



Revista Eureka sobre Enseñanza y Divulgación de las Ciencias  
ISSN: 1697-011X  
revista.eureka@uca.es  
Universidad de Cádiz  
España

## Problemas de validez y fiabilidad en los cuestionarios ROSE: revisión sistemática de la producción española

**Toma, Radu Bogdan**

Problemas de validez y fiabilidad en los cuestionarios ROSE: revisión sistemática de la producción española  
Revista Eureka sobre Enseñanza y Divulgación de las Ciencias, vol. 18, núm. 3, 2021  
Universidad de Cádiz, España  
Disponible en: <https://www.redalyc.org/articulo.oa?id=92066410005>  
DOI: [https://doi.org/10.25267/Rev\\_Eureka\\_ensen\\_divulg\\_cienc.2021.v18.i3.3102](https://doi.org/10.25267/Rev_Eureka_ensen_divulg_cienc.2021.v18.i3.3102)

## Problemas de validez y fiabilidad en los cuestionarios ROSE: revisión sistemática de la producción española

Validity and reliability issues in the ROSE questionnaires: A systematic review of Spanish literature

*Radu Bogdan Toma*

*Departamento de Didácticas Específicas. Área de Didáctica de las Ciencias Experimentales. Universidad de Burgos, España*  
rbtoma@ubu.es

 <https://orcid.org/0000-0003-4846-7323>

DOI: [https://doi.org/10.25267/Rev\\_Eureka\\_ensen\\_divulg\\_cienc.2021.v18.i3.3102](https://doi.org/10.25267/Rev_Eureka_ensen_divulg_cienc.2021.v18.i3.3102)

Rev\_Eureka\_ensen\_divulg\_cienc.2021.v18.i3.3102  
Redalyc: <https://www.redalyc.org/articulo.oa?id=92066410005>

Recepción: 04 Diciembre 2020

Revisado: 13 Enero 2021

Aprobación: 18 Marzo 2021

### RESUMEN:

Los cuestionarios ROSE ocupan un lugar destacado en la investigación relacionada con las actitudes hacia la ciencia, sin embargo, su análisis psicométrico en población de habla hispana ha sido descuidado. Esta revisión sistemática recopila estudios que han empleado los instrumentos ROSE y presenta un análisis de sus evidencias psicométricas. Los hallazgos revelan lagunas en términos de validez y fiabilidad. Concretamente, los resultados de validez estructural y fiabilidad documentados en la bibliografía son poco satisfactorios, insuficientes y contradictorios. Asimismo, existe una ausencia de aporte de evidencias fundamentales de validez (validez convergente, discriminante, concurrente y predictiva) y de fiabilidad test-retest (estabilidad temporal). Por lo tanto, la calidad psicométrica de los cuestionarios ROSE es preocupante y se desconoce en qué medida evalúan adecuadamente las actitudes hacia la ciencia.

**PALABRAS CLAVE:** Actitudes hacia la ciencia, imagen de la ciencia, ciencia escolar, ciencia y tecnología, educación científica.

### ABSTRACT:

The ROSE questionnaires stand out in research related to attitudes towards science; however, its psychometric analysis in the Spanish-speaking population has been neglected. This systematic review compiles research that used the ROSE instruments and provides an analysis of their psychometric evidence. The findings reveal gaps in terms of validity and reliability evidence. In particular, the results of structural validity and reliability documented in the literature are unsatisfactory, insufficient, and contradictory. Furthermore, fundamental psychometric properties are absent, such as convergent, discriminant, concurrent, and predictive validity; no evidence of temporal stability reliability was found neither. Therefore, the psychometric quality of the ROSE questionnaires is of concern and the extent to which they adequately measure attitudes towards science is unknown.

**KEYWORDS:** Attitudes toward science, image of Science, school science, science and technology, science education.

### INTRODUCCIÓN

*A menos que se tenga el cuidado necesario para determinar las características de rendimiento de los instrumentos utilizados en investigación, (...) los resultados en sí mismos no pueden considerarse dignos de confianza.* Munby (1983, p. 162)

El término *actitud hacia la ciencia* es un constructo amplio, de gran importancia en la educación científica, empleado para hacer referencias a las concepciones psicológicas y afectivas que los ciudadanos poseen sobre la ciencia (Tytler y Osborne 2012). Sin embargo, la investigación desarrollada en este campo se ha visto sometida a reiteradas críticas relacionadas con la poca validez y fiabilidad de los instrumentos de recolección de datos. En este sentido, Munby (1983) concluyó que la discrepancia de los resultados en esta línea de investigación se debe a problemas psicométricos de los cuestionarios empleados. Esta crítica

fue posteriormente respaldada por Blalock, Lichtenstein, Owen, Pruski, Marshal y Toepperwein (2008) al señalar que son pocos los instrumentos con la suficiente validez psicométrica como para poder ser utilizados con garantías. Recientemente, Toma (2020) y Toma y Lederman (2020) plantearon interrogantes sobre la validez y la fiabilidad de los instrumentos actitudinales publicados en las últimas dos décadas.

En la bibliografía española, los cuestionarios ROSE son indudablemente los más populares de su categoría, con numerosos estudios influyentes que los han empleado para estudiar las actitudes hacia la ciencia. Estos cuestionarios siguen estando ampliamente presentes en la investigación y se postulan como instrumentos válidos y fiables a pesar de que su análisis psicométrico se ha descuidado notablemente. Dado que las revisiones existentes de los instrumentos de actitudes hacia la ciencia no prestaron atención a los cuestionarios ROSE (v.g., Blalock *et al.* 2008, Toma 2020; Toma y Lederman 2020), el objetivo de este estudio es revisar y sintetizar las evidencias de validez y fiabilidad disponibles para estos instrumentos y proporcionar recomendaciones sobre su uso en la población de habla hispana.

Este propósito se abordará mediante un análisis exhaustivo de los estudios que hacen uso de los cuestionarios ROSE, analizando en qué medida han presentado u omitido aspectos sobre su validez y fiabilidad. Al leal saber del autor, el presente estudio es el primero en emprender tal esfuerzo. Los hallazgos de este trabajo ofrecen información valiosa para la investigación en didáctica de las ciencias española e hispanoamericana, máxime si se considera el gran impacto que estos cuestionarios han tenido hasta la fecha en el conocimiento que se posee sobre las actitudes hacia la ciencia del alumnado.

## FUNDAMENTOS TEÓRICOS

### Definición del constructo actitudes hacia la ciencia

El desarrollo de actitudes favorables hacia la ciencia se ha considerado perennemente un objetivo central de la educación científica (Tytler y Osborne 2012). La importancia de las actitudes ha cobrado impulso en los últimos años debido a la disminución del número de estudiantes interesados en proseguir una carrera relacionada con la ciencia (DeWitt y Archer 2015; Kennedy, Lyons y Quinn 2014). En la actualidad, la importancia de desarrollar una actitud positiva hacia la ciencia se refleja en la literatura que sugiere una estrecha relación entre las actitudes y los resultados académicos, las intenciones de seguir participando en experiencias de aprendizaje de la ciencia, y en la confianza y apoyo de los ciudadanos a la investigación científica (Besley 2016; Newell, Tharp, Vogt, Moreno y Zientek 2015).

El concepto de actitud ha sido sometido durante décadas a distintas conceptualizaciones. No obstante, una definición comúnmente adoptada en la literatura refiere a esta como una “(...) tendencia psicológica que se expresa mediante la evaluación de una entidad u objeto concreto, con cierto grado de favorabilidad o desfavorabilidad” (Eagly y Chaiken 1995, p. 414). Por lo tanto, las actitudes se conciben como evaluaciones favorables o desfavorables, de tipo interno y no observable, que se forman a medida que los individuos desarrollan creencias sobre el objeto de estudio (Ajzen 1991). En el contexto de las actitudes hacia la ciencia, se definen como la inclinación positiva o negativa del alumnado hacia el aprendizaje de la ciencia, las disciplinas científicas o la ciencia escolar (Tytler 2014). Asimismo, suelen concebirse como un constructo multidimensional que incluye tres componentes, a saber, el componente cognitivo, el afectivo, y el conativo o conductual (Khine 2015).

El primer componente comprende los pensamientos y creencias sobre la ciencia, reflejadas en ítems del tipo “La ciencia contribuye a mejorar la calidad de vida” o “El dinero que se invierte en ciencia es dinero bien invertido” (Navarro, Förster, González y González-Pose 2016, p. 1479-1480). Por otro lado, el componente afectivo refiere a los sentimientos y las emociones. En este sentido, ítems como “Pienso que Ciencias de la Naturaleza es aburrida” (Toma y Meneses-Villagrá 2019, p. 8) o “Mi clase de ciencia es interesante” (Wang

y Berlin 2010, p. 2423) evalúan la dimensión afectiva de las actitudes. Finalmente, la dimensión conativa se relaciona con las experiencias pasadas que afectan el comportamiento actual, y con las intenciones, tendencias y disposiciones conductuales hacia el objeto actitudinal investigado. Un ejemplo de ítems que miden esta dimensión podría ser “Me convertiré en un científico en el futuro” o “Me matricularé en cursos de ciencia adicionales en el futuro” (Summers y Abd-El-Khalick 2018, p. 196).

Dado que las actitudes deben ser evaluadas con respecto a un objeto de actitud determinado (v.g., la ciencia; los museos científicos) o una acción concreta (v.g., *estudiar* ciencias; *visitar* museos científicos) (Ajzen 1991), los didactas de las ciencias agruparon numerosas dimensiones actitudinales bajo el epígrafe común *actitudes hacia la ciencia*. Por ejemplo, Tytler y Osborne (2012) concluyeron que las investigaciones existentes se centraban en la medición de las actitudes hacia la ciencia, los científicos, la ciencia escolar, la medida en que los estudiantes disfrutan de las experiencias de aprendizaje de la ciencia, y sus intenciones de participar en actividades, estudios o carreras relacionadas con la ciencia. Análogamente, Reid (2015) sugirió que son cuatro los objetivos actitudinales de interés: la ciencia como disciplina, el aprendizaje de la asignatura de ciencias, el contenido de los cursos de ciencias, y los métodos de la ciencia.

## Los cuestionarios ROSE

Una cuestión importante en la evaluación de las actitudes hacia la ciencia es que se trata de un constructo que no es directamente observable y, por tanto, ha de ser inferido en base a la puntuación de un instrumento. Por ello, la existencia de herramientas de medición válidas y fiables constituye un aspecto fundamental en la investigación educativa. En este sentido, Schreiner y Sjøberg (2004) lideraron un esfuerzo internacional relacionado con la evaluación de las actitudes hacia la ciencia en el que participaron más de 40 países. El proyecto, denominado Relevancia de la Educación Científica (en inglés, *Relevance of Science Education* - ROSE), tenía por propósito recopilar información sobre las actitudes de alumnado de 15 a 16 años, matriculado en el último curso de la educación secundaria obligatoria.

El instrumento ROSE comprende un total de 245 ítems agrupados en 8 cuestionarios distintos. No obstante, para su uso en población española, se adaptaron al castellano únicamente cuatro de estos cuestionarios debido a que la administración completa del mismo no era viable de ser realizada en una sola sesión de clase<sup>[1]</sup>. Dos de estos cuestionarios están específicamente relacionados con las actitudes hacia la ciencia y, en consecuencia, son de interés para el presente estudio. El primero se denomina *Mis clases de ciencia* y comprende 18 ítems que proporcionan información sobre la percepción que poseen los estudiantes acerca de sus clases de ciencia.

El segundo cuestionario, titulado *Mis opiniones sobre la ciencia y la tecnología* consta de 16 ítems que sondean diferentes aspectos de la forma en que los estudiantes perciben el papel y la función de la ciencia y la tecnología en la sociedad. Todos los ítems se puntuán en una escala de tipo Likert con cuatro opciones de respuesta (1 – Nada de acuerdo; 4 – Totalmente de acuerdo). Si bien ambos cuestionarios constan en su versión original de 16 ítems cada uno, la versión en español de *Mis clases de ciencias* incluye dos ítems adicionales. El análisis de los estudios recuperados en la presente revisión falló en encontrar un fundamento o justificación para esta decisión.

## METODOLOGÍA

Se trata de una revisión sistemática de la literatura, acompañada de un comentario narrativo, elaborada conforme a las directrices de Petticrew y Roberts (2006), siendo este uno de los protocolos más consolidados en ciencias sociales. Los enfoques sistemáticos de las revisiones de la literatura recurren a un procedimiento de búsqueda, selección y síntesis de la bibliografía estructurado, transparente y cuidadosamente definido

que puede ser fácilmente reproducible, adoptando así una postura más objetiva en comparación con las revisiones no sistemáticas. De este modo, el presente estudio incluye una exhaustiva búsqueda bibliográfica, una selección de publicaciones basada en criterios de inclusión y exclusión, y una evaluación de las evidencias psicométricas recogidos en las investigaciones recuperadas. Con ello, se pretende dar una visión global de las evidencias de validez y fiabilidad existentes para los cuestionarios ROSE, proporcionando recomendaciones de uso y conclusiones libres de sesgo. El informe se presenta acorde a la declaración PRISMA (*Preferred Reporting Items for Systematic reviews and Meta-Analyses*).

## Estrategia de búsqueda bibliográfica

Los estudios potencialmente relevantes se recuperaron mediante una estrategia de búsqueda exhaustiva que incluye búsquedas en bases de datos y manuales. En primer lugar, se rastreó sistemáticamente en las bases de datos *Web of Science Core Collection*, *Scopus* y *Dialnet* estudios que investigaran de forma directa las actitudes hacia la ciencia. Los términos de búsqueda fueron ““attitud\* toward\* science” OR “actitud\* hacia la\* ciencia\*”, utilizados en el título de la publicación, el resumen y las palabras clave. Se recurrió a asteriscos (\*) para permitir declinaciones (v.g., actitud/es/inal).

Además, se efectuó una búsqueda manual en las siguientes revistas del ámbito de la didáctica de las ciencias en español: Enseñanza de las Ciencias, Revista Eureka sobre Enseñanza y Divulgación de las Ciencias, Revista Electrónica de Enseñanza de las Ciencias, Didácticas de las Ciencias Experimentales y Sociales, y Enseñanza de la Física. Por último, se adoptó un enfoque de ascendencia, que consiste en revisar el listado de referencias bibliográficas de los estudios recuperados para incluir aquellas investigaciones pertinentes que no han sido identificadas mediante la estrategia de búsqueda.

## Criterios de elegibilidad

Se han considerado relevantes para el objetivo de este estudio aquellas investigaciones que satisfacían todos los siguientes criterios de elegibilidad:

- Estudios publicados como artículos de revistas. Se excluyeron las actas de conferencias, capítulos de libros e informes.
- Publicados entre 2004 –año del primer informe ROSE (Schreiner y Sjøberg 2004)– y junio de 2020, momento en el que se realizó la búsqueda bibliográfica.
- Escritos en los idiomas inglés o español.
- Estudios que incluyen población española en la muestra.
- Emplean alguno de los dos cuestionarios actitudinales ROSE para la recolección de datos.

## Procedimiento de extracción de datos

La información relevante de los estudios incluidos en esta revisión se recopilaron mediante un formulario de extracción de datos que incluye: (i) la referencia del estudio, (ii) el cuestionario ROSE utilizado, (iii) el diseño del estudio, (iv) la muestra encuestada, (v) las dimensiones evaluadas según informan los autores de cada estudio, (vi) el número total de ítems administrados, (vii) el tipo y escala de respuesta empleada, y el reporte de (viii) evidencias de validez y (vix) fiabilidad. Los datos de los estudios se resumieron en distintas tablas (véase el Anexo 1) y, seguidamente, se sintetizaron de manera narrativa.

## Evaluación de las propiedades psicométricas

Este estudio se fundamenta el marco de referencia de Trochim y Donnelly (2006) para la evaluación de las propiedades psicométricas de los instrumentos en ciencias sociales. Por consiguiente, los instrumentos deben evaluarse en términos de validez de contenido (los ítems representan el constructo objeto de estudio) y validez del criterio (los ítems proporcionan las conclusiones que se esperan, basado en fundamentos teóricos). A continuación, se ofrece una breve definición de estos criterios psicométricos en el contexto de los cuestionarios objeto de estudio<sup>[2]</sup>.

Según el informe inicial del proyecto, la validez de contenido de los cuestionarios ROSE ha sido evaluada mediante un panel de expertos, discusión y entrevistas con los estudiantes, y tres aplicaciones piloto (Schreiner y Sjøberg 2004). Por lo tanto, se descarta el análisis de esta propiedad psicométrica del presente estudio. No obstante, Schreiner y Sjøberg (2004) no proporcionaron otras evidencias de validez y fiabilidad. Por tanto, y en consecuencia con el marco de referencia adoptado, se espera encontrar en la bibliografía de habla española relacionada con los cuestionarios ROSE, evidencias psicométricas pertenecientes a la validez del criterio, que incluye los siguientes aspectos. En primer lugar, los ítems del cuestionario deben reflejar la dimensionalidad hipotética del constructo objeto de estudio, a través de análisis factorial (validez estructural). Además, los ítems de una dimensión actitudinal (v.g., dimensión afectiva) debieran estar altamente correlacionados entre sí (validez convergente), y mínimamente correlacionados con los ítems de otras dimensiones distintas que evalúen otros aspectos (validez discriminante).

Por otro lado, las puntuaciones del instrumento han de distinguir entre los grupos de población que teóricamente difieren (validez concurrente; v.g., chicos y chicas) y predecir algo que teóricamente debiera predecir (validez predictiva; v.g., rendimiento académico). Por último, se debe aportar el coeficiente de alfa de Cronbach ( $\alpha$ ) para cada dimensión extraída mediante análisis factorial (consistencia interna), así como evidencias de estabilidad temporal (fiabilidad test-retest) en la medida en que el instrumento se vaya a emplear en estudios longitudinales o de intervenciones educativas con diseños pretest-posttest<sup>[3]</sup>.

## RESULTADOS

### Resultados de la búsqueda bibliográfica

La Figura 1 presenta el diagrama de flujo de selección de los estudios. La estrategia de búsqueda empleada en las bases de datos recuperó 2038 registros, de los cuales 1910 fueron excluidos por no cumplir los cuatro primeros cuatro criterios de elegibilidad (v.g., año de publicación, tipo de publicación, idioma, y muestra española). Los 128 artículos restantes fueron revisados mediante la lectura del apartado de metodología para comprobar si cumplían con el último criterio de elegibilidad (v.g., emplean algún cuestionario ROSE), de los cuales sólo 11 fueron retenidos. Mediante la búsqueda manual en revistas seleccionadas, se identificaron 5 nuevos artículos que satisfacían los criterios de inclusión. No se encontraron estudios relevantes al emplear el enfoque de ascendencia. De este modo, se han recopilado 16 estudios que emplearon los cuestionarios ROSE para evaluar actitudes hacia la ciencia en población española y, por consiguiente, fueron incluidos para su lectura en profundidad y evaluación de las evidencias psicométricas.

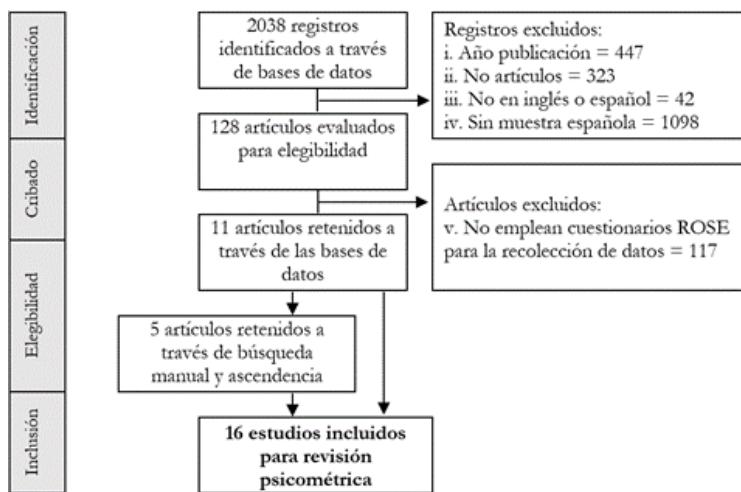


FIGURA 1  
Diagrama PRISMA de selección de estudios

## Evidencias de validez y fiabilidad

La Figura 2 sintetiza las evidencias de validez y fiabilidad identificadas en los estudios analizados. En el Anexo se recogen dos tablas exhaustivas que amplían estos datos. En cuanto a la validez de ambos cuestionarios ROSE, apenas dos estudios han aportado evidencias de validez estructural para cada uno de ellos, con resultados insatisfactorios. Empleando análisis de componentes principales, Vázquez y Manassero (2005) reportaron una estructura de cuatro dimensiones para los 18 ítems originales del cuestionario *Mis clases de ciencias*, sin embargo, un total de siete ítems fueron retenidos a pesar de presentar cargas cruzadas en dos o más factores, siendo por tanto candidatos para su eliminación de la solución latente final. De manera similar, Vázquez y Manassero (2004) retuvieron los ítems del cuestionario *Mis opiniones sobre la ciencia y la tecnología* en cuatro dimensiones, aunque tres de estos presentaron cargas cruzadas en dos factores. Al eliminar estos ítems problemáticos, ambos cuestionarios presentan dimensiones con apenas dos ítems, siendo esta una cantidad inferior al mínimo recomendado para instrumentos de medidas de autoinforme (DeVellis 2017).

Además de insatisfactorias, las evidencias de validez estructural resultan contradictorias. En un estudio posterior, Vázquez y Manassero (2009a) sometieron nuevamente ambos cuestionarios a análisis factorial, reteniendo esta vez hasta un total de 20 dimensiones que difieren de las obtenidas en estudios anteriores. Este hecho revela que los cuestionarios ROSE se ven afectados por las falacias denominadas Jingle-Jangle, que consisten en que cuestionarios nombrados de la misma manera miden, en realidad, dimensiones distintas (falacia Jingle), o que los cuestionarios miden la misma dimensión a pesar de ser nombrados de manera diferente (falacia Jangle). De este modo, mientras que en Vázquez y Manassero (2005) los ítems del cuestionario *Mis clases de ciencias* son empleados para medir finalidades educativas, agrado por asignaturas de ciencias, vocaciones científico-tecnológicas, o facilidad de aprendizaje de la ciencia, en Vázquez y Manassero (2009a) estos mismos ítems han sido agrupados en una sola dimensión relacionada con percepciones de la ciencia escolar. Asimismo, en Vázquez y Manassero (2004) el cuestionario *Mis opiniones sobre la ciencia y la tecnología* miden progreso social, problemas sociales, cientifismo, e impactos negativos, sin embargo, en Vázquez y Manassero (2009a) se consideran reflejo de otras cuestiones como, por ejemplo, imagen de progreso de la ciencia y la tecnología, o imagen social de la ciencia y la tecnología.

	Estructural	Convergente	Discriminante	Concurrente	Predictiva	Consistencia interna	Estabilidad temporal
Vázquez y Manassero (2004)	✓					✓	
Vázquez y Manassero (2005)	✓					✓	
Vázquez y Manassero (2008a)							
Vázquez y Manassero (2008b)							
Vázquez y Manassero (2009a)	✓					✓	
Vázquez y Manassero (2009b)							
Vázquez y Manassero (2009c)							
Vázquez y Manassero (2009d)						✓	
Vázquez y Manassero (2010)							
Marbá-Tallada y Márquez (2010)							
Vázquez y Manassero (2011)						✓	
Fernández-Cézar y Pinto-Solano (2017)						✓	
Mazas y Bravo Torija (2018)							
Fernández-Cézar et al. (2018)						✓	
Dapía et al. (2019)						✓	
Queiruga-Díos et al. (2020)							

FIGURA 2  
Síntesis de las evidencias de validez y fiabilidad

En relación con la fiabilidad de los instrumentos ROSE, seis estudios han aportado el Alfa de Cronbach ( $\alpha$ ) para el cuestionario *Mis clases de Ciencias* y cinco para *Mis opiniones sobre la ciencia y la tecnología*. No obstante, al igual que sucede con las evidencias de validez estructural, los valores aportados son insuficientes y, en definitiva, poco concluyentes. Por ejemplo, para la estructura de cuatro dimensiones del cuestionario *Mis clases de Ciencias*, compuesta por 18 ítems, Vázquez y Manassero (2004) obtuvieron valores marginales para dos dimensiones ( $\alpha = 0.56$  y  $\alpha = 0.66$ ), claramente por debajo de los estándares psicométricos contemporáneos (Taber 2018; Tavakol y Dennick 2011). Estos valores han sido más aceptables en investigaciones que han empleado menos ítems y que han considerado el cuestionario como reflejo de una sola dimensión. Así, Vázquez y Manassero (2009d) han obtenido un  $\alpha = 0.89$  para 15 ítems, y Fernández-Cézar, Pinto-Solano y Muñoz-Hernández (2018) reportaron un  $\alpha = 0.81$  para 8 ítems. Sin embargo, es necesario señalar que los resultados derivados del análisis factorial sugieren que este cuestionario mide, al menos, cuatro dimensiones distintas y, por tanto, no debería emplearse como una escala unidimensional.

La fiabilidad del cuestionario *Mis opiniones sobre la ciencia y la tecnología* también es insuficiente. La estructura de cuatro dimensiones presenta valores para el Alfa de Cronbach que oscilan entre 0.32 y 0.77 (Vázquez y Manassero 2004). Al emplearse este cuestionario como una única escala unidimensional, contraria a los resultados del análisis factorial, los valores de fiabilidad apenas mejoran. Así, Vázquez y Manassero (2009a, 2009d) reportan un exiguo  $\alpha = 0.57$ , y Dapía, Escudero-Cid y Vidal (2019) un valor residual de  $\alpha = 0.75$ . Considerando que el valor de Alfa de Cronbach depende en gran medida del número total de ítems (v.g., a mayor número de ítems, mayor será su valor), y que este indicador aporta resultados poco fiables en instrumentos de tipo Likert, estos hallazgos no son satisfactorios (Tavakol y Dennick 2011). Por último, no se han encontrado evidencias para los restantes tipos de validez, ni tampoco evidencias para la fiabilidad de temporal (v.g., test-retest) de ambos cuestionarios. Asimismo, cabe mencionar que ningún estudio aportó evidencias de fiabilidad basadas en el coeficiente Omega de McDonald. Por consiguiente, se

desconoce en qué medida los instrumentos ROSE poseen validez convergente, discriminante, concurrente y predictiva, así como cuál es su estabilidad temporal y consistencia interna.

## DISCUSIÓN

Esta revisión sistemática evaluó las propiedades psicométricas de los cuestionarios ROSE a partir de 16 estudios identificados por medio de una búsqueda bibliográfica exhaustiva. En términos generales, el análisis de la producción española revela que las evidencias de validez y fiabilidad de los cuestionarios *Mis clases de ciencia* y *Mis opiniones sobre la ciencia y la tecnología* no se ajustan a las recomendaciones contemporáneas. En efecto, si bien los estudios examinados presentaron los cuestionarios ROSE como instrumentos válidos y fiables para medir actitudes hacia la ciencia, en su mayor parte, esta afirmación solo se vio respaldada por un aporte insuficiente de evidencias de validez y fiabilidad. Por consiguiente, la calidad psicométrica de los instrumentos ROSE es motivo de preocupación y es discutible en qué medida evalúan adecuadamente las actitudes de la población de habla hispana. Si se atiende al presente análisis, los hallazgos sugieren que las evidencias psicométricas publicadas no son suficientes de acuerdo con los estándares actuales y, por tanto, ambos instrumentos parecerían producir resultados cuya validez y fiabilidad se desconoce.

## Propiedades psicométricas de los cuestionarios ROSE

En general, las evidencias de validez identificadas en la bibliografía fueron insuficientes, encontrándose estructuras factoriales contradictorias y cuya validez convergente y discriminante no ha sido explorada. Tanto si los instrumentos ROSE se utilizan en estudios observacionales o cuasiexperimentales, la puntuación derivada de su uso posee un significado sobre la que los investigadores forman un juicio o una interpretación de las actitudes de los individuos. Sin embargo, sin un conocimiento sólido de sus estructuras latentes, la validez de dichas conclusiones sería limitada. Del mismo modo, aunque los instrumentos fueron extensivamente empleados para hacer comparaciones entre grupos de interés (v.g., niñas y niños, o alumnos de grados inferiores y superiores del sistema educativo), no se encontró ninguna evidencia de validez concurrente para ninguno de los dos cuestionarios analizados. Análogamente, ambos cuestionarios se utilizaron en numerosas investigaciones de tipo predictivo-transversal; sin embargo, en ninguno de los estudios examinados se midió la capacidad de los cuestionarios para predecir algo que teóricamente debería poder predecir, esto es, su validez predictiva.

Por otro lado, las evidencias de fiabilidad fueron asimismo insuficientes, con resultados muy inferiores a los valores recomendados. Muchos estudios presentaron un valor moderado o bajo para el alfa de Cronbach, sin embargo, el índice se calculó para todos los ítems asumiendo una estructura unidimensional que contradice los resultados del análisis factorial. En la bibliografía psicométrica se ha comprobado que los cuestionarios compuestos por una alta cantidad de ítems arrojan un valor de Alfa aceptable en la mayoría de los casos, con independencia de si el instrumento está compuesto por una o más dimensiones (Taber 2018). Por consiguiente, la fiabilidad ha de ser evaluada mediante el alfa de Cronbach únicamente si, en primera instancia, se ha confirmado la unidimensionalidad del instrumento mediante análisis factorial (Tavakol y Dennick 2011). Si el análisis factorial muestra la existencia de múltiples dimensiones, como en el caso de ambos cuestionarios ROSE, este índice debe calcularse para cada subescala/dimensión por separado y no para todo el cuestionario (DeVellis 2017). Desafortunadamente, los resultados de esta revisión muestran que sólo dos estudios se adhirieron a este criterio, con resultados poco satisfactorios (v.g., Vázquez y Manassero 2004, 2005). Además, ningún estudio ha evaluado la fiabilidad de los instrumentos ROSE empleando el coeficiente Omega de McDonald, siendo este un estadístico más apropiado para las características de los cuestionarios evaluados y, por lo tanto, un aspecto a considerar en futuros estudios.

Por último, en el presente análisis también se comprobó que ningún estudio presentó evidencias de estabilidad temporal para los cuestionarios ROSE, a pesar de que estos fueron empleados en varias investigaciones de tipo manipulativo cuasiexperimental que requieren de un instrumento con una robusta fiabilidad temporal para obtener datos fidedignos. Aquellos estudios cuyo diseño constan de varias fases de recopilación de datos (v.g., diseños quasi-experimentales de tipo pretest – posttest), requieren del uso de un instrumento que reproduzca sistemáticamente las actitudes de los estudiantes, suponiendo que estas no han cambiado a lo largo del tiempo (Toma y Lederman 2020). Sin evidencias de fiabilidad test-retest, se desconoce si los datos proporcionados por los instrumentos son una representación exacta de las actitudes de los estudiantes o si, por el contrario, son fruto de una falta de estabilidad temporal. Por consiguiente, el uso de instrumentos sin fiabilidad temporal puede enmascarar (v.g., errores de tipo II o falsos negativos) o potenciar (v.g., errores de tipo I o falsos positivos) los verdaderos efectos de una intervención educativa.

## Limitaciones potenciales

El procedimiento empleado en este estudio para evaluar las evidencias psicométricas de los cuestionarios ROSE puede estar sujeto a debate. Específicamente, algunos investigadores podrían argumentar que los estándares de calidad psicométrica aquí adoptados son demasiado exigentes. No obstante, cabe señalar que este estudio es consistente con otras revisiones de la literatura, e inclusive podría considerarse menos estricto. Por ejemplo, si se emplea la rúbrica de Blalock *et al.* (2008) para evaluar las evidencias de validez y fiabilidad existentes para los cuestionarios ROSE, se observa que ambos obtendrían apenas 3 de los 28 puntos posibles, un valor que se sitúa muy por debajo de la mediana de 9.5 puntos obtenidos por la mayoría de los 62 instrumentos analizados por estos autores. También cabe resaltar que el procedimiento de evaluación empleado se fundamenta en un marco de referencia consolidado en didáctica de las ciencias (v.g., Trochim y Donnelly 2006), y que las propiedades psicométricas examinadas son coherentes con los *Estándares para las evaluaciones educativas y psicológicas* (AERA, APA y NCME 2014), según los cuales, se necesita aportar evidencias para todos los tipos distintos de validez y fiabilidad a menos que se especifique claramente por qué son irrelevantes, inapropiados o técnicamente inviables para una determinada población o uso.

## Directrices para investigaciones futuras

Este estudio proporciona una mirada profunda sobre las actuales evidencias de validez y fiabilidad de los cuestionarios ROSE (*Mi clase de ciencias y Mis opiniones sobre la ciencia y la tecnología*), y señala las direcciones futuras que han de abordarse para estudiar más a fondo sus propiedades psicométricas. Así, es necesario examinar la validez convergente, discriminante, concurrente y predictiva de ambos. Sin embargo, esto requiere, en primer lugar, de resultados robustos y no contradictorios con respecto a la validez estructural de los cuestionarios (DeVellis 2017). De manera similar, es necesario determinar la estabilidad temporal de ambos cuestionarios. Ello permitiría que se pudiesen utilizar de manera fiable en estudios de intervención o longitudinales que requieran varias fases de recolecta de datos. Sin este aspecto, los efectos reales de las intervenciones educativas podrían quedar camuflados y no será posible determinar con certeza en qué medida produjeron el resultado deseado. Por último, los cuestionarios ROSE fueron principalmente empleados con estudiantes de 15-16 años de edad. A pesar de que algunos estudios recolectaron datos en educación primaria o niveles universitarios (v.g., Dapía, Escudero-Cid y Vidal 2019; Mazas y Bravo-Torija 2018), las evidencias de validez estructural de ambos cuestionarios identificadas son aplicables únicamente a la población de educación secundaria. Por lo tanto, si se pretenden emplear en otras etapas educativas, es necesario desarrollar, en primera instancia, un análisis exhaustivo de sus propiedades psicométricas en la nueva población objeto de estudio.

## Significancia de este estudio

Esta revisión sistemática es una contribución singular y original que pone de manifiesto varias lagunas relacionadas con la validez y fiabilidad de los dos instrumentos más utilizados en la investigación sobre las actitudes hacia la ciencia en población española. Las implicaciones educativas originarias en esta línea de investigación están sujetas a la robustez psicométrica de los instrumentos de medida empleados. Dado que los cuestionarios analizados presentan evidencias psicométricas insuficientes, la calidad de y la confianza que se puede depositar en las conclusiones derivadas de su uso podrían estar en entredicho.

Asimismo, esta revisión señala las cuestiones que deben mejorarse en lo que respecta a la validez y fiabilidad de los cuestionarios ROSE e invita a los investigadores a someter estos instrumentos a rigurosos procedimientos de validación que puedan ayudar a verificar o refutar muchas de las afirmaciones que se realizaron en los estudios analizados sobre estas herramientas y acerca de las actitudes de los estudiantes españoles hacia la ciencia. De acuerdo con la calidad psicométrica que figura en la bibliografía, no se pueden ofrecer recomendaciones inequívocas sobre su uso, pues el problema fundamental reside en que ambos cuestionarios no poseen propiedades psicométricas satisfactorias. Por tanto, su uso, en su forma actual, debería evitarse hasta que no se avancen sólidas evidencias de validez y fiabilidad que indiquen lo contrario.

Únicamente con instrumentos validados y fiables se podrá determinar si las innovaciones educativas promocionadas en didáctica de las ciencias (v.g., indagación, enfoque STEM, aprendizaje basado en problemas) mejoran realmente las actitudes de los estudiantes hacia la ciencia. En consecuencia, se espera que este artículo inicie un diálogo fructífero entre los investigadores de España interesados en mejorar los estándares de calidad psicométrica de los instrumentos de autoinforme empleados. En este sentido, se subraya que, si bien este estudio se ha enfocado en los instrumentos ROSE, debido a que son, con mucho, los más empleados en el contexto español, la necesidad de comprobar la robustez psicométrica es relevante para cualquier otro instrumento que se decida emplear. Así, los cuestionarios ROSE no son más que un ejemplo entre los muchos instrumentos existentes en la producción española que afirman medir las actitudes hacia la ciencia a pesar de no aportar evidencias suficientes sobre su validez y fiabilidad.

## MATERIALES SUPLEMENTARIOS

Anexo I (pdf)

## REFERENCIAS

- AERA, APA, NCME. (2014) *Standards for educational and psychological testing*. Washington, DC: AERA.
- Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes*, 50 (2), 179–211.
- Besley J. C. (2016) The National Science Foundation's science and technology survey and support for science funding, 2006–2014. *Public Understanding of Science*, 27, 94–109.
- Blalock, C. L., Lichtenstein, M. J., Owen, S., Pruski, L., Marshall, C., Toepperwein, M. A. (2008) In pursuit of validity: A comprehensive review of science attitude instruments 1935–2005. *International Journal of Science Education*, 30 (7), 961–977. <https://doi.org/10.1080/09500690701344578>
- \*Dapía, M., Escudero-Cid, R., Vidal, M. (2019) ¿Tiene género la ciencia? Conocimientos y actitudes hacia la Ciencia en niñas y niños de Educación Primaria. *Revista Eureka Sobre Enseñanza y Divulgación de Las Ciencias*, 16 (3), 3201. [https://doi.org/10.25267/Rev\\_Eureka\\_ensen\\_divulg\\_cienc.2019.v16.i3.3302](https://doi.org/10.25267/Rev_Eureka_ensen_divulg_cienc.2019.v16.i3.3302)

- DeVellis, R. F. (2017) *Scale development. Theory and applications*. Los Angeles: SAGE.
- DeWitt, J., Archer, L. (2015) Who aspires to a science career? A comparison of survey responses from primary and secondary school students. *International Journal of Science Education*, 37 (13), 2170-2192. <https://doi.org/10.1080/09500693.2015.1071899>
- Eagly, A. H., Chaiken, S. (1995) Attitude strength, attitude structure, and resistance to change. En R. E. Petty y J. A. Krosnick (Eds.), *Ogio State University series on attitudes and persuasion, Vol. 4. Attitude strength: Antecedents and consequences* (pp. 413-432). New York: Lawrence Erlbaum Associates, Inc.
- \*Fernández-Cézar, R., Pinto-Solano, N. (2017) Actitud hacia las clases de ciencias naturales en la educación primaria en España. *Revista Electrónica de Investigación Educativa*, 19 (4), 112-123. <https://doi.org/10.24320/redie.2017.19.4.1393>
- \*Fernández-Cézar, R., Pinto-Solano, N., Muñoz-Hernández, M. (2018) ¿Mejoran los proyectos de divulgación con experimentación la actitud hacia las clases de ciencias? *Revista de Educación*, 381, 285-307. <https://doi.org/10.4438/1988-592X-RE-2017-381-389>
- Kennedy, J. P., Lyons, T., Quinn, F. (2014) The continuing decline of science and mathematics enrolments in Australian high schools. *Teaching Science*, 60 (2), 34-46.
- Khine, S. M. (2015) *Attitude measurements in science education: Classic and contemporary approaches*. Charlotte, NC: Information Age Publishing, INC.
- \*Marbá-Tallada, A., Márquez, B. C. (2010) ¿Qué opinan los estudiantes de las clases de ciencias? Un estudio transversal de sexto de primaria a cuarto de ESO. *Enseñanza de Las Ciencias*, 28 (1), 19-30.
- \*Mazas, B., Bravo Torija, B. (2018) Actitudes hacia la ciencia del profesorado en formación de educación infantil y educación primaria. *Profesorado, Revista de Currículum y Formación Del Profesorado*, 22 (2), 329-348. <https://doi.org/10.30827/profesorado.v22i2.7726>
- Munby, H. (1983) Thirty studies involving the "Scientific Attitude Inventory": What confidence can we have in this instrument? *Journal of Research in Science Teaching*, 20 (2), 141-162.
- Navarro, M., Förster, C., González, C., González-Pose, P. (2016) Attitudes toward science: Measurement and psychometric properties of the test of science-related attitudes for its use in Spanish-speaking classrooms. *International Journal of Science Education*, 38 (9), 1459-1482. <https://doi.org/10.1080/09500693.2016.1195521>
- Newell, A. D., Tharp, B. Z., Vogt, G. L., Moreno, N. P., Zientek, L. R. (2015) Students' attitudes toward science as predictors of gains on student content knowledge: Benefits of an after-school program. *School science and Mathematics*, 115 (5), 216-225. <https://doi.org/10.1111/ssm.12125>
- Petticrew, M., Roberts, R. (2006) *Systematic reviews in the social sciences: A practical guide*. Oxford: Blackwell.
- \*Queiruga-Dios, M. Á., López-Iñesta, E., Diez-Ojeda, M., Sáiz-Manzanares, M. C., Dorrío, J. B. V. (2020) Citizen science for scientific literacy and the attainment of sustainable development goals in formal education. *Sustainability*, 12 (10), 4283. <https://doi.org/10.3390/su12104283>
- Reid, N. (2015) Attitude research in science education. En M.S. Khine (ed.), *Attitude measurements in science education. Classic and contemporary approaches* (pp. 3-46). Charlotte, NC: Information Age Publishing, INC
- Schreiner, C., Sjøberg, S. (2004) *ROSE: The Relevance of Science Education. Sowing the seeds of Rose. Background, rationale, questionnaire development and data collection for ROSE (The Relevance of Science Education) - a comparative study of students' views of science and science education*. Oslo: Acta Didactica.
- Summers, R., Abd-El-Khalick, F. (2018) Development and validation of an instrument to assess student attitudes toward science across grades 5 through 10. *Journal of Research in Science Teaching*, 55 (2), 172-205. <https://doi.org/10.1002/tea.21416>
- Taber, K. S. (2018) The use of Cronbach's Alpha when developing and reporting research instruments in science education. *Research in Science Education*, 48 (6), 1273-1296. <https://doi.org/10.1007/s11165-016-9602-2>
- Tavakol, M., Dennick, R. (2011) Making sense of Cronbach's alpha. *International Journal of Medical Education*, 2, 53-55. <https://doi.org/10.5116/ijme.4dfb.8dfd>

- Toma, R. B. (2020) Revisión sistemática de instrumentos de actitudes hacia la ciencia (2004-2016). *Enseñanza de las Ciencias*, 38 (3), 143-159. <https://doi.org/10.5565/rev/ensciencias.2854>
- Toma, R. B., Lederman, N. G. (2020) A comprehensive review of instruments measuring attitudes toward science. *Research in Science Education*. <https://doi.org/10.1007/s11165-020-09967-1>
- Toma, R. B., Meneses-Villagrá, J. A. (2019). Validation of the single-items Spanish-school science attitude survey (S-SSAS) for elementary education. *PLoS ONE*, 14 (1), e0209027. <https://doi.org/10.1371/journal.pone.0209027>
- Trochim, W. M., Donnelly, J. P. (2006) *The research methods knowledge base*. Cincinnati: Atomic Dog Publishing Inc.
- Tytler, R., Osborne, J. (2012) Student attitudes and aspirations towards science. En B. J. Fraser, K. G. Tobin, y C. J. McRobbie (Eds.), *Second International Handbook of Science Education* (pp. 597-625). Netherland: Springer.
- Tytler, R. (2014) Attitudes, identity, and aspirations toward science. En N. G. Lederman y S. K. Abell (Eds.), *Handbook of research on science education, II Edition* (pp. 82-103). New York: Routledge.
- \*Vázquez, A., Manassero, M. A. (2004) Imagen de la ciencia y la tecnología al final de la educación obligatoria. *Cultura y Educación*, 16 (4), 385-398.
- \*Vázquez, A., Manassero, M. A. (2005) La ciencia escolar vista por los estudiantes. *Bordón. Revista de Pedagogía*, 57 (5), 125-144.
- \*Vázquez, A., Manassero, M. A. (2008a) El declive de las actitudes hacia la ciencia de los estudiantes: un indicador inquietante para la educación científica. *Revista Eureka Sobre Enseñanza y Divulgación de Las Ciencias*, 5 (3), 274-292. [https://doi.org/10.25267/rev\\_eureka\\_ensen\\_divulg\\_cienc.2008.v5.i3.03](https://doi.org/10.25267/rev_eureka_ensen_divulg_cienc.2008.v5.i3.03)
- \*Vázquez, A., Manassero, M. A. (2008b) La vocación científica y tecnológica de las chicas en secundaria y la educación diferenciada. *Bordón. Revista de Pedagogía*, 60 (3), 149-163.
- \*Vázquez, A., Manassero, M. A. (2009a) Factores actitudinales determinantes de la vocación científica y tecnológica en secundaria. *Cultura y Educación*, 21 (3), 319-330. <https://doi.org/10.1174/113564009789052280>
- \*Vázquez, A., Manassero, M. A. (2009b) La relevancia de la educación científica: actitudes y valores de los estudiantes relacionados con la Ciencia y la Tecnología. *Enseñanza de Las Ciencias*, 27 (1), 33-48.
- \*Vázquez, A., Manassero, M. A. (2009c) La vocación científica y tecnológica: predictores actitudinales significativos. *Revista Eureka Sobre Enseñanza y Divulgación de Las Ciencias*, 6 (2), 213-231. [https://doi.org/10.25267/rev\\_eureka\\_ensen\\_divulg\\_cienc.2009.v6.i2.03](https://doi.org/10.25267/rev_eureka_ensen_divulg_cienc.2009.v6.i2.03)
- \*Vázquez, A., Manassero, M. A. (2009d) Patrones actitudinales de la vocación científica y tecnológica en chicas y chicos de secundaria. *Revista Iberoamericana de Educación*, 50 (4), 1-15.
- \*Vázquez, A., Manassero, M. A. (2010) Perfiles actitudinales de la elección de ciencias en secundaria según el sexo y el tipo de educación. *Revista Electrónica de Enseñanza de Las Ciencias*, 9 (1), 242-260.
- \*Vázquez, A., Manassero, M. A. (2011) El descenso de las actitudes hacia la ciencia de chicos y chicas en la Educación Obligatoria. *Ciência & Educação*, 17 (2), 249-268.
- Viladrich, C., Angulo-Brunet, A., Doval, E. (2017) Un viaje alrededor de alfa y omega para estimar la fiabilidad de consistencia interna. *Anales de Psicología*, 33(3), 755-782. <http://dx.doi.org/10.6018/analesps.33.3.268401>
- Wang, T., Berlin, D. (2010) Construction and validation of an instrument to measure Taiwanese elementary students' attitudes toward their science class. *International Journal of Science Education*, 32 (18), 2413-2428. <https://doi.org/10.1080/09500690903431561>
- Watkins, M. W. (2017) The reliability of multidimensional neuropsychological measures: From alpha to omega. *The clinical neuropsychologist*, 31(6-7), 1113-1126. <http://dx.doi.org/10.1080/13854046.2017.1317364>

## NOTAS

- [1] Véase el informe español para más detalles (p. 2): <https://roseproject.no/network/countries/spain/spain.html>
- [2] Una explicación detallada de estos conceptos está fuera del alcance de este trabajo, por lo que se invita al lector a consultar Trochim y Donnelly (2006)

- [3] La exactitud del coeficiente Alfa de Cronbach depende de varios supuestos que se vulneran en la gran mayoría de los cuestionarios actitudinales, especialmente al utilizar ítems de tipo Likert con menos de siete alternativas, siendo este el caso de los cuestionarios ROSE (Viladrich, Angulo-Brunet y Doval 2017). Un índice de fiabilidad más apropiado para este tipo de instrumentos sería el coeficiente Omega de McDonald (Watkins 2017)

## INFORMACIÓN ADICIONAL

*Para citar este artículo:* Toma R. B. (2021) Problemas de validez y fiabilidad en los cuestionarios ROSE: revisión sistemática de la producción española. *Revista Eureka sobre Enseñanza y Divulgación de las Ciencias* 18(3), 3102. doi: 10.25267/Rev\_Eureka\_ensen\_divulg\_cienc.2021.v18.i3.3102