REVISTA MULTIDISCIPLINAR EPISTEMOLOGÍA DE LAS CIENCIAS

Volumen 3, Número 1
Enero-Marzo 2026

Edición Trimestral

CROSSREF PREFIX DOI: 10.71112

ISSN: 3061-7812, www.omniscens.com

Revista Multidisciplinar Epistemología de las Ciencias

Volumen 3, Número 1
enero-marzo 2026

Publicación trimestral
Hecho en México

La Revista Multidisciplinar Epistemología de las Ciencias acepta publicaciones de cualquier área del conocimiento, promoviendo una plataforma inclusiva para la discusión y análisis de los fundamentos epistemológicos en diversas disciplinas. La revista invita a investigadores y profesionales de campos como las ciencias naturales, sociales, humanísticas, tecnológicas y de la salud, entre otros, a contribuir con artículos originales, revisiones, estudios de caso y ensayos teóricos. Con su enfoque multidisciplinario, busca fomentar el diálogo y la reflexión sobre las metodologías, teorías y prácticas que sustentan el avance del conocimiento científico en todas las áreas.

Contacto principal: admin@omniscens.com

9773061781003

Cintillo legal

Revista Multidisciplinar Epistemología de las Ciencias Vol. 3, Núm. 1, enero-marzo 2026, es una publicación trimestral editada por el Dr. Moises Ake Uc, C. 51 #221 x 16B , Las Brisas, Mérida, Yucatán, México, C.P. 97144 , Tel. 9993556027, Web: https://www.omniscens.com, admin@omniscens.com, Editor responsable: Dr. Moises Ake Uc. Reserva de Derechos al Uso Exclusivo No. 04-2024-121717181700-102, ISSN: 3061-7812, ambos otorgados por el Instituto Nacional del Derecho de Autor (INDAUTOR). Responsable de la última actualización de este número, Dr. Moises Ake Uc, fecha de última modificación, 1 enero 2026.

Revista Multidisciplinar Epistemología de las Ciencias

Volumen 3, Número 1, 2026, enero-marzo

DOI: https://doi.org/10.71112/yqay0d64

ANÁLISIS DE CLASES LATENTES COMO ENFOQUE DE MODELAMIENTO NO OBSERVADO: FUNDAMENTOS TEÓRICOS, SUPUESTOS Y RELACIONES CON LOS MODELOS DE MEDICIÓN DE RASCH

LATENT CLASS ANALYSIS AS AN UNOBSERVED MODELING APPROACH: THEORETICAL FOUNDATIONS, MODEL ASSUMPTIONS, AND RELATIONSHIPS WITH RASCH MEASUREMENT MODELS

Jeremias Willmore Metivier

República Dominicana

DOI: https://doi.org/10.71112/yqay0d64

Análisis de Clases Latentes como Enfoque de Modelamiento no Observado: fundamentos teóricos, supuestos y relaciones con los modelos de medición de Rasch

Latent Class Analysis as an Unobserved Modeling Approach: Theoretical Foundations, Model Assumptions, and Relationships with Rasch Measurement Models

Jeremias Willmore Metivier

jeremiaswillmore@gmail.com

https://orcid.org/0000-0002-2759-2830

Universidad Católica del Cibao (UCATECI)

República Dominicana

RESUMEN

El análisis de clases latentes (Latent Class Analysis, LCA) constituye un enfoque estadístico orientado a la identificación de subpoblaciones no observadas a partir de patrones de respuesta en variables categóricas. Se diferencia de los modelos continuos de rasgos latentes, porque el LCA asume que la heterogeneidad poblacional puede representarse mediante un número finito de clases cualitativamente distintas. Este artículo presenta una revisión teórica del análisis de clases latentes, abordando fundamentos conceptuales, formalización estadística, supuestos del modelo y los criterios comúnmente utilizados para la determinación del número óptimo de clases. Se discute su relación conceptual y metodológica con los modelos de medición de Rasch, destacando similitudes, diferencias y posibles usos complementarios en investigación educativa. El escrito concluye con una reflexión crítica sobre

DOI: https://doi.org/10.71112/yqay0d64

las ventajas, limitaciones y desafíos actuales del enfoque de clases latentes en el estudio de fenómenos educativos complejos.

Palabras clave: clases latentes, modelos de mezcla, heterogeneidad poblacional, Rasch, medición educativa.

ABSTRACT

Latent Class Analysis (LCA) is a statistical approach aimed at identifying unobserved subpopulations based on patterns of responses to categorical variables. Unlike continuous latent trait models, LCA assumes that population heterogeneity can be represented through a finite number of qualitatively distinct latent classes. This article presents a theoretical review of latent class analysis, addressing its conceptual foundations, statistical formalization, model assumptions, and the criteria commonly used to determine the optimal number of classes. The conceptual and methodological relationship between LCA and Rasch measurement models is also examined, highlighting their similarities, differences, and potential complementary uses in educational research. The paper concludes with a critical reflection on the advantages, limitations, and current challenges of latent class approaches in the study of complex educational phenomena.

Keywords: latent classes, mixture models, population heterogeneity, Rasch, educational measurement.

Recibido: 10 enero 2026 | Aceptado: 26 enero 2026 | Publicado: 27 enero 2026

DOI: https://doi.org/10.71112/yqay0d64

INTRODUCCIÓN

En la investigación educativa y social, su problemática central se sitúa en la adecuada representación de la heterogeneidad de las poblaciones estudiadas. Históricamente, muchos modelos estadísticos han asumido implícita o explícitamente la homogeneidad de los individuos respecto a los procesos generadores de datos. Esta suposición resulta frecuentemente insostenible cuando se analizan fenómenos complejos como el aprendizaje, el rendimiento académico o los estilos de respuesta en pruebas estandarizadas.

El análisis de clases latentes (Latent Class Analysis, LCA), introducido por Lazarsfeld y Henry en la década de 1960, surge como una alternativa para modelar dicha heterogeneidad mediante la identificación de subgrupos no observables directamente, denominados clases latentes. Desde esta perspectiva, se asume que las asociaciones observadas entre las variables manifiestas pueden explicarse por la pertenencia de los individuos a un número finito de clases cualitativamente distinta. Se diferencia de los modelos de rasgo latente continuo —como los modelos de la Teoría de Respuesta al Ítem y, en particular, el modelo de Rasch—, el análisis de clases latentes conceptualiza la variable latente como categórica. Esta diferencia no es meramente técnica, sino que implica concepciones distintas sobre la naturaleza del constructo subyacente, la forma de la variabilidad individual y los objetivos de la inferencia estadística.

El presente artículo tiene como objetivo desarrollar una revisión teórica del análisis de clases latentes, en función de los fundamentos conceptuales, la formulación estadística y los criterios de decisión asociados a la selección del número de clases. Finalmente, se exploran las relaciones entre el enfoque de clases latentes y los modelos de Rasch, con el propósito de analizar sus posibles usos, en conjunto, en estudios de medición educativa y evaluación del aprendizaje.

DOI: https://doi.org/10.71112/yqay0d64

DESARROLLO

Fundamentos Conceptuales de las Clases Latentes

El análisis de clases latentes (Latent Class Analysis, LCA) pertenece a la familia de los modelos de mezcla finita y constituye un enfoque estadístico orientado a la identificación de subpoblaciones no observables a partir de patrones de respuesta en variables de manifestación categóricas (Lazarsfeld & Henry, 1968; McCutcheon, 1987). Desde una mirada conceptual, el LCA parte del supuesto de que la heterogeneidad observada en una población puede explicarse mediante la existencia de un número finito de clases cualitativamente distintas, cada una de las cuales presenta un perfil de probabilidades de respuesta. Su origen se remonta a los trabajos de Paul Lazarsfeld, quien introdujo el concepto de variables latentes categóricas como una alternativa a los enfoques continuos dominantes en la psicometría clásica (Lazarsfeld, 1950; Lazarsfeld & Henry, 1968). En este marco, las clases latentes no representan niveles cuantitativos de un rasgo subyacente, sino tipologías discretas que agrupan individuos con patrones similares de comportamiento o respuesta.

Desde el punto de vista epistemológico, el LCA se apoya en una concepción tipológica de los constructos latentes, en contraste con la concepción dimensional propia de los modelos de rasgo latente continuo (Collins & Lanza, 2010). Mientras que los modelos continuos asumen que las diferencias individuales se distribuyen a lo largo de un continuo unidimensional, el análisis de clases latentes postula que dichas diferencias reflejan pertenencias cualitativamente diferenciadas a subgrupos latentes. Un concepto central en el LCA, y que es común a ambos enfoques, es el de independencia local, el cual establece que, condicionadas a la clase latente, las variables observadas son estadísticamente independientes entre sí (Goodman, 1974; Hagenaars & McCutcheon, 2002). Este supuesto implica que toda la asociación entre los ítems o indicadores se explica exclusivamente por la variable latente categórica, lo que confiere al modelo una interpretación clara y parsimoniosa.

DOI: https://doi.org/10.71112/yqay0d64

Asimismo, el LCA se fundamenta en la noción de heterogeneidad no observada, entendida como la variabilidad entre individuos que no puede ser explicada por variables observables incluidas explícitamente en el modelo (Vermunt & Magidson, 2002; Oberski, 2016). En este sentido, las clases latentes actúan como variables auxiliares que capturan patrones estructurales subyacentes en los datos. Las clases latentes no se conciben como categorías deterministas, sino como constructos definidos en términos de probabilidades de pertenencia (Goodman, 2002). En tal sentido, cada individuo posee una probabilidad estimada de pertenecer a cada clase, lo que permite una interpretación flexible y coherente con la incertidumbre inherente a la inferencia estadística (Bolck, Croon, & Hagenaars, 2004).

En las ciencias sociales y educativas, el análisis de clases latentes ha sido ampliamente utilizado para identificar perfiles de aprendizaje, estilos cognitivos, trayectorias de desarrollo y patrones de respuesta en pruebas estandarizadas (Nylund-Gibson & Choi, 2018; Masyn, 2013). Su gran potencial radica en la posibilidad de describir poblaciones complejas sin asumir supuestos de continuidad o normalidad sobre los constructos latentes. El LCA se distingue de otros enfoques de clasificación tradicionales, como el análisis de clúster, en que se basa en un modelo estadístico explícito y en criterios de ajuste bien definidos, en lugar de medidas ad hoc de similitud o distancia (Vermunt & Magidson, 2005; Magidson & Vermunt, 2004). Esta característica refuerza su estatus como modelo inferencial más que como técnica exploratoria puramente descriptiva.

Finalmente, diversos autores han señalado que el análisis de clases latentes debe entenderse no como una herramienta de "descubrimiento automático" de grupos reales, sino como un modelo teórico que propone una representación simplificada y útil de la estructura latente de los datos (Bauer & Curran, 2003; Muthén, 2004). En consecuencia, la interpretación de las clases latentes requiere siempre un anclaje sustantivo en el marco teórico y empírico del estudio, evitando lecturas esencialistas o reificadas de las categorías obtenidas.

DOI: https://doi.org/10.71112/yqay0d64

Formalización Estadística del Modelo de Clases latentes

El análisis de clases latentes (*Latent Class Analysis*, LCA) se formaliza como un modelo de mezcla finita en el que la distribución conjunta de un conjunto de variables observadas se explica mediante la pertenencia a una variable latente categórica no observada. Sea una muestra de N individuos, donde para cada individuo i se observan J variables manifiestas categóricas $Y_{ij} (j = 1, \ldots, J)$. Se postula la existencia de una variable latente C_i que representa la clase latente del individuo i, con K posibles categorías mutuamente excluyentes: $C_i \in \{1, \ldots, K\}$.

La probabilidad de pertenencia a cada clase latente se denota por

$$\pi_k = P(C_i = k),$$

con la restricción

$$\sum_{k=1}^{K} \pi_k = 1.$$

Estas probabilidades representan la proporción esperada de individuos en cada clase latente dentro de la población.

Un supuesto central del modelo es el de independencia local, según el cual las variables observadas son estadísticamente independientes entre sí una vez condicionadas a la clase latente. Bajo este supuesto, la probabilidad conjunta de un patrón de respuestas $\mathbf{Y}_i = (Y_{i1}, \ldots, Y_{iJ})$, dado que el individuo pertenece a la clase k, puede expresarse como el producto de probabilidades condicionales:

$$P(\mathbf{Y}_i = \mathbf{y}_i \mid C_i = k) = \prod_{j=1}^{J} P(Y_{ij} = y_{ij} \mid C_i = k).$$

DOI: https://doi.org/10.71112/yqay0d64

Para cada variable manifiesta Y_{ij}, se definen parámetros de respuesta condicional por clase. En el caso de variables dicotómicas, estos parámetros corresponden a probabilidades del tipo $P(Y_{ij} = 1 \mid C_i = k)$. En el caso de variables politómicas, se definen probabilidades específicas para cada categoría de respuesta, sujetas a restricciones de no negatividad y suma a uno dentro de cada clase.

La probabilidad marginal de observar un patrón de respuestas $\mathbf{Y}_i$ se obtiene como una combinación ponderada de las probabilidades condicionales de cada clase, lo que da lugar a la siguiente expresión de mezcla:

$$P(\mathbf{Y}_i = \mathbf{y}_i) = \sum_{k=1}^{K} \pi_k \prod_{j=1}^{J} P(Y_{ij} = y_{ij} \mid C_i = k).$$

Asumiendo independencia entre individuos, la función de verosimilitud del modelo para la muestra completa se define como el producto de las probabilidades marginales individuales:

$$L(\boldsymbol{\pi}, \boldsymbol{\theta}) = \prod_{i=1}^{N} P(\mathbf{Y}_i = \mathbf{y}_i) = \prod_{i=1}^{N} \left[\sum_{k=1}^{K} \pi_k \prod_{j=1}^{J} P\left(Y_{ij} = y_{ij} \mid C_i = k\right) \right].$$

La estimación de los parámetros del modelo se realiza comúnmente mediante máxima verosimilitud, utilizando el procedimiento de Expectation–Maximization (EM). Este algoritmo permite manejar la naturaleza no observada de la variable latente al alternar entre la estimación de las probabilidades posteriores de pertenencia a cada clase (paso E) y la actualización de los parámetros del modelo (paso M).

Un resultado clave del modelo es la obtención de las probabilidades posteriores de clase, que representan la probabilidad de que un individuo pertenezca a cada clase latente dado su patrón de respuestas observado. Estas probabilidades constituyen la base para la

DOI: https://doi.org/10.71112/yqay0d64

asignación probabilística de individuos a clases y para la interpretación sustantiva de los perfiles latentes identificados. Desde un punto de vista inferencial, el modelo de clases latentes no impone una métrica continua sobre el constructo subyacente, sino que lo conceptualiza como una estructura discreta de estados latentes. En consecuencia, la formalización estadística del LCA ofrece un marco flexible para modelar heterogeneidad no observada, especialmente en contextos donde los supuestos de continuidad y normalidad resultan conceptualmente inapropiados o empíricamente insostenibles.

El modelo de clases latentes se sustenta en varios supuestos fundamentales. En primer lugar, se asume la existencia de una variable latente categórica finita $C_i \in \{1, \ldots, K\}$capaz de representar la heterogeneidad no observada mediante subpoblaciones discretas. En segundo lugar, se postula que la distribución conjunta de las respuestas observadas puede expresarse como una mezcla finita, $P(\mathbf{Y}_i = \mathbf{y}_i) = \sum_{k=1}^{K} \pi_k \; P(\mathbf{Y}_i = \mathbf{y}_i \mid C_i = k)$, con $\sum_{k=1}^{K} \pi_k = 1$. Un supuesto central es la independencia local, según el cual, condicionadas a la clase latente, las variables observadas son independientes: $P(\mathbf{Y}_i = \mathbf{y}_i \mid C_i = k) = \prod_{j=1}^{J} P(Y_{ij} = y_{ij} \mid C_i = k)$.

Adicionalmente, se asume homogeneidad condicional dentro de clase, de modo que los individuos en la misma clase comparten parámetros de respuesta condicionales comunes, así como independencia entre individuos o, en su defecto, que la dependencia por estructuras jerárquicas esté adecuadamente modelada. Finalmente, para garantizar estimabilidad se requiere identificabilidad del modelo y estabilidad numérica en la estimación, considerando que la función de verosimilitud puede presentar máximos locales y soluciones en el borde en presencia de separación casi perfecta.

Criterios para la Determinación del Número de Clases Latentes

La determinación del número de clases latentes (K) constituye una de las decisiones más relevantes en el análisis de clases latentes (LCA), debido a que define el grado de heterogeneidad no observada que el modelo representará y condiciona la interpretación sustantiva de los perfiles resultantes. Dado que no existe un único indicador concluyente, la selección de K debe basarse en una estrategia de triangulación que combine evidencia estadística, estabilidad numérica e interpretabilidad teórica.

En primer lugar, se utilizan criterios de información que comparan modelos con diferente número de clases penalizando la complejidad. Entre los más empleados se encuentran el AIC, el BIC y el BIC ajustado por tamaño muestral. Para un modelo con Kclases, log-verosimilitud máxima $\ell(\hat{\theta}_K)$, p_K parámetros y tamaño muestral N, estos criterios se definen como:

$$\mathrm{AIC}(K) = -2\,\ell(\hat{\theta}_K) + 2p_K,$$

$$\mathrm{BIC}(K) = -2\,\ell(\hat{\theta}_K) + p_K \log{(N)},$$

$$\mathrm{SABIC}(K) = -2\,\ell(\hat{\theta}_K) + p_K \log{\left(\frac{N+2}{24}\right)}.$$

En general, se prefiere el modelo que minimiza estos índices, con la consideración de que el BIC tiende a favorecer soluciones más parsimoniosas y suele emplearse como referencia principal en estudios de LCA.

En segundo lugar, pueden emplearse pruebas basadas en verosimilitud para comparar modelos anidados K versus $K-1$. El incremento de ajuste se expresa mediante la diferencia de verosimilitud:

$\Delta G^2 = -2[\ell(\hat{\theta}_{K-1}) - \ell(\hat{\theta}_K)]$.

Sin embargo, debido a que los supuestos regulares de las pruebas χ^2 no se cumplen plenamente en modelos de mezcla, se recomienda el uso de procedimientos específicos como el Bootstrap Likelihood Ratio Test (BLRT) y, cuando estén disponibles, las aproximaciones Lo–Mendell–Rubin (LMR) o Vuong–Lo–Mendell–Rubin (VLMR), seleccionando el mayor K para el cual la mejora respecto a $K-1$ permanezca estadísticamente significativa, siempre que la solución sea estable e interpretable.

En tercer lugar, se evalúa la calidad de clasificación mediante las probabilidades posteriores de pertenencia y medidas agregadas como la entropía. La probabilidad posterior de clase para el individuo i se define como:

$$\gamma_{ik} = P(C_i = k \mid \mathbf{Y}_i = \mathbf{y}_i) = \frac{\pi_k \prod_{j=1}^{J} P(Y_{ij} = y_{ij} \mid C_i = k)}{\sum_{h=1}^{K} \pi_h \prod_{j=1}^{J} P(Y_{ij} = y_{ij} \mid C_i = h)}.$$

A partir de estas probabilidades, puede calcularse una entropía normalizada como indicador global de separación entre clases:

$$\text{Entropy} = 1 - \frac{-\sum_{i=1}^{N} \sum_{k=1}^{K} \gamma_{ik} \log(\gamma_{ik})}{N \log(K)}.$$

Aunque valores altos de entropía sugieren asignaciones más nítidas, este indicador no debe interpretarse como criterio suficiente para elegir K, sino como evidencia complementaria sobre la claridad de la estructura de clases bajo cada solución. Adicionalmente, se recomienda considerar criterios operativos de tamaño mínimo de clase y estabilidad. En particular, deben evitarse soluciones con clases extremadamente pequeñas (por ejemplo, proporciones marginales muy bajas) si no existe una justificación teórica fuerte, y se debe verificar que la solución se replique bajo múltiples valores iniciales para reducir el riesgo de máximos locales en la estimación.

Por último, es esencial evaluar la consistencia del modelo con el supuesto de independencia local. En presencia de dependencia residual entre indicadores (por contenido, formato o método), el incremento de K puede producir clases espurias que capturan correlaciones no modeladas. Por ello, la selección del número de clases debe complementarse con diagnósticos de residuales y con la inspección de perfiles de probabilidades condicionales, priorizando soluciones que sean estadísticamente adecuadas y sustantivamente interpretables.

En síntesis, la determinación del número de clases latentes debe fundamentarse en una decisión integrada que contemple (a) el comportamiento de AIC/BIC/SABIC, (b) evidencia de pruebas de razón de verosimilitud cuando sea pertinente, (c) calidad de clasificación y separación entre clases, (d) estabilidad numérica y tamaños de clase razonables, y (e) coherencia con el marco teórico y el propósito de la investigación.

Interpretación Sustantiva de las Clases Latentes

La interpretación sustantiva de las clases latentes consiste en traducir la solución estadística del LCA (es decir, las estimaciones de π_k y de las probabilidades condicionales $P(Y_{ij} = r \mid C_i = k)$) en perfiles conceptualmente significativos y coherentes con el marco teórico del estudio (Collins & Lanza, 2010; Hagenaars & McCutcheon, 2002). En sentido estricto, las clases latentes no son “grupos observables” sino constructos inferenciales definidos por patrones probabilísticos; por tanto, su interpretación requiere evitar la reificación (Bauer & Curran, 2003; Masyn, 2013).

Un primer paso consiste en describir cada clase mediante su perfil de probabilidades de respuesta por indicador. En el caso dicotómico, la clase k se caracteriza por

$\theta_{jk} = P(Y_{ij} = 1 \mid C_i = k)$; en el caso politómico, por $\rho_{jkr} = P(Y_{ij} = r \mid C_i = k)$. La comparación

entre clases se realiza examinando contrastes sistemáticos: indicadores con alta discriminación entre clases (probabilidades muy diferentes) ofrecen mayor valor interpretativo que aquellos con perfiles similares (Magidson & Vermunt, 2004; Vermunt & Magidson, 2002).

Un segundo paso es asignar etiquetas sustantivas a las clases (p. ej., “procedimental básico”, “razonamiento estructurado”, “estilo impulsivo”), basadas en (a) la teoría del constructo, (b) evidencia curricular o cognitiva, y (c) el patrón empírico de respuestas. Para evitar etiquetas superficiales, es recomendable formular una breve “narrativa de clase” que explique qué sabe/hace el individuo típico de esa clase y qué evidencias lo sustentan (Collins & Lanza, 2010; Nylund-Gibson & Choi, 2018).

Un tercer paso es evaluar la separación entre clases mediante las probabilidades posteriores $\gamma_{ik} = P(C_i = k \mid \mathbf{Y}_i)$ y medidas agregadas como la entropía. Sin embargo, una

buena separación no garantiza validez sustantiva: una solución puede clasificar “bien” y aun así capturar efectos de método o dependencia local (Oberski, 2016). Por ello, la interpretación debe apoyarse también en diagnósticos de independencia local y coherencia teórica.

Finalmente, cuando se incorporan covariables, se recomienda distinguir entre (a) interpretación descriptiva de clases y (b) interpretación explicativa de su asociación con covariables, evitando lecturas causales no justificadas por el diseño (Vermunt, 2010; Asparouhov & Muthén, 2014). En síntesis, la interpretación sustantiva es un proceso de inferencia integradora: el modelo sugiere una tipología latente, y el investigador valida su sentido mediante teoría, evidencia contextual y análisis de robustez.

Relación entre Análisis de Clases Latentes y Modelos de Rasch

El LCA y los modelos de Rasch comparten el rasgo común de ser modelos latentes orientados a explicar patrones de respuesta, pero difieren de forma fundamental en la naturaleza del continuo latente y en el objetivo de la inferencia. En Rasch, la variable latente es típicamente continua (habilidad θ), y la medición busca construir una escala intervalar en logits con propiedades de objetividad específica bajo supuestos de unidimensionalidad e independencia local (Rasch, 1960/1980; Wright & Masters, 1982; Bond & Fox, 2015). En LCA, la variable latente es categórica ($C \in \{1, \ldots, K\}$), y el objetivo central es modelar heterogeneidad no observada como tipologías discretas (Lazarsfeld & Henry, 1968; Collins & Lanza, 2010).

A nivel de supuestos, ambos enfoques comparten una forma de independencia local, aunque con interpretaciones distintas. En Rasch, la independencia local se entiende condicionada a θ; en LCA, condicionada a C. En ambos casos, violaciones de dependencia local pueden inducir sesgos: en Rasch, distorsionar parámetros de ítems/personas y el ajuste; en LCA, inflar el número de clases o producir clases “método” (Linacre, 2002; Oberski, 2016). En términos de complementariedad, existen al menos cuatro formas relevantes de articulación metodológica:

- LCA como diagnóstico de heterogeneidad en medición: antes o después del ajuste Rasch, el LCA puede identificar subpoblaciones con patrones diferenciales (p. ej., estilos de respuesta, uso de estrategias, subgrupos con dependencia local), que luego se investigan con DIF o modelos más ricos (Embretson & Reise, 2000; Bond & Fox, 2015).

DOI: https://doi.org/10.71112/yqay0d64

- Modelos de mezcla Rasch (mixture IRT): integran la idea de clases latentes con medición IRT/Rasch, permitiendo que existan clases con parámetros distintos (por ejemplo, distintas dificultades o distintas estructuras), formalizando heterogeneidad latente sin abandonar el marco IRT (Rost, 1990; von Davier & Yamamoto, 2004). Conceptualmente, estos modelos conectan directamente LCA y Rasch al considerar C como una variable de mezcla y θ como rasgo dentro de clase.

Clases latentes para perfiles de aprendizaje: mientras Rasch proporciona una escala continua de habilidad, el LCA puede traducir patrones de desempeño en perfiles cualitativos útiles para intervención pedagógica (p. ej., clases de error conceptual), especialmente cuando el objetivo es segmentación educativa más que escalamiento (Masyn, 2013).

LTA y Rasch longitudinal: el Análisis de Transición Latente (LTA) modela cambios discretos de clase; Rasch longitudinal modela cambios continuos en logits. Son perspectivas complementarias: “cuánto cambia” (Rasch) versus “de qué perfil a cuál” (LTA) (Collins & Lanza, 2010; Embretson & Reise, 2000).

En resumen, Rasch es primariamente un marco de medición (escala continua) y LCA un marco de clasificación latente (tipologías). Su integración es especialmente potente en evaluación educativa cuando se requiere simultáneamente estimar habilidad y describir perfiles, o cuando se sospecha heterogeneidad que viola supuestos de medición homogénea.

Ventajas, Limitaciones y Debates Actuales

Ventajas. El LCA ofrece (a) una representación parsimoniosa de heterogeneidad no observada mediante un número finito de clases; (b) un marco probabilístico con verosimilitud explícita y criterios formales de comparación de modelos; y (c) interpretaciones orientadas a perfiles, útiles para decisiones educativas (Hagenaars & McCutcheon, 2002; Collins & Lanza, 2010). Frente a métodos de clúster, el LCA aporta estimación inferencial, manejo de incertidumbre de clasificación y extensiones naturales con covariables o longitudinalidad (Vermunt & Magidson, 2005; Masyn, 2013).

Limitaciones. Entre las limitaciones más importantes destacan: (1) sensibilidad a violaciones de independencia local, que puede inducir sobre extracción de clases; (2) dependencia de la selección de K, donde distintos criterios pueden sugerir soluciones diferentes; (3) presencia de máximos locales y problemas de convergencia, que exigen múltiples inicios; (4) clases pequeñas o soluciones degeneradas en muestras insuficientes; y (5) riesgo de reificación interpretativa (Bauer & Curran, 2003; Oberski, 2016; Nylund-Gibson & Choi, 2018).

Debates actuales. Tres debates son especialmente relevantes:

- Discreto versus continuo: si las clases representan tipologías “reales” o aproximaciones discretas a variación continua. Este debate ha impulsado modelos híbridos como factor mixture y mixture IRT (Masyn, 2013; von Davier & Yamamoto, 2004).
- Enumeración de clases y uso de índices: la práctica de seleccionar K por mínimos de BIC/AIC sin considerar teoría ni diagnósticos ha sido criticada; se recomienda triangulación y reporte transparente (Nylund et al., 2007; Collins & Lanza, 2010).
- Clasificación y uso posterior: cuando se usan clases como variables en análisis posteriores, existen sesgos por error de clasificación. Han surgido enfoques “tres pasos” mejorados (p. ej., BCH) para reducir sesgo, especialmente en modelos con covariables o distal outcomes (Bolck et al., 2004; Vermunt, 2010; Asparouhov & Muthén, 2014).

Implicaciones para la investigación educativa

En investigación educativa, el LCA permite avanzar desde promedios globales hacia una comprensión de distribuciones internas y perfiles de desempeño, alineándose con enfoques de equidad y heterogeneidad del aprendizaje. Entre sus implicaciones principales se incluyen:

1) Perfilamiento diagnóstico: identificación de subgrupos con patrones específicos de aciertos/errores, estrategias o concepciones alternativas, útil para retroalimentación e intervención focalizada (Collins & Lanza, 2010; Masyn, 2013).
2) Diseño y evaluación de intervenciones: el LTA permite estudiar cambios de perfil a lo largo del tiempo (p. ej., antes/después de un programa), complementando métricas continuas de progreso. Esto facilita responder quién mejora, quién se estanca y en qué transiciones se concentran los cambios.
3) Equidad y subpoblaciones invisibles: el LCA puede revelar subgrupos no capturados por variables sociodemográficas observables, ayudando a diseñar estrategias de apoyo más precisas. No obstante, el uso con covariables exige cautela interpretativa (Vermunt, 2010).
4) Medición y validez: combinado con Rasch/IRT, el LCA puede detectar heterogeneidad que afecta supuestos de medición (p. ej., estilos de respuesta, dependencia local, multidimensionalidad latente), fortaleciendo argumentos de validez al documentar estructuras internas alternativas (Bond & Fox, 2015; Embretson & Reise, 2000).
5) Política educativa y toma de decisiones: en contextos de rendición de cuentas o diagnóstico nacional, la segmentación por perfiles puede evitar lecturas simplistas del promedio, facilitando decisiones más informadas sobre focalización curricular, acompañamiento docente y recursos.

En síntesis, el LCA aporta un lenguaje metodológico para pasar de “cuánto” a “cómo” y “quién”, fortaleciendo la utilidad interpretativa de los datos educativos siempre que se sostenga con teoría y buenas prácticas de modelamiento.

CONCLUSIONES

El análisis de clases latentes constituye un enfoque robusto para modelar heterogeneidad no observada mediante tipologías latentes definidas probabilísticamente. su formalización como mezcla finita y su supuesto de independencia local proporcionan un marco claro para interpretar patrones de respuesta en variables categóricas, con extensiones naturales a modelos con covariables y longitudinalidad. no obstante, la validez sustantiva de las clases depende críticamente de la selección de K, de la evaluación de independencia local, de la estabilidad de la estimación y de una interpretación anclada en teoría.

En el campo de la medición educativa, la relación entre lca y rasch es fundamentalmente complementaria: rasch ofrece escalamiento continuo y propiedades de medición; lca ofrece perfilamiento discreto y segmentación interpretativa. la integración de ambos enfoques —incluyendo modelos de mezcla y diagnósticos de heterogeneidad— abre una vía fértil para fortalecer inferencias, apoyar decisiones pedagógicas y construir argumentos de validez más completos. en consecuencia, el uso del lca en educación debe orientarse por una práctica metodológica rigurosa: triangulación de criterios, reporte transparente y coherencia teórica, evitando tanto la sobre dependencia en índices numéricos como la reificación de clases.

Declaración de conflicto de interés

El autor declara no tener ningún conflicto de interés relacionado con esta investigación.

Declaración de contribución a la autoría

Esta sección se debe llenar con los roles de cada uno de los autores, hay 14 roles según la taxonomía de CRediT https://credit.niso.org/No es obligatorio cubrir los 14 roles, solo poner los roles que se involucraron en la investigación.

Nombre y apellidos de primer autor: conceptualización, curación de datos, análisis formal, adquisición de fondos, investigación, metodología, administración del proyecto, recursos, software, supervisión, validación, visualización, redacción del borrador original, revisión y edición de la redacción. Hacer lo mismo para cada uno de los autores.

Declaración de uso de inteligencia artificial

Los autores no utilizaron inteligencia artificial en ninguna parte del manuscrito.

REFERENCIAS

Asparouhov, T., & Muthén, B. (2014). Auxiliary variables in mixture modeling: Three-step approaches using Mplus. Structural Equation Modeling: A Multidisciplinary Journal, 21(3), 329–341.

Bauer, D. J., & Curran, P. J. (2003). Distributional assumptions of growth mixture models: Implications for overextraction of latent trajectory classes. Psychological Methods, 8(3), 338–363. https://doi.org/10.1037/1082-989X.8.3.338

Bolck, A., Croon, M., & Hagenaars, J. (2004). Estimating latent structure models with categorical variables: One-step versus three-step estimators. Political Analysis, 12(1), 3–27. https://doi.org/10.1093/pan/mph001

Bond, T. G., & Fox, C. M. (2015). Applying the Rasch model: Fundamental measurement in the human sciences (3rd ed.). Routledge.

Collins, L. M., & Lanza, S. T. (2010). Latent class and latent transition analysis: With applications in the social, behavioral, and health sciences. Wiley.

Embretson, S. E., & Reise, S. P. (2000). Item response theory for psychologists. Lawrence Erlbaum.

Goodman, L. A. (1974). Exploratory latent structure analysis using both identifiable and unidentifiable models. Biometrika, 61(2), 215–231. https://doi.org/10.1093/biomet/61.2.215

Goodman, L. A. (2002). Latent class analysis: The empirical study of latent types, latent variables, and latent structures. In J. A. Hagenaars & A. L. McCutcheon (Eds.), Applied latent class analysis (pp. 3–55). Cambridge University Press.

Hagenaars, J. A., & McCutcheon, A. L. (Eds.). (2002). Applied latent class analysis. Cambridge University Press. https://doi.org/10.1017/CBO9780511499531

La Lazarsfeld, P. F. (1950). The logical and mathematical foundation of latent structure analysis. In S. A. Stouffer et al. (Eds.), Measurement and prediction (pp. 362–412). Princeton University Press.

zarsfeld, P. F., & Henry, N. W. (1968). Latent structure analysis. Houghton Mifflin.

Linacre, J. M. (2002). What do infit and outfit, mean-square and standardized mean? Rasch Measurement Transactions, 16(2), 878.

Magidson, J., & Vermunt, J. K. (2004). Latent class models. In D. Kaplan (Ed.), The Sage handbook of quantitative methodology for the social sciences (pp. 175–198). Sage.

Masyn, K. E. (2013). Latent class analysis and finite mixture modeling. In T. D. Little (Ed.), The Oxford handbook of quantitative methods (Vol. 2, pp. 551–611). Oxford University Press.

McCutcheon, A. L. (1987). Latent class analysis. Sage. https://doi.org/10.4135/9781412984713

Muthén, B. (2004). Latent variable analysis: Growth mixture modeling and related techniques for longitudinal data. In D. Kaplan (Ed.), The Sage handbook of quantitative methodology for the social sciences (pp. 345–368). Sage.

Nylund, K. L., Asparouhov, T., & Muthén, B. O. (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. Structural Equation Modeling, 14(4), 535–569. https://doi.org/10.1080/10705510701575396

Nylund-Gibson, K., & Choi, A. Y. (2018). Ten frequently asked questions about latent class analysis. Translational Issues in Psychological Science, 4(4), 440–461. https://doi.org/10.1037/tps0000176

Oberski, D. L. (2016). Beyond the number of classes: Separating substantive from non-substantive dependence in latent class analysis. Advances in Data Analysis and Classification, 10(2), 171–182. https://doi.org/10.1007/s11634-015-0219-0

Rasch, G. (1980). Probabilistic models for some intelligence and attainment tests (Expanded ed.). University of Chicago Press. (Original work published 1960)

Rost, J. (1990). Rasch models in latent classes: An integration of two approaches to item analysis. Applied Psychological Measurement, 14(3), 271–282.

Vermunt, J. K. (2010). Latent class modeling with covariates: Two improved three-step approaches. Political Analysis, 18(4), 450–469.

Vermunt, J. K., & Magidson, J. (2002). Latent class cluster analysis. In J. A. Hagenaars & A. L. McCutcheon (Eds.), Applied latent class analysis (pp. 89–106). Cambridge University Press.

Vermunt, J. K., & Magidson, J. (2005). Latent GOLD 4.0 user's guide. Statistical Innovations.

Von Davier, M., & Yamamoto, K. (2004). Mixture-distribution and hybrid Rasch models. In M. von Davier & C. H. Carstensen (Eds.), Multivariate and mixture distribution Rasch models (pp. 99–115). Springer.

Wright, B. D., & Masters, G. N. (1982). Rating scale analysis: Rasch measurement. MESA Press.