

Análisis del rendimiento estudiantil y propiedades psicométricas de cuestionarios de autoevaluación de Moodle en Metabolismo-Nutrición.

Analysis of student performance and psychometric properties of Moodle self-assessment quizzes in Metabolism-Nutrition.

Raisa Rodríguez Hernández,¹ Yasnay Jorge Sainz,² Jackeline López Báster,³ Mariela Dieguez Martínez,⁴ Sonia Navarro Arrieta,⁵ Alberto Alberteris Rodríguez.⁶

¹Doctora en Medicina. Especialista en Medicina General Integral y en Bioquímica Clínica. Máster en Ciencias. Profesor Auxiliar. Investigador Agregado. Universidad de Ciencias Médicas de Holguín. Holguín, Cuba. <https://orcid.org/0000-0002-7361-4324>

²Doctora en Medicina. Especialista en Medicina General Integral y en Bioquímica Clínica. Máster en Ciencias. Profesor Auxiliar. Investigador Agregado. Universidad de Ciencias Médicas de Holguín. Holguín, Cuba. <https://orcid.org/0000-0002-0755-727X>

³Doctora en Medicina. Especialista en Medicina General Integral y en Bioquímica Clínica. Máster en Ciencias. Profesor Asistente. Investigador Agregado. Universidad de Ciencias Médicas de Holguín. Holguín, Cuba. <https://orcid.org/0000-0002-9096-9071>

⁴Doctora en Medicina. Especialista en Medicina General Integral y en Bioquímica Clínica. Máster en Ciencias. Profesor Auxiliar. Investigador Auxiliar. Universidad de Ciencias Médicas de Holguín. Holguín, Cuba. <https://orcid.org/0000-0002-1673-8128>

⁵Doctora en Medicina. Especialista en Medicina General Integral y en Bioquímica Clínica. Máster en Ciencias. Profesor Auxiliar. Universidad de Ciencias Médicas de Holguín. Holguín, Cuba. <https://orcid.org/0000-0002-9067-9103>

⁶Doctor en Medicina. Especialista en Medicina General Integral y en Reumatología. Máster en Ciencias. Profesor Auxiliar. Investigador Agregado. Hospital Clínico Quirúrgico "Lucía Íñiguez Landín". Holguín, Cuba <https://orcid.org/0000-0002-8420-7492>

Correspondencia: raisarod@infomed.sld.cu

RESUMEN

Fundamento: Las autoevaluaciones son una herramienta pedagógica clave en entornos virtuales, pero es fundamental analizar su calidad y fiabilidad, no solo el rendimiento del alumno, para asegurar su eficacia formativa.

Objetivo: Analizar el rendimiento estudiantil y las características psicométricas de tres autoevaluaciones del curso virtual Metabolismo-Nutrición para estudiantes de ciencias de la salud.

Métodos: Se realizó un estudio cuantitativo con una cohorte de 53 estudiantes, analizando datos estadísticos del LMS. Se midió el rendimiento en los distintos intentos y la consistencia interna de cada cuestionario mediante el coeficiente alfa de Cronbach.

Resultados: Los estudiantes utilizaron los intentos múltiples como estrategia de aprendizaje, mejorando sus calificaciones. La Autoevaluación 3 fue más difícil pero mostró una consistencia interna aceptable ($\alpha=0,758$). En cambio, las dos primeras, aunque más sencillas para el alumnado, presentaron una fiabilidad muy baja ($\alpha\approx0,38$), cuestionando su validez.

Conclusiones: El análisis psicométrico es una herramienta diagnóstica crucial. No solo mide el rendimiento estudiantil, sino que evalúa la calidad y fiabilidad del instrumento de evaluación, permitiendo orientar mejoras pedagógicas y de diseño para lograr una evaluación más justa, válida y eficaz.

Palabras claves: autoevaluación; psicometría; rendimiento académico; analítica del aprendizaje; consistencia interna.

ABSTRACT

Background: Self-assessments are a key pedagogical tool in virtual environments, but it is crucial to analyze their quality and reliability, not just student performance, to ensure their formative effectiveness.

Objective: To analyze the student performance and psychometric characteristics of three self-assessments from the online course "Metabolism-Nutrition" for health science students.

Methods: A quantitative study was conducted with a cohort of 53 students, analyzing statistical data from the LMS. Performance across different attempts was measured, and the internal consistency of each questionnaire was calculated using Cronbach's alpha coefficient.

Results: Students used multiple attempts as a learning strategy, improving their scores. Self-assessment 3 was more difficult but showed acceptable internal consistency ($\alpha=0.758$). In contrast, the first two, although easier for students, showed very low reliability ($\alpha\approx0.38$), questioning their validity.

Conclusions: Psychometric analysis is a crucial diagnostic tool. It not only measures student performance but also evaluates the quality and reliability of the assessment instrument, guiding pedagogical and design improvements to achieve a fairer, more valid, and effective assessment.

Keywords: self-assessment; psychometrics; academic performance; learning analytics; internal consistency.

INTRODUCCIÓN

La transformación digital en la educación superior ha consolidado a las plataformas de gestión del aprendizaje (LMS, por sus siglas en inglés) como ecosistemas centrales para la docencia.¹ Dentro de este paradigma, las herramientas de evaluación en línea, particularmente los cuestionarios de autoevaluación, han ganado una prominencia significativa. Estos instrumentos trascienden su función sumativa para convertirse en potentes herramientas de evaluación formativa.² Su valor reside en la capacidad de proporcionar a los estudiantes retroalimentación inmediata y continua, un pilar para fomentar la autorregulación del aprendizaje.³ Al permitirles identificar brechas de conocimiento antes de las evaluaciones de alto impacto, se promueve un ciclo de estudio, práctica y reflexión.⁴

El curso Metabolismo-Nutrición, caracterizado por una alta densidad conceptual y la necesidad de integrar procesos bioquímicos y fisiológicos complejos, es un campo fértil para la aplicación de estas herramientas. Sin embargo, la mera implementación de tecnología evaluativa no garantiza su efectividad pedagógica. Es imperativo que los educadores no solo observen los resultados de los estudiantes, sino que también sometan los propios instrumentos a un escrutinio riguroso para asegurar su calidad psicométrica.⁵ Un instrumento de evaluación debe ser, ante todo, fiable y válido. La fiabilidad se refiere a la consistencia de la medida, es decir, su capacidad para producir resultados similares en condiciones análogas. La validez, por su parte, asegura que el instrumento mide realmente el constructo que pretende medir.⁶

Un instrumento con baja fiabilidad introduce un error de medición significativo, lo que significa que la calificación de un estudiante puede estar más influenciada por el azar o por fallas en el diseño del test que por su conocimiento real. Esto no solo distorsiona la evaluación del aprendizaje individual, sino que también compromete las conclusiones que se pueden extraer sobre la efectividad de la enseñanza.⁷ Indicadores estadísticos como el coeficiente de consistencia interna (comúnmente medido con el Alfa de Cronbach), la desviación estándar de las puntuaciones, y los índices de asimetría y curtosis ofrecen un diagnóstico profundo sobre la calidad del cuestionario y el comportamiento del grupo.⁸

A pesar de la disponibilidad de estos datos en la mayoría de los LMS, a menudo son subutilizados por los docentes. Por ello, el objetivo de este estudio es realizar un análisis descriptivo y comparativo del rendimiento estudiantil y las propiedades psicométricas de tres cuestionarios de autoevaluación utilizados en el curso Metabolismo-Nutrición. Con ello se busca diagnosticar su efectividad como herramientas de aprendizaje y, fundamentalmente, como instrumentos de medición fiables, para guiar futuras mejoras instruccionales.

METODOLOGÍA

2.1. Diseño del estudio

Se llevó a cabo un estudio cuantitativo, descriptivo y de tipo ex post facto. El análisis se basó en datos secundarios extraídos de los informes estadísticos generados por la plataforma LMS (basada en Moodle) utilizada en la institución.

2.2. Participantes y contexto

La muestra estuvo constituida por los datos de rendimiento de una cohorte completa de 53 estudiantes ($N=53$) matriculados en la asignatura Metabolismo-Nutrición, perteneciente a un programa de ciencias de la salud de una universidad. Se incluyeron los datos de todos los estudiantes que completaron al menos el primer intento en las tres autoevaluaciones obligatorias del curso.

2.3. Instrumentos

Los instrumentos analizados fueron tres cuestionarios en línea denominados Autoevaluación 1, Autoevaluación 2 y Autoevaluación 3. Cada cuestionario consistía en 20 preguntas de opción múltiple con retroalimentación automática. Fueron diseñados para evaluar la comprensión de diferentes unidades temáticas del curso y estaban configurados para permitir múltiples intentos, registrándose la calificación más alta.

2.4. Recolección y análisis de datos

Se recopilaron los informes estadísticos agregados generados automáticamente por el LMS para cada cuestionario. Las variables analizadas incluyeron:

- **Estadísticas de participación:** Número de primeros intentos y número total de intentos.
- **Medidas de tendencia central:** Promedio de los primeros intentos, promedio de todos los intentos, calificación media de los mejores intentos y mediana de los mejores intentos.
- **Medidas de dispersión:** Desviación estándar y error estándar de los mejores intentos.
- **Forma de la distribución:** Asimetría y curtosis de la distribución de las mejores calificaciones, para evaluar la normalidad y posible efecto techo o suelo.
- **Indicadores de fiabilidad:** Coeficiente de consistencia interna (calculado como Alfa de Cronbach por el LMS) y el ratio de error asociado ($1 - \text{Alfa}$), que representa la proporción de la varianza total atribuible al error.

El análisis se centró en la comparación de estos indicadores entre los tres cuestionarios para identificar patrones de rendimiento y diferencias en la calidad psicométrica de los instrumentos.

RESULTADOS

Los datos estadísticos clave para cada una de las tres autoevaluaciones se resumen en la Tabla 1.

Tabla 1. Resumen de estadísticas de rendimiento y psicométricas de las autoevaluaciones

| Métrica | Autoevaluación 1 | Autoevaluación 2 | Autoevaluación 3 |
|----------------------------|------------------|------------------|------------------|
| Nº primeros intentos | 53 | 53 | 53 |
| Nº total de intentos | 114 | 110 | 100 |
| Promedio 1er intento | 79,47 % | 77,50 % | 62,55 % |
| Promedio mejores intentos | 87,80 % | 86,48 % | 75,47 % |
| Mediana (mejores intentos) | 90,00 % | 89,39 % | 76,67 % |
| Desv. Estándar (mejores) | 11,37 % | 11,88 % | 20,86 % |
| Asimetría (mejores) | -0,6241 | -1,4258 | -0,7091 |
| Curtosis (mejores) | -0,5823 | 2,4940 | -0,4001 |
| Consistencia Interna (α) | 38,61 % (0,386) | 38,93 % (0,389) | 75,77 % (0,758) |
| Ratio de Error (1-α) | 78,35 % | 78,15 % | 49,22 % |
| Error Estándar | 8,91 % | 9,29 % | 10,27 % |

Fuente: Informes de la plataforma Moodle.

3.1. Rendimiento y participación

El número de intentos totales por cuestionario (114, 110 y 100) fue consistentemente el doble del número de estudiantes (53), lo que indica un uso generalizado de la opción de múltiples intentos como estrategia de aprendizaje. En las tres evaluaciones, se observa una mejora sustancial entre el promedio del primer intento y el promedio de los mejores intentos. La ganancia de puntuación fue de 8,33 puntos porcentuales para la Autoevaluación 1, 8,98 para la 2, y un notable incremento de 12,92 para la Autoevaluación 3.

3.2. Dificultad comparativa

La Autoevaluación 3 destaca por su mayor nivel de dificultad objetiva. El promedio en el primer intento (62,55%) fue marcadamente inferior al de las Autoevaluaciones 1 (79,47%) y 2 (77,50%). Esta diferencia se mantuvo en las calificaciones de los mejores intentos. Además, la desviación estándar de la Autoevaluación 3 (20,86%) fue casi el doble que la de las otras dos, indicando una mayor dispersión en las puntuaciones y una mayor heterogeneidad en el dominio de la materia por parte del grupo.

3.3. Propiedades psicométricas

El hallazgo más significativo reside en las métricas de fiabilidad. Las Autoevaluaciones 1 y 2 presentaron un coeficiente de consistencia interna (Alfa de Cronbach) extremadamente bajo ($\alpha = 0,386$ y $\alpha = 0,389$, respectivamente). Estos valores se sitúan muy por debajo del umbral comúnmente aceptado de $\alpha \geq 0.70$ para evaluaciones de aula.^{9,10} Esto se corresponde con un ratio de error muy elevado (superior al 78%), lo que sugiere que la mayor parte de la varianza en las puntuaciones se debe a factores aleatorios y no al conocimiento de los estudiantes. En marcado contraste, la Autoevaluación 3, pese a ser más difícil, mostró una consistencia interna aceptable ($\alpha = 0,758$) y un ratio de error considerablemente menor (49,22%).

En cuanto a la distribución, las puntuaciones de los mejores intentos en los tres casos mostraron una asimetría negativa, indicando un sesgo hacia las calificaciones altas (efecto techo). Este efecto fue particularmente pronunciado en la Autoevaluación 2 (asimetría = -1,4258), que también presentó una distribución leptocúrtica (curtosis = 2,4940), sugiriendo una alta concentración de estudiantes con puntuaciones máximas o cercanas al máximo.

DISCUSIÓN

Los resultados de este estudio ofrecen una doble perspectiva: por un lado, el comportamiento de los estudiantes frente a las autoevaluaciones y, por otro, la calidad intrínseca de dichos instrumentos.

4.1. Las autoevaluaciones como herramienta de aprendizaje activo

El incremento sistemático de las calificaciones entre los primeros y los mejores intentos confirma que los estudiantes utilizan estos cuestionarios de forma formativa. La posibilidad de repetir la evaluación, combinada con la retroalimentación inmediata, crea un ciclo de aprendizaje en el que los estudiantes pueden identificar errores, revisar el material de estudio y consolidar su conocimiento. Este comportamiento es un claro indicador de aprendizaje autorregulado³ y metacognición.¹¹ La mayor ganancia de puntuación en la Autoevaluación 3 puede atribuirse a su mayor dificultad inicial, lo que genera una "dificultad deseable" que impulsa un esfuerzo cognitivo más profundo para mejorar el resultado.¹²

4.2. Dificultad del ítem versus calidad del instrumento

La disparidad entre la dificultad percibida y la calidad psicométrica es el aspecto más revelador del análisis. Las Autoevaluaciones 1 y 2, aunque aparentemente "fáciles" para la mayoría de los estudiantes, son instrumentos psicométricamente débiles. Una consistencia interna tan baja ($\alpha < 0.40$) es inaceptable e indica que las preguntas del cuestionario no miden un constructo unificado y coherente.⁷ Las posibles causas incluyen ítems ambiguos, preguntas que miden constructos diferentes, distractores poco plausibles o errores en la clave de corrección.¹³ Un instrumento con

baja fiabilidad es como una balanza que da un peso distinto cada vez que se mide el mismo objeto; sus resultados no son confiables y amenazan la validez de cualquier inferencia sobre el aprendizaje.

Por el contrario, la Autoevaluación 3, si bien representa un mayor desafío académico, es un instrumento de medida mucho más robusto. Su consistencia interna ($\alpha = 0,758$) es aceptable para evaluaciones formativas,⁹ indicando que sus ítems trabajan de manera coherente para medir el dominio del tema. La combinación de una mayor dificultad y una buena fiabilidad sugiere que este cuestionario contiene ítems mejor diseñados, que discriminan adecuadamente entre los estudiantes que han alcanzado los objetivos de aprendizaje y los que no.⁵

4.3. Implicaciones pedagógicas y para el diseño instruccional

Estos hallazgos tienen implicaciones directas para la práctica docente y el diseño instruccional:

1. **Revisión y análisis de ítems:** Las Autoevaluaciones 1 y 2 deben ser revisadas en profundidad. Se recomienda realizar un análisis de ítems detallado, examinando la dificultad y el índice de discriminación de cada pregunta para identificar y corregir o eliminar los ítems problemáticos que degradan la fiabilidad del test.¹⁴
2. **Modelo de calidad y andamiaje:** La Autoevaluación 3 puede ser considerada un modelo a seguir en términos de diseño de preguntas. No obstante, su mayor dificultad y la amplia dispersión de las notas sugieren que los estudiantes podrían beneficiarse de un mayor andamiaje pedagógico (p. ej., más ejemplos, lecturas de apoyo, foros de discusión) para prepararse mejor para el desafío.
3. **Cultura de calidad basada en datos:** Se recomienda a los equipos docentes institucionalizar la práctica de analizar las estadísticas psicométricas de todas las evaluaciones. Esto transforma los datos del LMS de un simple registro de calificaciones a una potente herramienta de analítica del aprendizaje para la mejora continua de la calidad docente y evaluativa.¹⁵

4.4. Limitaciones y futuras líneas de investigación

Este estudio presenta ciertas limitaciones. Primero, la muestra se restringe a una única cohorte de un curso específico, lo que limita la generalización de los resultados. Segundo, el diseño es correlacional y no permite establecer relaciones de causalidad. Tercero, el análisis cuantitativo no captura las percepciones o estrategias cualitativas de los estudiantes.

Futuras investigaciones podrían abordar estas limitaciones mediante estudios longitudinales que rastreen la evolución de la calidad de los instrumentos a lo largo del tiempo. Sería valioso implementar diseños mixtos que combinen el análisis estadístico con entrevistas o grupos focales para comprender por qué los estudiantes utilizan los múltiples intentos y cómo perciben la dificultad y claridad de las preguntas.

CONCLUSIONES

Este estudio demuestra que un análisis estadístico detallado de las autoevaluaciones en línea proporciona información valiosa que trasciende el mero seguimiento del rendimiento estudiantil. Ha revelado una preocupante desconexión entre la aparente facilidad de un test y su calidad psicométrica. Si bien los estudiantes utilizan activamente estas herramientas para mejorar su aprendizaje, no todos los instrumentos son igualmente fiables. La identificación y corrección de cuestionarios con baja consistencia interna es un paso crítico para mejorar la justicia, la validez y la precisión de la evaluación. Para el curso Metabolismo-Nutrición, y por extensión para cualquier disciplina, es fundamental que los educadores se apropien de estas herramientas analíticas para asegurar que los instrumentos evaluativos no solo faciliten el aprendizaje, sino que también lo midan de una manera psicométricamente sólida y defendible.

REFERENCIAS BIBLIOGRÁFICAS

1. Rasheed RA, Kamsin A, Abdullah NA. Challenges in the online component of blended learning: A systematic review. Comput Educ. 2020;144:103701. Disponible en: <https://doi.org/10.1016/j.compedu.2019.103701>
2. Yan Z, Wang X, Boud D, Lao H. The effect of self-assessment on academic performance and the role of explicitness: A meta-analysis. Assess Eval High Educ. 2023;48(3):334-51. Disponible en: <https://doi.org/10.1080/02602938.2022.2091929>
3. Broadbent J, Panadero E, Fuller-Tyszkiewicz M. The impact of online formative assessment on self-regulated learning: A systematic review. Comput Educ. 2021;172:104253. Disponible en: <https://doi.org/10.1016/j.compedu.2021.104253>
4. Rakoczy K, Pinger P, Hochweber J, Klieme E, Schütze B, Bircan G. Formative assessment in mathematics: The effect of task-specific feedback and strategy-focused feedback on students' learning behavior. Z Erziehungswiss. 2019;22(4):859-78. Disponible en: <https://doi.org/10.1007/s11618-019-00898-w>
5. Kane MT. The validation of measurement-based interpretations. Measurement. 2021;19:100147. Disponible en: <https://doi.org/10.1016/j.measurement.2021.100147>
6. An C, Runyon C. Review of an introduction to psychological science, third canadian edition. Psychol Learn Teach. 2022;21(3):344-7. Disponible en: <https://doi.org/10.1177/14757257221105943>
7. Jardina JR, Ogunyemi D. Psychometric analysis of online learning platform quiz data: a comparison of classical test theory and item response theory. Med Sci Educ. 2021;31(2):635-40. Disponible en: <https://doi.org/10.1007/s40670-021-01227-8>

8. Cho E. Neither Cronbach's alpha nor McDonald's omega: A commentary on life and death of coefficient alpha. *Health Serv Outcomes Res Method.* 2021;21(1):71-8. Disponible en: <https://doi.org/10.1007/s10742-020-00227-4>
9. Sendk G, Padilha M, Santos L, Brandão C, Lira C, Pithon K, et al. Psychometric analysis of multiple-choice questions in a formative assessment for medical students. *BMC Med Educ.* 2021;21(1):521. Disponible en: <https://doi.org/10.1186/s12909-021-02947-6>
10. Sijtsma K, van der Ark LA. A general framework for measurement with multiple outcomes. *Psychometrika.* 2020;85(1):217-36. Disponible en: <https://doi.org/10.1007/s11336-020-09699-3>
11. Panadero E, Andres P, Fitjar RD, Pinya C. The role of self-assessment and self-praise in self-regulated learning. *Z Psychol.* 2019;227(1):3-9. Disponible en: <https://doi.org/10.1027/2151-2604/a000350>
12. Bjork RA, Bjork EL. A new theory of disuse and an old theory of stimulus fluctuation. In: Healy AF, Kosslyn SM, Shiffrin RM, editors. From learning processes to cognitive processes: Essays in honor of William K. Estes. Vol. 2. Hillsdale, NJ: Lawrence Erlbaum Associates; 1992. p. 35-67. Disponible en: https://bjorklab.psych.ucla.edu/wp-content/uploads/sites/13/2016/07/RBjork_EBjork_1992.pdf
13. Tarrant M, Ware J. A framework for improving the quality of multiple-choice questions. *Nurse Educ.* 2020;45(5):E79-E83. Disponible en: <https://doi.org/10.1097/NNE.0000000000000762>
14. Sam AH, Field SM, Collares CF, van der Vleuten CPM, Wass VJ, Melville C, et al. Very-short-answer questions: reliability, discrimination and validity. *BMC Med Educ.* 2019;19(1):453. Disponible en: <https://doi.org/10.1186/s12909-019-1886-z>
15. Matcha W, Gašević D, Pardo A. A systematic review of empirical studies on learning analytics dashboards: A self-regulated learning perspective. *IEEE Trans Learn Technol.* 2020;13(2):224-45. Disponible en: <https://doi.org/10.1109/TLT.2019.2916892>

Conflicto de intereses: Los autores declaran no tener ningún conflicto de intereses.