

Exámenes de certificación evaluados con casos clínicos seriados: planteamiento teórico y problemática

Víctor H. Olmedo-Canchola,¹ * Pierre Jean-Aurelius,²  Miguel Á. Vásquez-Luna³  y José I. Santos-Preciado⁴

¹División de Estudios de Posgrado, Facultad de Medicina, Universidad Nacional Autónoma de México; ²Servicio de Trasplantes, Hospital de Pediatría, Centro Médico Nacional Siglo XXI, Instituto Mexicano del Seguro Social; ³Comité Normativo Nacional de Consejos de Especialidades Médicas; ⁴Unidad de Medicina Experimental, Facultad de Medicina, Universidad Nacional Autónoma de México. Ciudad de México, México

Resumen

Se presenta un planteamiento teórico de las desventajas de realizar los exámenes de certificación de las especializaciones médicas mediante casos clínicos seriados y analizarlos con la Teoría Clásica del Test (TCT). Los exámenes de certificación de las especializaciones médicas corresponden a exámenes sumativos criterioles de altas consecuencias, por lo tanto, es imperativo mantener los máximos estándares de calidad en todos los procesos implicados en su elaboración. Actualmente, se ha detectado que algunos consejos de especialidades médicas realizan los exámenes de certificación con pruebas que incluyen casos clínicos seriados y que el análisis psicométrico está sustentado en la TCT; sin embargo, la estructura de este tipo de pruebas infringe los supuestos fundamentales de dicha teoría. La violación de los supuestos fundamentales de la TCT en las pruebas construidas mediante casos clínicos seriados puede conducir a sesgos o interpretación errónea de los resultados. Por lo anterior, es recomendable utilizar otros modelos psicométricos para el análisis de este tipo de pruebas, o evitar el uso de los casos clínicos seriados en los exámenes de certificación de las especialidades médicas.

PALABRAS CLAVE: Certificación médica. Consejos médicos. Evaluación médica. Psicometría. Teoría Clásica del Test. Testlet.

Certification exams evaluated with serial clinical cases: theoretical and problematic approach

Abstract

The objective of this issue is a theoretical approach to the disadvantages involved in taking the certification exams of medical specializations through serial clinical cases and analyzing them with the Classical Test Theory (TCT). The certification exams of medical specializations correspond to criterial summative exams with high consequences. Therefore, it is imperative to maintain the highest quality standards in all the processes involved in preparing the exam. At present, it has been detected that some councils of medical specialties conduct the certification exams with tests that contain serial clinical cases and the psychometric analysis that they conduct is supported by the TCT; however, the structure of this type of test violates the fundamental assumptions of the TCT. The violation of the fundamental assumptions of the TCT in the tests constructed through serial clinical cases can lead to biases or misinterpretations of the results. Due to the above, it is advisable to use other psychometric models for the analysis of this type of test, or to avoid the use of serial clinical cases in the certification exams of medical specialties.

KEYWORDS: Medical certification. Medical council. Medical evaluation. Psychometrics. Classical test theory. Testlet.

*Correspondencia:

Víctor H. Olmedo-Canchola

E-mail: victor.olmedo@fmposgrado.unam.mx

0016-3813/© 2024 Academia Nacional de Medicina de México, A.C. Publicado por Permanyer. Este es un artículo *open access* bajo la licencia CC BY-NC-ND (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Fecha de recepción: 11-10-2023

Fecha de aceptación: 27-03-2024

DOI: 10.24875/GMM.23000411

Gac Med Mex. 2024;160:128-134

Disponible en PubMed

www.gacetamedicademexico.com

Antecedentes

El objetivo de los exámenes de certificación en las especialidades médicas es identificar a los examinados que poseen los conocimientos mínimos necesarios para ejercer la especialidad correspondiente dentro de los estándares de calidad y seguridad para el paciente.

En México, el Comité Normativo Nacional de Consejos de Especialidades Médicas (CONACEM), constituido por la Academia Nacional de Medicina, la Academia Mexicana de Cirugía y los consejos de especialidades miembros, tiene la naturaleza de organismo auxiliar de la administración pública federal a efecto de supervisar los procesos de certificación y recertificación de los diferentes especialistas de la medicina a través de los consejos de especialidades médicas (el CONACEM es la instancia encargada de otorgar la declaratoria de idoneidad a los consejos de especialidades médicas, y los consejos integrantes del CONACEM son los únicos acreditados para la expedición de los certificados de especialidad). Lo anterior ha quedado establecido en el artículo 81 de la Ley General de Salud.¹

El CONACEM se ha dado a la tarea de identificar áreas de oportunidad para la mejora de la calidad en los procesos de certificación y recertificación por examen que aplican los consejos de especialidades médicas; una de esas áreas de oportunidad corresponde a la utilización de casos clínicos seriados en los exámenes y sus problemas asociados.

El objetivo de este ensayo es analizar, desde un punto de vista teórico, las desventajas que implica realizar exámenes de certificación mediante casos clínicos seriados en las especialidades médicas, particularmente cuando se trata de exámenes de altas consecuencias, tales como la certificación y recertificación en una especialidad. La idea primigenia de este trabajo apareció en 2019, cuando se presentaron resultados preliminares en el 2° Congreso de Evaluación: Competencias e Innovación en Evaluación,² en la Facultad de Medicina de la Universidad Nacional Autónoma de México.

Casos clínicos seriados

Las pruebas que se elaboran con base en casos clínicos seriados generan grupos de preguntas a partir de una viñeta clínica; es decir, son preguntas de opción múltiple asociadas a un caso. Habitualmente, constan

de una breve presentación de un caso, seguida de varias preguntas de opción múltiple (generalmente cinco). En teoría, cada pregunta se refiere a un aspecto distinto del caso, ya que se considera la situación clínica desde diferentes perspectivas; sin embargo, en la realidad pocas veces sucede de esta forma.

En 1987, Wainer y Kiely propusieron el nombre *testled* para este tipo de pruebas, y señalaron que el examen incluye diversos elementos al interior de un subconjunto que comparten un solo estímulo de contenido;³ es decir, se conforma un miniexamen dentro de un macroexamen. Los ítems que comparten el mismo estímulo se agrupan en una unidad y se les denomina *testlet* o preguntas seriadas. Dicho de otro modo, el conjunto de ítems (*testlet*), o preguntas seriadas, corresponde a una subunidad de puntuación dentro de una prueba mayor. Una muestra de un caso clínico seriado se presenta en la Tabla 1.

Cuando los casos clínicos seriados están bien elaborados y apegados a las tablas de especificaciones⁴ presentan ventajas y desventajas.

Ventajas de los casos clínicos seriados

- Ayudan a explorar la toma de decisiones en el examinado, por ejemplo, decidir qué paso se debe seguir en el manejo de un paciente.
- Permiten enfocarse en problemas potencialmente graves o comunes.
- Evitan situaciones clínicas que deberían ser tratadas por un subespecialista.
- Son de gran utilidad en las evaluaciones formativas.

Desventajas de los casos clínicos seriados

- Conforme se incrementa el número de preguntas seriadas o el número de distractores, su elaboración se hace cada vez más difícil.
- Puede haber información cruzada que proporcione “pistas” al sustentante. La información cruzada implica que los datos de un reactivo pueden, inadvertidamente, proporcionar información que ayude a obtener la respuesta correcta de otra pregunta.
- Los elementos dentro de los *testlets* o preguntas seriadas dependen localmente entre sí, porque están asociados a un mismo estímulo.
- La dependencia de elementos locales introduce dimensiones no intencionadas.⁵
- Dado que los reactivos seriados comparten una relación fuerte, se infringe el supuesto de independencia local de la Teoría de Respuesta al Ítem (TRI) y de la independencia global de la Teoría Clásica de los Test (TCT).⁶

Tabla 1. Ejemplo de caso clínico seriado

<p>Caso clínico para evaluar diabetes en el embarazo Mujer de 37 años, nulípara; sin antecedentes familiares ni personales de interés. Vida sexual activa con medidas anticonceptivas de barrera de forma irregular. Acude a la consulta por amenorrea de dos semanas; menciona que sus ciclos menstruales siempre han sido muy regulares (28 × 3 días). A la exploración física, destaca tensión arterial de 120/65 mm Hg, índice de masa corporal de 30.12 kg/m². La exploración por aparatos es normal sin datos relevantes.</p>
<p>Con estos datos, ¿cuál considera que sea el mejor método para diagnosticar o descartar embarazo en esta paciente? Cuantificación en sangre de la fracción beta de la gonadotropina coriónica. Prueba cualitativa en orina de la fracción beta de gonadotropina coriónica. Cuantificación en orina de la fracción beta de la gonadotropina coriónica. Ultrasonido pélvico. A través de síntomas.</p>
<p>Según las guías de práctica clínica, si estuviera embarazada la paciente, ¿cómo la clasificaríamos en función del riesgo para desarrollar diabetes a lo largo del embarazo? Sin riesgo. Riesgo bajo. Riesgo moderado. Riesgo alto. Riesgo moderado-bajo.</p>
<p>¿Cuándo se le debe realizar el tamizaje con la prueba de O'Sullivan? En la semana 24-28 de gestación En la primera visita y entre las semanas 24-28 y 32-36 de gestación. En la primera visita y entre las semanas 24-28 de gestación. Entre las semanas 24-28 y 32-36 de gestación. En la primera visita y entre las semanas 32-36 de gestación.</p>
<p>En caso de que la paciente tuviera diabetes gestacional, ¿cuál sería el tratamiento farmacológico de primera elección para el control glucémico? Glibenclamida. Acarbosa. Metformina Insulina. Vildagliptina.</p>

Métodos fundamentales para el análisis psicométrico de una prueba

Los métodos para analizar los reactivos varían de acuerdo con las necesidades de la evaluación, las características del estudio y las capacidades para generar el análisis. Las dos corrientes principales para el análisis psicométrico de las pruebas son la TCT y la TRI.⁷

La TCT es, probablemente, el modelo de análisis más conocido en la psicometría. Este modelo clásico fue introducido por el psicólogo británico Charles

Spearman, quien, a principios del siglo XX, publicó una serie de argumentos matemáticos en los que expuso que las calificaciones de una prueba son medidas poco precisas de los rasgos humanos. A lo largo del tiempo, la TCT ha avanzado y se ha adaptado a las necesidades del momento; sin embargo, sus principios básicos se siguen manteniendo como la forma de análisis más popular de la evaluación educativa moderna, lo cual no necesariamente implica que sea la mejor o la más adecuada en todos los casos.⁸

La TRI es un conjunto de modelos matemáticos que describen el desempeño de los ítems. A diferencia de la TCT, esta teoría se enfoca en el reactivo propiamente dicho y no en la prueba. Los modelos de la TRI intentan descubrir la relación entre la habilidad del examinado y la probabilidad de responder correctamente el ítem, y describir dicha relación con base en un modelo matemático.⁹

La razón por la cual se emplea con frecuencia la TCT (y probablemente el único método que conocen algunos evaluadores) es la facilidad para conseguir sus índices (*dificultad* y *discriminación*); sin embargo, el uso indiscriminado de cualquier modelo matemático puede conducir al desarrollo de sesgos al tratar de analizar algo con un modelo que no es el adecuado. Precisamente eso sucede cuando se pretende analizar los casos clínicos seriados con el modelo de la TCT.

TCT versus TRI

En esta sección se indican algunas ventajas conceptuales y prácticas de los modelos TRI sobre los modelos TCT con respecto a la construcción y el desarrollo de pruebas o escalas. El objetivo principal es señalar que los modelos TRI tienen características únicas que complementan las medidas basadas en TCT en lugar de comparar exhaustivamente los dos enfoques.

En primer lugar, una limitación del TCT estriba en que no son distinguibles las características del ítem y de la persona, como los parámetros de dificultad del ítem y las puntuaciones de la persona. Dependiendo de la subpoblación en cuestión, las características de los reactivos pueden cambiar. Si se considera una subpoblación de alta capacidad para la seriación de reactivos, todos los ítems de la prueba parecerían fáciles. Pero cuando se considera una subpoblación con baja capacidad para la seriación, el mismo conjunto de ítems sería difícil. Esta limitación dificulta la evaluación de las capacidades de los individuos mediante el uso de diferentes versiones de las pruebas. En la TRI, las características de los ítems y

las habilidades personales se formulan mediante parámetros distintivos. Una vez calibrados los ítems para una población, las puntuaciones de los sujetos de esa población se pueden comparar directamente, incluso si responden a diferentes subconjuntos de ítems. Algunos investigadores se refieren a esta cualidad como la *propiedad invariante* de los modelos TRI.¹⁰

En segundo lugar, la definición de *confiabilidad* en la TCT se basa en pruebas paralelas, difíciles de lograr en la práctica. La precisión de la medición es la misma para todas las puntuaciones de una muestra particular. En la TCT, las pruebas más largas suelen ser más fiables que las pruebas más cortas. Sin embargo, la confiabilidad en la TRI se define como una función condicionada a las puntuaciones del constructo latente medido. La precisión de la medición difiere a lo largo del continuo del constructo latente y puede generalizarse a toda la población objetivo. En la TRI, la precisión de las mediciones suele representarse mediante curvas de información. Estas curvas pueden tratarse como una función del factor latente condicionado a los parámetros del ítem, y se pueden calcular para un ítem individual (curva de información del ítem) o para toda la prueba (curva de información de la prueba). La curva de información de la prueba se puede utilizar para evaluar el desempeño de la prueba.¹⁰

Múltiples resultados de las pruebas

Al planificar cualquier prueba para evaluar los conocimientos de una persona, es fundamental recordar que solo se va a obtener lo que la prueba permite observar. Para conocer la puntuación verdadera o la habilidad del examinado se deberán tomar en consideración otros aspectos (Figura 1):

- *Puntuación observada*: lo que realmente se obtiene en una prueba.
- *Puntuación verdadera*: la puntuación que probablemente se obtendría teniendo en cuenta la incertidumbre y el error de la estimación.
- *Habilidad*: lo que realmente puede hacer o saber el sustentante acerca de un dominio, independiente de lo que está en una prueba.

Relación entre los resultados de la prueba y las puntuaciones

La habilidad es relativamente constante o invariante y solo cambia gradualmente con el aprendizaje y la enseñanza (*prueba independiente*).

Por su parte, la puntuación real varía según la dificultad de los ítems seleccionados en la prueba (*prueba dependiente*).

Las puntuaciones observadas varían según la calidad de los ítems y su precisión en la medición del constructo (*prueba dependiente*).

Teoría clásica del test (TCT)

- Ha servido bien para analizar las pruebas durante la mayor parte del siglo XX.
- Implica suposiciones teóricas relativamente débiles-fáciles de cumplir.
- Es fácil de aplicar, ya que las medidas estadísticas se pueden calcular manualmente.
- Su limitación principal estriba en que su enfoque es a nivel de la prueba (*test*) como un todo.

El problema del resultado de las pruebas

Con la finalidad de exponer la problemática al realizar el análisis psicométrico de las pruebas y cómo interpretar sus resultados, a continuación se presenta un ejemplo hipotético.

Ejemplo: Los examinados A, B y C sustentan el mismo examen. La prueba consta de 10 reactivos de opción múltiple con una sola respuesta correcta para cada reactivo. Los tres sustentantes responden seis respuestas correctas de las 10 preguntas planteadas. ¿Qué conclusión podemos extraer?

Resultados bajo la perspectiva de la TCT

Véase la Figura 2. Bajo la TCT, se considera que cada ítem es igual de difícil y tiene el mismo peso en la puntuación total. La puntuación total basada en la suma de los reactivos correctos es una buena estimación de la habilidad real, por lo que podría inferirse que los estudiantes con igual puntuación son igualmente capaces. Sin embargo, ¿qué sucede si los reactivos no tienen el mismo grado de dificultad? Ante este hecho, las cosas se complican en virtud de que el enfoque clásico implica que los ítems tienen el mismo peso en la composición de las estadísticas y puntuaciones.

Los ítems de la prueba son como ladrillos en una pared. Los reactivos solo significan algo en el contexto de la prueba en la que están y deben ser una muestra aleatoria del dominio que se está evaluando (Figura 3).

Los estadísticos básicos son el promedio de la prueba (media), la dispersión de las puntuaciones (desviación estándar) y la distribución de los ítems hacia las personas (discriminación).



Figura 1. Los múltiples resultados de las pruebas.

Dificultades

Las medidas estadísticas de personas e ítems son dependientes de la muestra. La puntuación total es simplemente la suma del número de ítems contestados correctamente

Si se analiza una prueba construida con casos clínicos seriados (Figura 4), con la TCT se incurre en múltiples fallos a los supuestos que sustentan esta teoría, entre ellos:

- Los ítems están relacionados, por lo tanto no corresponden a muestras aleatorias del dominio que se está evaluando.
- Las estadísticas que podrían usarse son el análisis de covarianza y el análisis factorial.
- La distribución de los ítems y las personas se tendría que analizar para cada caso clínico seriado.
- Las estadísticas de personas e ítems son dependientes de la muestra de sujetos examinados y de los casos clínicos seriados.
- La puntuación total no debe ser simplemente la suma del número de ítems contestados correctamente

Discusión

Con la finalidad de percibir el problema de los casos clínicos seriados, podría realizarse el ejercicio de imaginar una prueba con dos casos clínicos seriados (Figura 5), de cada uno de los cuales se desprenden cinco preguntas. El caso clínico 1 corresponde a un enfermo con rinofaringitis, respecto al cual se exploran aspectos relacionados con el diagnóstico, fisiopatología, tratamiento, prevención y pronóstico. Dado que la rinofaringitis es una enfermedad con una prevalencia elevada y ubicua, los examinados estarán más familiarizados con esta patología, las preguntas les resultarán familiares y, por lo tanto, más fáciles de

Ítem	1	2	3	4	5	6	7	8	9	10	porcentaje
Sujeto A	✓	✓	✓	✓	✓	✓					60
Sujeto B	✓		✓	✓	✓	✓		✓			60
Sujeto C			✓		✓		✓	✓	✓	✓	60

Figura 2. Resultados de un examen bajo la perspectiva de la Teoría Clásica de los Test.

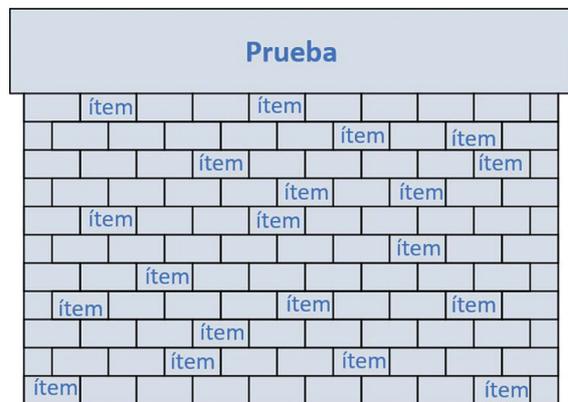


Figura 3. Prueba con enfoque de la Teoría Clásica de los Test.

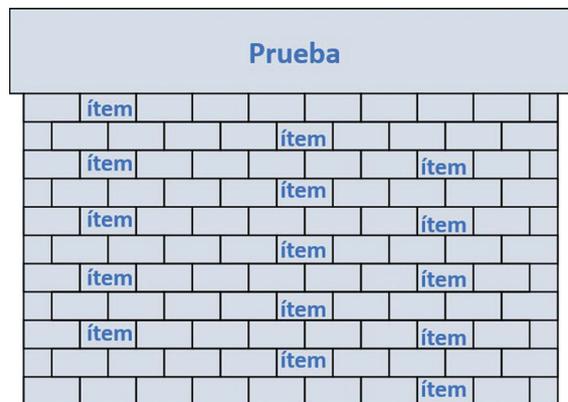


Figura 4. Prueba con casos clínicos seriados.

responder correctamente. El caso clínico 2 corresponde a dengue hemorrágico, del que se exploran las

Ítem	RINOFARINGITIS Caso clínico 1					DENGUE HEMORRAGICO Caso clínico 2					porcentaje
	1	2	3	4	5	6	7	8	9	10	
Sujeto											
A	✓	✓	✓	✓	✓	✓					60
B	✓		✓	✓	✓	✓		✓			60
C			✓		✓		✓	✓	✓	✓	60
	Diagnóstico	Fisiopatología	Tratamiento	Prevención	Pronóstico	Diagnóstico	Morfología	Fisiopatología	Tratamiento	Complicaciones	

Figura 5. Caso hipotético de una prueba con dos casos clínicos seriados.

áreas de diagnóstico, morfología, fisiopatología, tratamiento y complicaciones. Puesto que el dengue hemorrágico es endémico, su incidencia en algunas zonas es muy baja o, incluso, nula, lo que genera que las preguntas resulten con mayor grado de dificultad para los examinados que no habitan en zonas donde se presenta, y viceversa.

Al comparar el comportamiento de tres sujetos examinados, el examinado A respondió acertadamente a todas las preguntas acerca de la rinofaringitis; sin embargo, con relación al dengue únicamente contestó bien la pregunta sobre diagnóstico. Por otro lado, el sujeto C contestó correctamente las preguntas referentes al tratamiento y al pronóstico de la rinofaringitis; mientras que respecto al dengue hemorrágico, la única errónea fue la relativa al diagnóstico.

Reflexionemos un poco sobre el sujeto C. No resulta lógico que alguien desconozca el diagnóstico de una patología y pueda saber sobre morfología, tratamiento, fisiopatología, pronóstico y complicaciones. Resulta ilógico porque todo gira en torno a saber, de primera instancia, cuál es el diagnóstico. Con base en ello, la única explicación de esta situación es que el examinado C acertó a causa del azar. Por lo tanto, sustentar el análisis psicométrico de esta prueba bajo el modelo de la TCT conducirá irremediablemente a múltiples sesgos y fallos al intentar obtener los índices de dificultad, discriminación, análisis de distractores y confiabilidad.

Conclusiones

Bajo un planteamiento teórico, analizar las pruebas construidas con casos clínicos seriados mediante la TCT viola los supuestos fundamentales de la esta teoría y de la TRI, lo que podría conducir a sesgos o a malas interpretaciones de los resultados. Por tanto, es recomendable utilizar otros modelos estadísticos para el análisis de este tipo de pruebas, tales como los siguientes:

- *Análisis de subconjuntos*:¹¹ dado que cada caso clínico puede considerarse como un subconjunto de la prueba, se puede realizar un análisis independiente que permita identificar patrones específicos o tendencias en las respuestas de los sustentantes. Esto puede incluir la revisión de las respuestas a cada pregunta dentro del subconjunto, así como la correlación entre las respuestas a las diferentes preguntas dentro del mismo caso clínico.
- *Análisis factorial*:¹² el análisis factorial es una técnica estadística multivariante que se aplica a un único conjunto de variables cuando el investigador está interesado en determinar qué variables del conjunto forman subconjuntos lógicos relativamente independientes entre sí. En otras palabras, el análisis factorial es particularmente útil para identificar los factores subyacentes a las variables mediante la agrupación de variables relacionadas en el mismo factor.

- *Cálculo de los índices psicométricos por reactivo encadenado*:¹³ esto permite verificar si el sustentante acertó el primer reactivo y obtener probabilidades condicionales; por ejemplo, ¿cuál es la probabilidad de responder correctamente el/los subsecuente(s)?

Si no es factible usar la prueba estadística adecuada, lo mejor es evitar construir pruebas de altas consecuencias (como la certificación de especialistas) con base en casos clínicos seriados.

Agradecimientos

Los autores agradecen al doctor Francisco Antonio Casasola Méndez, médico pasante en servicio social, por el apoyo con el formato del presente artículo.

Financiamiento

No se recibió financiamiento público ni privado.

Conflicto de intereses

Los autores declaran que no existe ningún conflicto de intereses (económico, profesional o personal) real, potencial o potencialmente percibido que pueda resultar en un sesgo en la publicación de este trabajo.

Responsabilidades éticas

Protección de personas y animales. Los autores declaran que para esta investigación no se realizaron experimentos en seres humanos ni en animales.

Confidencialidad de los datos. Los autores declaran que en este artículo no aparecen datos de pacientes.

Derecho a la privacidad y consentimiento informado. Los autores declaran que en este artículo no aparecen datos de pacientes.

Uso de inteligencia artificial para generar textos.

Los autores declaran que no han utilizado ningún tipo de inteligencia artificial generativa en la redacción de este manuscrito ni para la creación de figuras, gráficos, tablas o sus correspondientes pies o leyendas.

Bibliografía

1. Acuerdo por el que se emiten los lineamientos a que se sujetarán el Comité Normativo Nacional de Consejos de Especialidades Médicas y los Consejos de Especialidades Médicas a los que se refiere el artículo 81 de la Ley General de Salud, para la aplicación de lo dispuesto por el artículo 272 bis y el título cuarto de dicha Ley. Diario Oficial de la Federación. 2015 Mar 25. Disponible en: https://www.dof.gob.mx/nota_detalle.php?codigo=5386686&fecha=25/03/2015#gsc.tab=0
2. Olmedo-Canchola VH, Heinze-Martin G, Fuentes-Venegas A. Problemática de la evaluación médica con casos clínicos seriados: planteamiento teórico. Trabajo oral presentado en el "2º Congreso de evaluación: competencias e innovación en evaluación". México: Facultad de Medicina, Universidad Nacional Autónoma de México; 2019. Disponible en: <https://docplayer.es/172323605-Memorias-trabajos-libres-carteles-y-mesas-de-trabajo.html>
3. Wainer H, Bradlow ET, Wang X. What's a testlet and why do we need them? En: Testlet Response Theory and Its Applications. Reino Unido: Cambridge University Press; 2007. pp. 44-59.
4. Aiken LR. Test psicológicos y evaluación. Décima primera edición. México: Pearson Educación; 2003.
5. Baker FB. The Basics of Item Response Theory. Segunda edición. Maryland, Estados Unidos: ERIC Clearinghouse on Assessment and Evaluation; 2001
6. ASC [Internet]. Minnesota, Estados Unidos: Thompson N. What is a testlet? Assessment Systems Corporation; 2024. Disponible en: <https://assess.com/what-is-a-testlet/#:-:text=In%20the%20real%20of%20education%20and%20assessment%2C%20the,an%20individual%E2%80%99s%20abilities%20compared%20to%20traditional%20testing%20methods>
7. Brown GT. Seminar at EARLI SIG 1. Summer School. August 2018, Helsinki, Finlandia.
8. Olea J, Ponsoda V. Tests adaptativos informatizados. Madrid, España: UNED-Universidad Nacional de Educación a Distancia; 2013.
9. Dogan E. The future is here: the (new) standards for educational and psychological testing. 2016. Disponible en: https://www.niss.org/sites/default/files/news_attachments/New%20standards%20Jan%2012%202015%20final_0.pdf
10. SAS [Internet]. Carolina del Norte, Estados Unidos: An X, Yung YF. Item response theory: what it is and how you can use the IRT procedure to apply it. SAS Institute Inc.; 2014. Disponible en: <https://support.sas.com/resources/papers/proceedings14/SAS364-2014.pdf>
11. Hastie T, Tibshirani R, Tibshirani R. best subset, forward stepwise or lasso? Analysis and recommendations based on extensive comparisons. Statist Sci. 2020;35(4):579-592. DOI: 10.1214/19-STS733
12. Shrestha N. Factor analysis as a tool for survey analysis. Am J Appl Math Stat. 2021;9(1):4-11. DOI: 10.12691/ajams-9-1-2
13. Rohmer J. Uncertainties in conditional probability tables of discrete bayesian belief networks: a comprehensive review. Eng Appl Artif Intell. 2020;88: 103384. DOI: 10.1016/j.engappai.2019.103384