, do fazer e do ser. As descobertas do estudo podem contribuir no fortalecimento da pesquisa educativa, por meio da reflexão sobre sua avaliação.

Palavras-chave: Pesquisa educacional; relação pesquisa-ensino; capacitação de pesquisadores; escala de medição; instrumentos de avaliação; avaliação de competências.



Introducción

La formación para la investigación ha sido un tema vinculado a la construcción de políticas educativas, dado el papel que desempeña en las prácticas pedagógicas y en el escenario académico como principio de transformación social (Rojas Granada y Aguirre Cano, 2015). La formación para la investigación educativa busca alcanzar la comprensión de fenómenos y procesos, mediante la “investigación-reflexión”, cuyo desarrollo parte de experiencias que se someten a una reelaboración permanente con base en cuestionamientos, para orientarlas al servicio y a la capacidad de resolución de problemas en el ámbito académico (Córdoba, 2016). Sujeto a esto, la evaluación de los procesos de formación investigativa, en el escenario de la educación superior, busca fortalecer la relación educación-investigación a través de la valoración del desempeño de los futuros grupos de profesionales frente a sus prácticas pedagógicas (Velandia-Mesa et al., 2017), para lo cual la investigación educativa se convierte en un reto que orienta sus acciones al mejoramiento del aprendizaje, transformadas en oportunidades para enriquecer la enseñanza.

Actualmente, en educación superior se requiere llevar a cabo aproximaciones significativas respecto a la evaluación de los procesos de formación investigativa, factor asociado a un marco curricular que considera la investigación como parte del desempeño y perfil de un personal docente transformador de un escenario social, cultural o educativo (Da Cunha, 2015). En consecuencia, es oportuno comprender la investigación formativa como un saber integral, no solo basado en conocimientos y metodologías, lo que conlleva a centrar la evaluación desde las dimensiones del saber, saber hacer y saber ser.

Uno de los elementos que se estima como importante para fortalecer la valoración del desempeño en investigación está asociado a los instrumentos que se aplican durante la evaluación de la formación investigativa. Si bien, la evaluación del trabajo de fin de grado y su defensa son una aproximación al desempeño del estudiantado en metodología de la investigación, este no es el único referente (Correa Bautista, 2009). Esto conduce al personal docente e investigador a concentrar esfuerzos en la construcción y validación de escalas de valoración, focalizadas en la evaluación de conocimientos, habilidades y actitudes que se procuran consolidar durante la formación investigativa.

El diseño, validación y aplicación de escalas de valoración conforma un momento fundamental en los procesos de investigación educativa (Valdés Cuervo et al., 2018). La construcción rigurosa de instrumentos constituye una etapa importante en la mejora de procesos de diagnóstico, intervención y evaluación de procesos formativos (Romero et al., 2006). Si bien existen diversas consideraciones relacionadas con la evaluación de los procesos de formación con relación a su pertinencia, capacidad discriminante, claridad, entre otros, la validación de instrumentos contempla una aproximación significativa frente al interés de identificar y valorar conocimientos, resultados de aprendizaje y desempeños, como parte de

una formación integral ([Velandia-Mesa et al., 2017](#)). Por esta razón, el diseño y validación de una escala de valoración para la formación investigativa es un proceso articulado al fortalecimiento de la investigación formativa, que trasciende la medida de los alcances logrados a orientar la estimación de la formación como parte de una experiencia conducente a mejores desempeños en la labor profesional ([González Campos et al., 2016](#)).

Este estudio considera el diseño, construcción y validación de una escala de estimación acumulativa, recurso que permite valorar los procesos de formación en investigación educativa a través de constructos relacionados con competencias investigativas y reactivos (indicadores binarios diferenciales) asociados a resultados de aprendizaje. Esta escala es una herramienta operativa y sistemática que permite una aproximación a la valoración del proceso de formación investigativa; sin embargo, conviene resaltar que tal estimación es producto de una articulación entre las consideraciones formativas traducidas en competencias y desempeños desde las dimensiones del ser, el hacer y el saber ([Piña-Osorio y Aguayo-Rousell, 2017](#)). En suma, la evaluación de la formación investigativa es un aspecto importante debido a que, a través de la valoración y su estimación, se evidencian, en gran medida, los logros y los alcances respecto a los procesos de enseñanza y aprendizaje asociados a la investigación como estrategia pedagógica ([García Bejarano et al., 2013](#)).

Las metodologías de evaluación de procesos educativos evolucionan constantemente en busca del progreso y su representatividad ([González-Espino, 2017](#)). Los constructos educativos respecto al saber, el hacer y el ser resultan ser concepciones teóricas que no son accesibles y valorables fácilmente, más aún, cuando reconocerlas presupone múltiples dimensiones para identificarlas. No obstante, frecuentemente los instrumentos de evaluación son focalizados en la optimización de su consistencia y estabilidad, los cuales se manifiestan por la medición del constructo, al imprimir el sello de la teoría clásica que establece una correlación lineal entre la medición obtenida y el nivel de constructo evaluado ([Hambleton et al., 1991](#)).

En otras palabras, basados en la teoría clásica, se estima que la calificación obtenida por el estudiantado es el resultado de la suma hipotética ejecutada por el juicio del ente evaluador más el error de medida propio del constructo. Este modelo presupone resistencia frente a las características evaluadas, ya que los constructos valorados, por lo general, no solo dependen de la esencia y propiedades de los instrumentos, sino también de quien los ejecuta. De igual modo, los resultados de la valoración de un proceso formativo también están ligados la capacidad discriminante del instrumento aplicado, ya que el coeficiente de dificultad estimado dependerá de quien lo construya; esto supone que existe una dependencia entre el instrumento, quien lo ejecuta y a quien se evalúa, y, por tanto, confrontar los resultados del aprendizaje entre el mismo grupo de estudiantes podría resultar arbitrario o parcial. Tales consideraciones problemáticas convergen en un bajo índice de concordancia (estabilidad) entre las mediciones determinadas por parte de los jurados evaluadores, ejecutadas con el mismo instrumento y en diferentes momentos.



En los últimos años, en el ámbito educativo, se ha producido el interés de reconocer, fortalecer y consolidar determinadas competencias investigativas (Aular et al., 2009; Reiban Barrera et al., 2017; Colás Bravo et al., 2009; García-Gutiérrez y Aznar-Díaz, 2019; Jaik Dipp, 2013; Velandia-Mesa et al., 2019) debido a que la formación en investigación hace parte de la identidad universitaria y permite el agenciamiento de la calidad educativa. Otros estudios como los de Balbo et al., (2015), Correa Bautista (2009); Pal et al. (2014); y González-Espino (2017), por su parte, han centrado su atención en la construcción de instrumentos con fines de evaluación de productos investigativos a partir de criterios focalizados en la valoración de trabajos de fin de grado y maestría. Estos estándares, así como algunos instrumentos de evaluación basados en objetivos de aprendizaje, constituyen un valioso recurso para el personal docente; sin embargo, basados en los antecedentes y las experiencias en el contexto del estudio (análisis de índices de reproducibilidad y repetitividad, capacidad discriminante, estabilidad, entre otros) se considera que a través de estos modelos se reduce la valoración del aprendizaje a los conocimientos transmitidos cuyo producto está determinado por un documento estandarizado y la defensa del proyecto.

En respuesta, se propone construir y validar una escala de valoración de la formación en investigación educativa basada en el modelo matemático en torno a las respuestas del ítem (IRT) y adaptada a la escala acumulativa de Guttman. A través de la escala multidimensional (producto del estudio), se pretende valorar la formación investigativa a partir de resultados de aprendizaje, mediante constructos asociados a declaraciones (ítems) donde cada uno de ellos cuenta con categorías diferenciales que se valoran de forma binaria, por medio de reactivos focalizados con la intención de establecer la probabilidad de discriminación y grado de dificultad de los ítems, atendiendo las consideraciones (IRT) de Fischer y Ponocny-Seliger (1998).

La revisión y el análisis documental ha contribuido a la elaboración y validación de la escala que presenta este estudio, así como al razonamiento respecto a la interpretación de los resultados del proceso formativo. En este sentido, el discernimiento sobre la situación del aprendizaje se ha basado en la evaluación criterial e ideográfica (Mesías Maraví y Frisancho León, 2013), que establece principios esenciales frente a la validez de contenido y justifica la relevancia de criterio (valorar el rendimiento del estudiante de manera absoluta comparando situación inicial y situación actual) más allá del precepto o el estándar que involucra un carácter netamente estadístico relativo.

Durante la dinámica de validación se ha considerado la evaluación como parte de un rol orientador que manifiesta a cada estudiante la posibilidad para identificar aspectos por mejorar y desempeños por consolidar, de modo antagónico a la concepción de la evaluación selectiva o como medio de clasificación de estudiantes. Así mismo, a través del estudio se abordó la posibilidad de desarrollar constructos, ítems y reactivos discriminantes que conllevan a descubrir y promover aptitudes durante los procesos de acompañamiento y seguimiento a la formación para la investigación educativa.

En el contexto del estudio presentado en este artículo, se ha abordado la formación en investigación educativa a través de seminarios presenciales y la ejecución de acciones en contextos de prácticas pedagógicas, mediadas a través de recursos tecnológicos para asistir y evaluar los procesos formativos durante cuatro momentos académicos (rutas de formación investigativa) correspondientes a: 1) planteamiento del proceso investigativo, 2) reconstrucción teórica, 3) trabajo de campo y 4) construcción para el futuro ([García Bejarano et al., 2013](#)), etapas de la formación para la investigación, no interdependientes y consecuentes con los procesos de evaluación. Si bien, la evaluación de la formación investigativa es un proceso complejo dada su naturaleza diversa y dinámica propia de la investigación, se consideró la optimización del grado de estabilidad de la escala, por medio de comparaciones en el tiempo como elemento inherente al proceso evaluativo, confrontando el rendimiento logrado con los objetivos previstos, basados en la premisa de que mientras mayor sea la información que aporta determinado ítem, mayor es la rigurosidad de la estimación del constructo y, por consiguiente, mayor es la estimación de la evolución del estudiante en función del aprendizaje ([Pal et al., 2014](#)).

Con el fin de orientar la lectura del artículo, se presenta, en la sección Material y método, la descripción del proceder metodológico por fases de investigación (exploratoria, descriptiva, relacional y explicativa). Posteriormente en la sección de Resultados, se describen los hallazgos de las fases mencionadas. Se culmina el artículo abordando discusiones y reflexiones finales.

Material y método

Atendiendo los niveles de investigación científica ([Supo-Condori, 2013](#)), el estudio fue desarrollado en cuatro momentos: La primera fase denominada validez de contenido, que corresponde a una etapa exploratoria basada en información documental y datos de contexto a través de entrevistas semiestructuradas y grupos focales. La segunda fase correspondiente a la validez de constructo, que responde a una etapa transversal descriptiva, cuyo fin fue la formulación, clasificación (conglomerados) y evaluación a través de la concordancia de jueces expertos. La tercera etapa mencionada como fase de análisis de consistencia interna, corresponde a una etapa de estudio relacional, en la cual se abordó un proceso inferencial respecto al análisis de fiabilidad y solidez de la escala. Finalmente, la fase explicativa corresponde a una etapa correlacional causal, que comprende un estudio de reproducibilidad y repetitividad (R&R) por medio del análisis de varianza y concordancia.

Fases

Fase exploratoria - validez de contenido

Dada la complejidad del proceso de valoración de la formación investigativa, debido a la dinámica y multidimensionalidad propia de los fenómenos educativos, se dio lugar al proceso de validación de contenido de la siguiente forma:



Identificando objetivos de la formación, resultados de aprendizaje, dimensiones pedagógicas, estándares y criterios de calidad en investigación, competencias investigativas, rubricas, instrumentos y patrones de evaluación. Para tal fin se llevó a cabo un análisis documental a través de muestreo teórico, que abordó factores de los modelos de calidad de la [Agencia Nacional de Evaluación de la Calidad y la Acreditación \(ANECA\) \(2013\)](#) de España y el [Consejo Nacional de Acreditación \(CNA\) \(2013\)](#) de Colombia y colaboraciones bibliográficas de autoridades e instituciones de educación superior respecto a la valoración de la formación en investigación educativa ([Balbo et al., 2015; Correa Bautista, 2009; González-Espino, 2017; Pal et al., 2014](#)). Este proceso originó una primera aproximación frente al registro de contenidos, objetivos y resultados de aprendizaje estimados como fundamentales en la formación investigativa, parcialmente definidos y abordados en un listado inicial de 144 resultados de aprendizaje asociados a los contenidos propuestos en los *syllabus* de las asignaturas de metodología de investigación educativa.

Posteriormente, se procedió a llevar a cabo un taller en mesas de trabajo con un panel del profesorado del área de investigación, representados por unidades académicas (departamentos, facultades y vicerrectoría), que manifestaron elementos que respaldaron la articulación entre la teoría y las necesidades contextuales frente a los procesos de formación investigativa. En correspondencia, la recolección de información de esta etapa se llevó a cabo a través de un protocolo de entrevista colectiva ejecutada en el grupo focal asociado a categorías de análisis como: resultados de aprendizaje, recursos, perfil del personal investigador en formación, estrategias de enseñanza, acompañamiento, seguimiento y evaluación. Las categorías y constructos emergen como parte de un proceso de análisis de necesidades contextuales y autoevaluación, que busca identificar competencias y desempeños en la formación para la investigación educativa.

Este proceso fue sistematizado y asistido con el apoyo del software AtlasTi8, se identificaron códigos representados por densidades (D - número de asociaciones entre categorías) y el peso (E- prevalencia en voces representativas asociadas a cada código).

Fase descriptiva - validez de constructo

Una vez reconocidas las competencias en investigación educativa e identificados los resultados de aprendizaje (RA) para la formación en investigación ([Velandia-Mesa et al., 2019](#)) se procedió al estudio de validez de constructo por medio de:

- Primera formulación de reactivos (ítems): Construida por parte del grupo investigador para responder a 4 indicadores binarios (diferenciales) atendiendo al análisis documental, la argumentación teórica, y basados en el supuesto que resalta una relación lineal frente a la información de ítem y la rigurosidad de la estimación del constructo.



- Clasificación de reactivos: Durante este proceso se solicitó a los integrantes del grupo de investigación, vincular cada uno de los ítems a los criterios de evaluación en el que consideraba mayor correspondencia; este proceso se denominó clasificación categórica. Esto generó una matriz ordinal que se analizó por cada una de las cuatro rutas de formación, acudiendo a la reducción de dimensiones parte del método de extracción de componentes principales, con un número fijo de factores correspondiente a dos.
- Evaluación del constructo: Luego de la elaboración y clasificación de reactivos, criterios de evaluación e indicadores binarios diferenciales (categórica y factorial), se llevó cabo una primera valoración de la escala por parte de los equipos evaluadores (jueces y juezas, personas expertas y estudiantes) donde se estimaron los criterios que se describen en la [Tabla 1](#):

Tabla 1: Ítems relacionados a criterios de evaluación. Ejemplo criterio: planteamiento de la investigación

Ítems crit.	Pertinencia ítem	Suficiencia ítems por criterio	Claridad ítem	Nivel de dificultad	Capacidad discriminante	Objetividad diferencial
DIM	Mantener (1) Retirar (0)	Suficiente (1) Incompleto (0)	Comprensible (1) Incomprensible (0)	Adecuado (1) Inadecuado (0)	Adecuado (1) Inadecuado (0)	Objetivo (1) Subjetivo (0)
1	1	1	1	1	1	1
2	1	0	0	1	1	1
3	0	1	1	0	1	0
4...	1	1	0	0	1	1

- Pertinencia: Estimación de conservación o exclusión de ítems en la escala.
- Suficiencia: Analiza si el número de ítems conlleva a evaluar adecuadamente el objetivo de aprendizaje.
- Claridad: Estimación respecto a la terminología y redacción de los ítems.
- Nivel de dificultad: Nivel de estimación en indicadores binarios.
- Capacidad discriminante: Categorías nominales que permiten identificar dificultades y fortalezas.
- Objetividad: Indicador que permite reconocer si las categorías diferenciales están expresadas en conductas observables o medibles.

Nota: Elaboración propia.



Fase relacional - consistencia interna

Conforme a las fases anteriores, se logró construir un primer borrador de la escala de valoración que conllevó a ejecutarla. Para ello, se evaluó la consistencia interna de la escala y se obtuvo la calificación global de cada estudiante mediante una prueba piloto; para tal fin se valoró el proceso de formación investigativa a través del planteamiento de investigación, referentes de investigación, la sistematización del trabajo de campo y la divulgación de los resultados en el documento final de grado.

Con el fin de fortalecer la capacidad discriminante de la escala de valoración, se optó por identificar, distinguir y determinar los niveles de logro en los procesos de formación investigativa; para ello, se construyeron indicadores binarios y categorías diferenciales cuyos valores en conjunto determinan los niveles de alcance de los objetivos de aprendizaje (destacado, satisfactorio, en proceso, en inicio, sin abordar) entre uno y cinco. Las calificaciones finales para cada uno de los resultados del aprendizaje en el proceso de formación investigativa, se obtuvieron en función del puntaje total alcanzado por cada individuo respecto a la suma de los indicadores binarios; de esta manera la escala permitió calcular un valor ordinal, con el fin de identificar la correlación de cada ítem respecto a la suma total del ítem.

Este proceso conllevó a determinar la consistencia interna de la escala a través de la técnica de dos mitades (debido a que los ítems están fragmentados por competencias investigativas) y corroborarla posteriormente con el cálculo del alfa de Cronbach. Para finalizar, se clasificaron los ítems respecto al índice de correlación (ítem-total) de mayor a menor identificando los reactivos con menor magnitud de asociación, cuyo proceso está relacionado con el "Alfa de Cronbach si se elimina el ítem" proceso que permitió analizar la consistencia interna y analizar la capacidad discriminante de la escala.

Fase explicativa - estabilidad

Para avanzar en la validación de la escala de valoración de la formación investigativa y teniendo en cuenta los resultados de las fases anteriores, se procedió al análisis de patrones de evaluación frente a los conceptos de repetitividad y reproducibilidad (R&R), basados en la aleatoriedad y la desviación de la medición sistemática. Generalmente, los instrumentos de recolección de información en educación consideran la validez de contenido a través de la evaluación documental y el juicio experto, así como la validez de constructo basados en los niveles de fiabilidad (Romero et al., 2006), sin embargo, en los procesos de valoración de la formación investigativa como parte del diagnóstico del presente estudio, se demostró que existe un bajo índice de concordancia ($Kappa < 0,4$) entre las mediciones determinadas por parte del sistema de evaluación usado, lo que conllevó a obtener diferencias estadísticamente significativas en la valoración por parte de jurados de proyectos con el mismo instrumento de evaluación durante el proceso de planteamiento de la investigación.



Frente a la condición mencionada, se condujo a la aplicación de la escala de valoración de la formación investigativa para determinar los índices de repetitividad y reproductibilidad (R&R); el primer concepto entendido como una medida de dispersión caracterizada por los resultados de evaluaciones sucesivas a un mismo grupo de estudiantes, con el mismo protocolo de valoración y el mismo evaluador; el segundo criterio entendido como el nivel de proximidad o de concordancia entre los resultados de las evaluaciones por parte de diferentes jurados evaluadores.

Los índices de R&R fueron expresados en términos de dispersión, cuyos métodos de determinación se basaron en la evaluación estadística a través de la representación de la desviación estándar y ANOVA, el cual segmenta la reproducibilidad en sus componentes de operador y operador por parte. Durante esta fase se persiguió el objetivo de determinar los índices de R&R sistematizando los resultados de la valoración de la formación investigativa por parte de tres jueces que evaluaron en dos momentos de evaluación, aplicando un estudio de medición de sistema cruzado a partir de las herramientas de calidad del software Minitab 17.

Cada grupo de estudiantes fue monitoreado y valorado por 3 miembros del grupo profesoral (de tutoría, de asesoría, de lectura). El profesorado y el alumnado evaluaron el proceso de formación investigativa durante 2 semestres académicos (cohorte) a través de la escala, sin retomar las valoraciones determinadas en el diagnóstico. Se obtuvo una calificación global por parte de los jurados en los momentos de evaluación determinados, cuyos insumos fueron determinantes para el análisis de R&R.

Participantes

Las personas participantes en el proceso de validación de la escala de valoración de la formación investigativa, estuvieron constituidas en tres grupos: 1) Un panel de 11 miembros del profesorado investigador, quienes actuaron por su experiencia en el grupo focal y la clasificación de ítems. 2) Grupo participante correspondiente a 8 integrantes del grupo de investigación y 5 egresadas del programa de pedagogía (asistentes de investigación), a totalidad procedió como el equipo evaluador. Finalmente, la aplicación del instrumento como prueba diagnóstica, se llevó a cabo en 26 estudiantes de formación investigativa en el programa de licenciatura. Dada la población objetivo, los criterios de selección fueron establecidos a partir del método de muestreo discrecional basado en las características contextuales de formación en investigación educativa. Durante el proceso se tuvo en cuenta las consideraciones éticas avaladas por el comité técnico *ad hoc* y la implementación de consentimientos informados.

Resultados

Estructuralmente en este apartado se exponen los hallazgos del estudio, correspondiendo a cada una de las fases metodológicas propuestas durante la validación de la escala para la



valoración de la formación investigativa. Para mayor claridad se ha segmentado la metodología y los resultados en fases; no obstante, las etapas de validación de la escala se encuentran en un diálogo permanente y actúan como elementos interdependientes:

Fase exploratoria - validez de contenido

Los resultados de esta etapa radican en la identificación de las competencias investigativas en un diálogo permanente con los resultados de aprendizaje asociados a la formación investigativa, los cuales se basaron en las dimensiones del saber, el saber hacer y el saber ser. De conformidad se encontró una tendencia hacia un rediseño curricular basado en el fortalecimiento de las prácticas investigativas como parte del perfil de egreso de profesionales en educación, lo que conlleva a focalizar los esfuerzos en la ética investigativa, el trabajo en equipo, la comunicación argumentativa, el desarrollo de pensamiento crítico y la resolución de problemas.

Fruto del muestreo teórico (antecedentes, estándares, políticas institucionales) y el análisis de datos provenientes del grupo focal (identificación de resultados del aprendizaje, construcción de un perfil de personal docente investigador en formación, el uso de estrategias de formación y métodos de evaluación) se determinaron 9 competencias investigativas (ética, tecnológica, argumentativa, relacional, comunicativa, crítica, propositiva, metodológica y analítica) agrupadas en 27 criterios de evaluación y definidos finalmente en 110 resultados de aprendizaje representados en la escala de valoración.

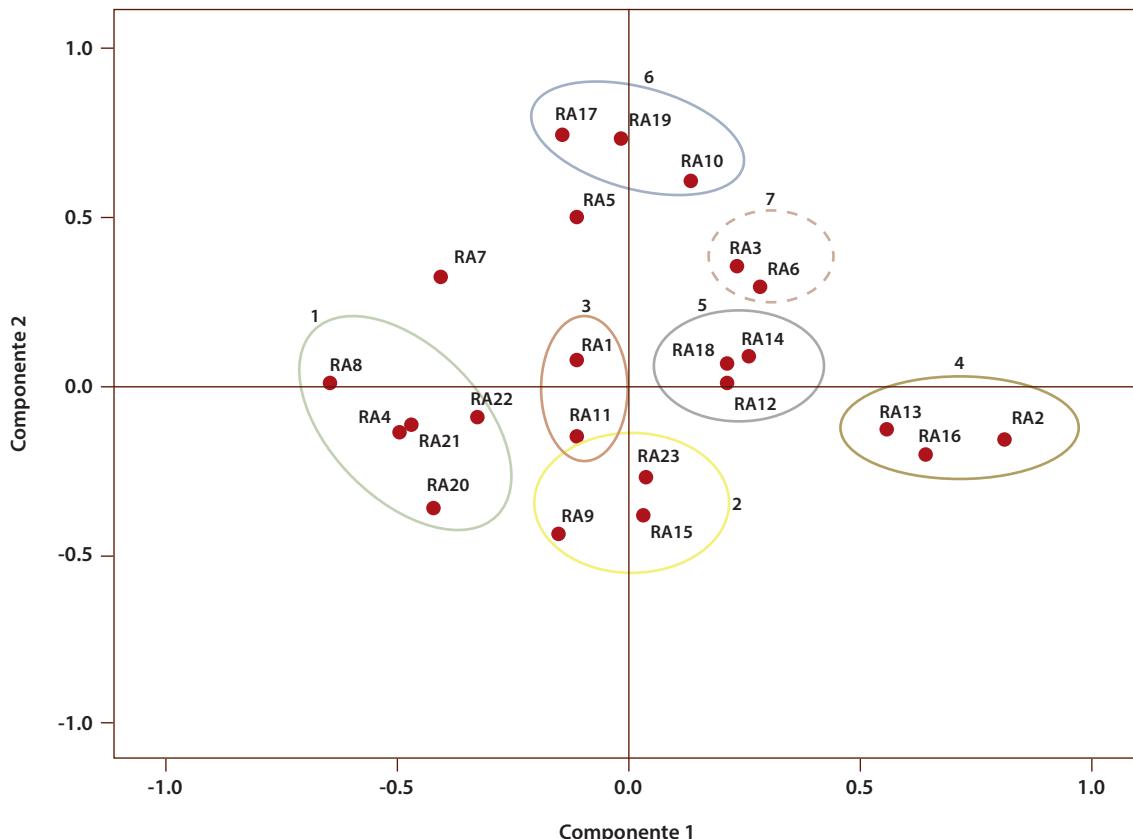
Fase descriptiva – validez de constructo

Los resultados de la validez de constructo están relacionados con la formulación de reactivos, su clasificación y la evaluación del constructo. Frente a la formulación de reactivos se lograron identificar 110 resultados de aprendizaje y 4 indicadores binarios diferenciales por ítem, asociados al perfil del personal docente investigador en formación, a los criterios de evaluación y a las dimensiones de las competencias investigativas (saber, saber hacer y saber ser).

La identificación los ítems llevó a desarrollar un proceso de clasificación. La organización de reactivos por parte de docentes-personal experto y el grupo investigador, generó una matriz ordinal que se analizó por cada una de las 4 rutas de aprendizaje (planteamiento, referenciación, trabajo de campo, construcción para el futuro), acudiendo a la reducción de dimensiones como parte del método de extracción de componentes a través de clúster jerárquico con un número fijo de factores correspondientes a dos ([ver Figura 1](#)), se obtuvo la clasificación de los resultados de aprendizaje, teniendo en cuenta la organización constituida por personal docente experto, como parte de la argumentación de la clasificación. Se obtuvieron las siguientes agrupaciones por similitud:



Figura 1: Clasificación de resultados de aprendizaje (RA Cluster) por similitud vinculados a los criterios del momento de fundamentación en la formación investigativa



Nota: Elaboración propia.

En el momento de formación denominado referenciación, acudiendo a la técnica de análisis de distancia mínima o similitud máxima (clúster jerárquico), se obtuvieron 26 resultados de aprendizaje (R26) en seis categorías. Finalmente, la evaluación de la escala por parte del equipo evaluador (jueces-docentes y jueces-estudiantes) arrojó valores de pertinencia, suficiencia, claridad, nivel de dificultad, capacidad discriminante y objetividad. Es importante resaltar que los descriptores se evaluaron de forma nominal binaria por parte de 15 jueces. Inicialmente se identificaron los ítems evaluados con puntajes bajos ($x \leq 5$) medios ($6 \leq x \leq 10$) y altos ($11 \leq x \leq 15$); dada las características de evaluación de cada criterio de forma nominal y con más de dos jueces (Fleiss y Cohen, 1973), en la Tabla 2 se expone la estimación del grado de concordancia absoluta de cada criterio independientemente a través de índice de Kappa de Fleiss con intervalo un de confianza (IC) del 95 %:



Tabla 2: Índice Kappa de Fleiss para los descriptores de evaluación

	Kappa	Asymptotic Standard Error	Z	P Value	Lower 95% Asymptotic CI Bound	Upper 95% Asymptotic CI Bound
Pertinencia	,681	,018	38,230	,000	,646	,716
Claridad	,659	,018	36,971	,000	,624	,694
Nivel de dificultad	,707	,018	39,695	,000	,672	,742
Suficiencia	,648	,018	36,356	,000	,613	,683
Capacidad discriminante	,708	,018	39,750	,000	,673	,736
Objetividad	,697	,018	39,792	,000	,663	,732

Nota: Elaboración propia.

Los valores del índice Kappa de Fleiss (1981) para los criterios de evaluación por jueces, muestran índices de concordancia considerables (0,61 a 1) cuyos límites asintóticos se encuentran entre 0,61 y 0,80. Debido a las manifestaciones de los jueces y a la óptima concordancia en respuesta, se suprimieron los ítems con menor calificación en el criterio de pertinencia; se replanteó la escritura y redacción de los reactivos con valoraciones bajas en el criterio de claridad. Respecto al nivel de dificultad se identificaron los indicadores binarios diferenciales cuyas valoraciones fueron muy altas y muy bajas, lo que conllevó a reconsiderar los indicadores de alcance, criterio ligado directamente a la capacidad discriminante. Finalmente, para el criterio de objetividad, se identificaron los ítems con menor valoración y se reconsideraron los indicadores de logro ajustando los criterios de evaluación y definiendo, con mayor precisión, los alcances de la evaluación de la formación para la investigación educativa.

Fase relacional - consistencia interna

Posterior al ajuste (contenido y constructo) y basados en la clasificación de ítems por parte del equipo experto y la evaluación por parte de jueces, se continuó el proceso dando lugar a la ejecución del instrumento con la aplicación de una prueba piloto para corroborar su idoneidad. Cada indicador binario de la escala basada en Guttman arrojó un valor que, en conjunto, determinó el nivel de logro de los objetivos a evaluar. Este proceso permitió cuantificar el nivel de logro de cada resultado de aprendizaje a través de variables ordinales.

Los índices de consistencia interna se obtuvieron en un primer momento haciendo uso del método de las dos mitades de Guttman, debido a que el instrumento fue construido de forma fraccionada mediante las dimensiones de competencias (saber, saber hacer, saber ser). Correspondiendo al momento de fundamentación, se obtuvieron índices de consistencia interna 0,707 y 0,777 respectivamente. Posteriormente, para la estimación global de la solidez



de la escala se aplicó el método de consistencia interna de Cronbach, se obtuvo un valor alfa de fiabilidad superior a 0,8. Tras la evaluación de las propiedades métricas de la escala a través de los índices de fiabilidad se revelaron los resultados representados en la [Tabla 3](#):

Tabla 3: Estadísticas de fiabilidad. Índice de consistencia interna Dos mitades y Cronbach

Dos mitades	Parte 1	Valor	,707
		N de elementos	12a
	Parte 2	Valor	,777
		N de elementos	11b
	N total de elementos		23
Alfa de Cronbach		Valor	,861
		N de elementos	23
Correlación entre formularios			,812
Coeficiente de Spearman-Brown	Longitud igual		,896
	Longitud desigual		,896
Coeficiente de dos mitades de Guttman			,894
a. Los elementos son: P1, P2, P3, P4, P5, P6, P7, P8, P9, P10, P11, P12.			
b. Los elementos son: P12, P13, P14, P15, P16, P17, P18, P19, P20, P21, P22, P23.			

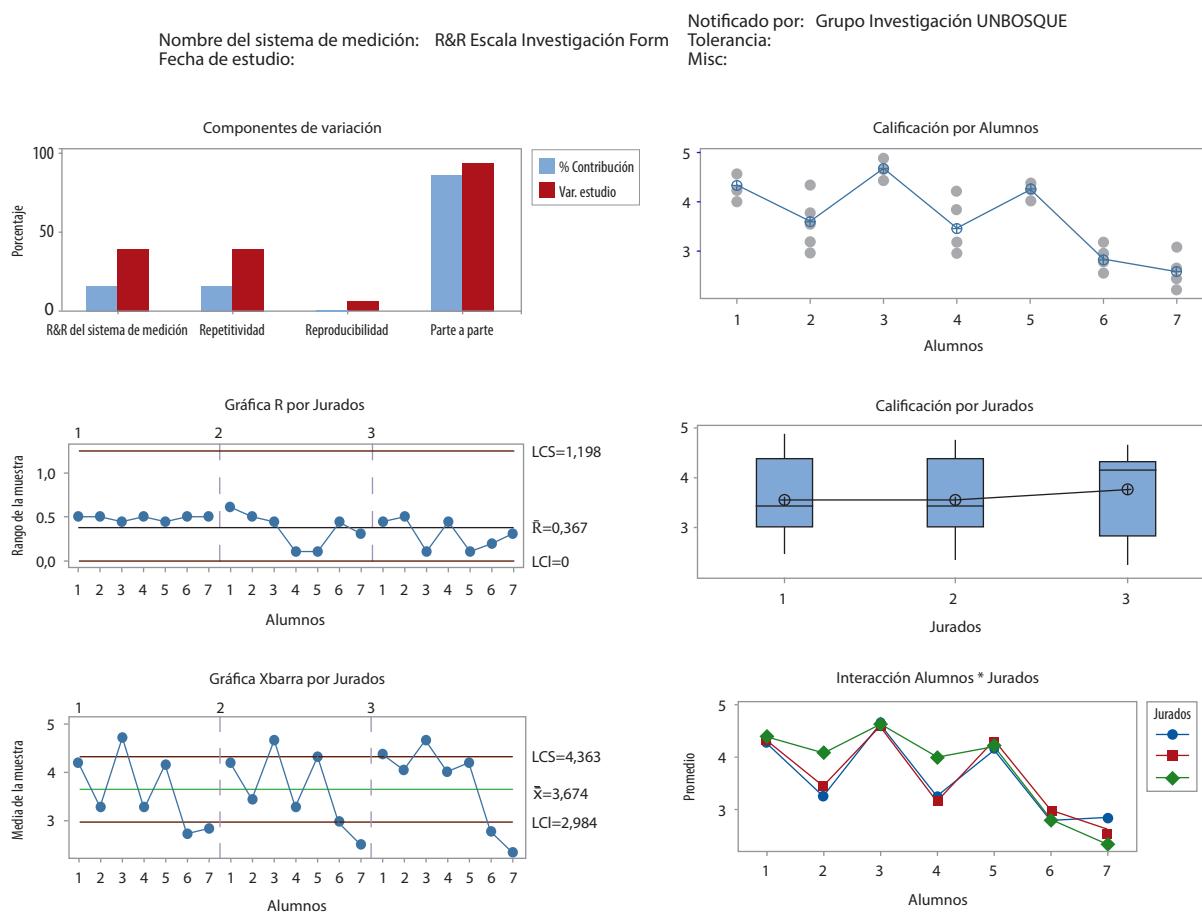
Nota: Elaboración propia.

De manera análoga, se procedió al análisis de fiabilidad “si se elimina el elemento [ítem]”; se observó que los coeficientes de confiabilidad de la escala, en tres de las cuatro rutas de evaluación formación investigativa, aumentaban ligeramente al eliminar alguno de los ítems. Los resultados evidencian que no cambia sustancialmente el índice de confiabilidad (0,861) se sugirió solo aplicar esta metodología en una de las secciones del instrumento, para fortalecer la validez de la escala.

Fase explicativa - estabilidad

En el proceso diagnóstico del presente estudio y durante el planteamiento de la investigación, se observó un alto índice de variabilidad en la valoración de la formación investigativa por parte de jurados externos que aplicaron el mismo instrumento de evaluación. Lo anterior, conllevó a plantear el diseño, construcción, validación y aplicación de la escala de valoración en mención. Como parte de la metodología de estudio se planteó el análisis de la estabilidad del instrumento, por medio del cálculo de patrones frente a los conceptos de repetitividad y reproducibilidad ([ver Figura 2](#)):



Figura 2: Informe de R&R del sistema de medición (ANOVA) para Calificación

Nota: Elaboración propia.

Los hallazgos más relevantes están vinculados a la variabilidad parte a parte proveniente de los grupos evaluados correspondiente a un 86,94 %. Esto conlleva a declarar que la escala de valoración de la formación en investigación educativa es satisfactoria debido a que el total de calibración corresponde a un 13,06 %, representado por un 4,58 % de la escala de valoración (repetitividad) y un 8,47 % concerniente al equipo evaluador (reproducibilidad), lo cual considera la mayor estabilidad en la sucesión de evaluación frente a la ejecución del instrumento. Esto conllevaría a pensar en un proceso de instrucción frente a la ejecución y aplicación de la escala de valoración. No obstante, los valores de estabilidad (R&R) presentados en la [Tabla 4](#), están conformes con la variabilidad natural del proceso de evaluación ($x=0,3588$).

Tabla 4: R&R del sistema de medición

Fuente	%Contribución CompVar	(de CompVar)
Gage R&R total	0,103095	13,06
Repetibilidad	0,036190	4,58
Reproducibilidad	0,066905	8,47
Jurados	0,006925	0,88
Jurados*Alumnado	0,059980	7,60
Parte a parte	0,686567	86,94
Variación total	0,789663	100,00

Nota: Elaboración propia.

Conclusiones

La ejecución de la escala para la valoración de la formación en investigación, basada en el modelo matemático en torno a las respuestas del ítem (IRT) y adaptada a la escala acumulativa de Guttman, proporciona una valiosa aproximación a la valoración del proceso de formación investigativa. Lo anterior está fundamentado en:

- Los constructos desarrollados, respecto a las dimensiones del saber, el hacer y el ser, resultan accesibles y valorables a través de la escala, puesto que la estimación acumulativa a través de indicadores binarios diferenciales permite la identificación de fortalezas y debilidades durante el proceso de formación.
- Valorar la formación en investigación educativa, por medio de constructos relacionados con competencias y reactivos asociados a resultados de aprendizaje, no solo permite evaluar productos de investigación sino también inferir en el desempeño centrado en el estudiantado durante el tiempo del proceso de formación; por tanto, la escala de valoración es un recurso que fortalece el acompañamiento y el seguimiento educativo, a través de categorías diferenciales que llevan a tomar acciones formativas y correctivas para la mejora de procesos de diagnóstico, intervención y evaluación.
- La capacidad discriminante está asociada a los niveles de logro del estudiantado y no presenta alta dependencia del ente evaluador (criterial e ideográfica) ya que a través de la escala se sugieren pautas y elementos iniciales que llevan a desarrollarse teniendo en cuenta el diagnóstico y las metas finales; así, el coeficiente de dificultad se obtiene en función del puntaje total alcanzado respecto al puntaje obtenido en la evaluación anticipatoria; de esta manera, a través de la escala, se calcula un valor ordinal, que identifica la relación de cada ítem respecto a la suma total del ítem.

- El muestreo teórico y el análisis de necesidades contextuales conllevaron a incorporar, fundamentalmente, en la escala de valoración, competencias asociadas a la ética investigativa, el trabajo en equipo, la comunicación argumentativa, el desarrollo de pensamiento crítico y la resolución de problemas; esto indica que es importante reconfigurar escenarios de prácticas investigativas como parte del perfil del estudiantado de licenciaturas en educación.
- De acuerdo con la evaluación de constructo, los índices de concordancia son pertinentes; esto nos llevó a reconfigurar y a estimar índices adecuados de pertinencia, suficiencia, claridad, nivel de dificultad, capacidad discriminante y objetividad en la escala de valoración.
- Es posible considerar que la escala tiene consistencia interna debido a que hay alta correlación entre ítems; es decir, la escala muestra un alto grado de homogeneidad según los análisis de fiabilidad realizados posterior a la prueba piloto; no obstante, es importante resaltar que la consistencia interna de la escala no es una propiedad inherente de esta misma; la consistencia interna responde a la medida patrón de las valoraciones obtenidas durante el proceso de formación para la investigación educativa.
- Los resultados obtenidos en el análisis *R&R* de la escala conllevan a considerar que existe proximidad en las medidas de dispersión evaluadas de manera longitudinal (cohortes académicas). Lo anterior supone cierto grado de independencia respecto al tiempo y quien la ejecuta. Debido a que los niveles de reproducibilidad (evaluación personal tutor/estudiantes) generó mayores niveles de dispersión, se estima llevar a cabo procesos de instrucción para la ejecución y aplicación de la escala.
- Teniendo en cuenta que mientras mayor sea la información que aporta determinado ítem (valoración del proceso formativo en función del aprendizaje) mayor es la rigurosidad de la estimación del constructo ([Pal et al., 2014](#)). Se presenta como producto de estudio la escala para la valoración de la formación en investigación educativa con 27 criterios de evaluación definidos en 110 resultados de aprendizaje determinados a través de 4 operadores binarios.

La argumentación teórica y contextual, la clasificación y organización de reactivos, la evaluación por parte de jueces con experticia, los resultados de consistencia interna y el análisis de estabilidad (*R&R*) permitieron fundamentar la validación de la escala. Sin embargo, en esencia el estudio busca contribuir y participar en el fortalecimiento y consolidación de los procesos de formación para la investigación educativa con base en la reflexión de la evaluación en virtud de las consideraciones plasmadas en el artículo. Conforme a ello, la formación investigativa supone alcanzar nuevos retos que sobrepasan las dimensiones del saber, y focalizan los esfuerzos en el desarrollo del ser y el hacer.

Declaración de Material complementario

Este artículo tiene disponible, como material complementario:

-La versión preprint del artículo en <https://doi.org/10.5281/zenodo.3421775>

Referencias

- Agencia Nacional de Evaluación de la Calidad y la Acreditación. (2013). *Guía para la redacción, puesta en práctica y evaluación de los resultados del aprendizaje*. Autor. <https://goo.gl/8Dg2Aa>
- Aular de Durán, J., Marcano, N. y Moronta, M. (2009). Competencias investigativas del docente de educación básica. *Revista Laurus*, 15(30), 138-165. <https://www.redalyc.org/pdf/761/76120651007.pdf>
- Balbo, J., Pacheco, M. y Rangel, Z. (2015). Medición de las competencias investigativas en los docentes adscritos al departamento de ciencias sociales de la Universidad Nacional Experimental del Táchira. *Aibi Revista de Investigación, Administración e Ingeniería*, 3(2), 27-35. <https://doi.org/10.15649/2346030X.513>
- Colás Bravo, M. P., Buendía Eisman, L. y Hernández Pina, F. (Coords.). (2009). *Competencias científicas para la realización de una tesis doctoral. Guía metodológica de elaboración y presentación*. Davinci.
- Consejo Nacional de Acreditación. (2013). *Lineamientos para la acreditación de programas de pregrado*. Autor. <https://goo.gl/JXiDaC>
- Córdoba, M. E. (2016). Reflexión sobre la formación investigativa de los estudiantes de pregrado. *Revista Virtual Universidad Católica del Norte*, 47, 20-37. <https://goo.gl/dCQ11U>
- Correa Bautista, J. E. C. (2009). Medición de las competencias investigativas en docentes de fisiología: Una aproximación empírica. *Revista de la Facultad de Medicina*, 57(3), 205-2017. <https://revistas.unal.edu.co/index.php/revfacmed/article/view/14378/15172>
- Da Cunha, M. I. (2015). Investigación y docencia: Escenarios y senderos epistemológicos para la evaluación de la educación superior. *REDU: Revista de Docencia Universitaria*, 13(1), 79-94. <https://doi.org/10.4995/redu.2015.6447>
- Fischer, G. H. y Ponocny-Seliger, E. (1998). *Structural rasch modeling, Handbook of the usage of LPCM-Win 1.0*. Progamma.
- Fleiss, J. L. (1981). *Statistical methods for rates and proportions*. John Wiley and Sons.

Fleiss, J. L. y Cohen, J. (1973). The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability. *Educational and Psychological Measurement*, 33(3), 613-619. <https://doi.org/10.1177/001316447303300309>

García Bejarano, A., Angarita, J. y Velandia, C. (2013). Implicaciones pedagógicas del uso de las TICs en la educación superior. *Revista de Tecnología*, 12(3), 36-56. [10.18270/rt.v12i3.1750](https://doi.org/10.18270/rt.v12i3.1750)

García-Gutiérrez, Z. del P. y Aznar-Díaz, I. (2019). El desarrollo de competencias investigativas, una alternativa para formar profesionales en pedagogía infantil como personal docente investigador. *Revista Electrónica Educare*, 23(1), 1-22. <https://doi.org/10.15359/ree.23-1.15>

González Campos, J., Carvajal Muquillaza, C. y Viveros Reyes, F. (2016). Coeficientes edumétricos para la validez y dificultad de un test: Propuesta. *Estudios Pedagógicos*, 42(3), 467-481. <https://doi.org/10.4067/S0718-07052016000400025>

González-Espino, Y. (2017). ¿Cómo evaluar la competencia investigativa desde la responsabilidad social universitaria? *Revista Cubana de Educación Superior*, 36(2), 4-13. http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S0257-43142017000200001

Hambleton, R. K., Swaminathan, H. y Rogers, H. J. (1991). *Fundamentals of item response theory* (2.^a ed.). Sage publications.

Jaik Dipp, A. (2013). *Competencias investigativas. Una mirada a la educación superior*. Redie. https://www.academia.edu/8510103/Competencias_investigativas._Una_mirada_a_la_educaci%C3%B3n_superior

Mesías Maraví, T. R. y Frisancho León, A. F. (2013). Evaluación psicométrica y evaluación edumétrica. *Investigación educativa*, 17(31), 93-108. <https://revistasinvestigacion.unmsm.edu.pe/index.php/educa/article/view/8027/7004>

Pal, M., Roy, U. K., Sau, S. y Datta, S. (2014). Formation of a scale for evaluation of thesis quality in md curriculum in west bengal. *Journal of Drug Delivery and Therapeutics*, 4(3), 129-131. <https://doi.org/10.22270/jddt.v4i3.865>

Piña-Osorio, J. M. y Aguayo-Rousell, H. B. (2017). La ética en la investigación de posgrado. *Revista Electrónica Educare*, 21(2), 1-25. <https://doi.org/10.15359/ree.21-2.12>

Reiban Barrera, R. E. R., De la Rosa Rodríguez, H. y Zeballos Chang, J. M. (2017). Competencias investigativas en la educación superior. *Revista Publicando*, 4(10), 395-405. <https://revistapublicando.org/revista/index.php/crv/article/view/439>

Romero, S. J., Ponsoda, V. y Ximénez, C. (2006). Validación de la estructura cognitiva del test de signos mediante modelos de ecuaciones estructurales. *Psicothema*, 18(4), 835-840. <https://goo.gl/Qcjqlu>



<http://doi.org/10.15359/ree.25-1.3>
<http://www.una.ac.cr/educare>
educare@una.ac.cr

Rojas Granada, C. y Aguirre Cano, S. (2015). La formación investigativa en la educación superior en América Latina y el Caribe: Una aproximación a su estado del arte. *Revista Eleuthera*, 12, 197-222. [10.17151/eleu.2015.12.11](https://doi.org/10.17151/eleu.2015.12.11)

Supo-Condori, J. (2013). Importancia del empleo de la bioestadística en las investigaciones biomédicas actuales. *Revista Médico-Científica "Luz y Vida"*, 4(1), 63-64. <https://www.redalyc.org/articulo.oa?id=325029251013>

Valdés Cuervo, Á. A., Carlos Martínez, E. A. y Torres Acuña, G. M. (2018). Propiedades psicométricas de una escala para medir cibervictimización en universitarios. *Revista Electrónica de Investigación Educativa*, 20(4), 36-48. <https://doi.org/10.24320/redie.2018.20.4.1841>

Velandia-Mesa, C., Serrano-Pastor, F. J. y Martínez-Segura, M. J. (2017). La investigación formativa en ambientes ubicuos y virtuales en Educación Superior. *Comunicar*, 25(51), 09-18. <https://doi.org/10.3916/C51-2017-01>

Velandia-Mesa, C., Serrano-Pastor, F. J. y Martínez-Segura, M. J. (2019). The challenge of competencies in training for educational research: A conceptual approach. *Revista Actualidades Investigativas en Educación*, 19(3), 1-27. <https://doi.org/10.15517/AIE.V19I3.38738>