

CARACTERIZACIÓN DEL RENDIMIENTO ACADÉMICO DE ESCOLARES DE CUARTO DE EDUCACIÓN PRIMARIA, ESPAÑA

ELENA VERNAZZA ^a, ALAR URRUTICOECHEA ^b, DIANA DEL CALLEJO CANAL ^c,
RAMÓN ÁLVAREZ VAZ ^a, MARGARITA CANAL MARTÍNEZ ^c

^aInstituto de Estadística, Facultad de Ciencias Económicas y de Administración
Universidad de la República (UdelaR), Eduardo Acevedo 1139, C.P. 11200, Montevideo Uruguay
e-mail: evernazza@iesta.edu.uy & ramon@iesta.edu.uy

^bDepartamento de Neurocognición, Facultad de Ingeniería y Tecnologías y Facultad de Ciencias de la Salud
Universidad Católica del Uruguay, Comandante Braga 2715, C.P. 1600, Montevideo, Uruguay
e-mail: alar.urruticoechea@ucu.edu.uy

^cInstituto de Investigación de Estudios Superiores, Económicos y Sociales
Universidad Veracruzana, Dr. Castelazo Ayala S/N, Colonia Industrial Ánimas, C.P. 91190, Xalapa, Veracruz, México
e-mail: ddelcallejo@uv.mx & mcanal@uv.mx

En España, el Ministerio de Educación aplica a nivel nacional una prueba conocida como Evaluación General de Diagnóstico (EDG). Dicha prueba mide el rendimiento académico de escolares en las competencias básicas de comunicación lingüística, matemática, conocimiento e interacción con el mundo físico, y competencia social-ciudadana. Para fines de esta investigación se utilizó la información, recabada en el año 2009, de 25295 escolares de cuarto año de primaria, con el propósito de proponer una medida fiable del rendimiento académico general, que recoja los puntajes de las cuatro competencias mencionadas y permita clasificar a los escolares en función de dicha medida. Para ello, se aplicó un análisis de clases latentes que fue verificado por un análisis discriminante cuadrático con un 90 % de casos coincidentes y un análisis CHAID con un 92.7 % de casos coincidentes. Los resultados muestran que una clasificación con tres clases latentes es suficiente para explicar el rendimiento académico general. Así, 32 % de los escolares de la muestra fueron clasificados con un rendimiento académico alto, caracterizado por presentar valores mayoritariamente altos en las 4 competencias, aunque se detectan también escolares con valores medios en comunicación lingüística y matemática; un 51 % con un rendimiento académico medio, caracterizado por presentar un nivel medio de todas las cuatro competencias; y un 17 % con un rendimiento académico bajo, caracterizado por presentar valores bajos en tres competencias y registros medios en matemáticas. Por último, se detecta que la mayor proporción de escolares en escuelas privadas reportan un nivel de rendimiento alto, mientras que la mayor proporción de escolares en escuelas públicas reportan un rendimiento académico medio. Además, la mayor proporción de escolares no repetidores cuenta con un rendimiento académico alto, mientras que la mayor proporción de escolares repetidores reportan un rendimiento académico bajo.

Keywords: Educación, análisis de clases latentes, CHAID, análisis discriminante, rendimiento.

1. Introducción

Se entiende por rendimiento académico el nivel de conocimientos, habilidades y destrezas que el alumno adquiere a lo largo del ciclo educativo (10). El grado de adquisición de éstos suele ser cuantificado en pruebas que miden el logro académico. Por ejemplo, a nivel internacional se mide mediante las pruebas PISA, a nivel nacional (España), mediante la Evaluación General de

Diagnóstico (EGD) (7) y a nivel de centro mediante las calificaciones globales en exámenes escolares (13). Cada una de estas pruebas mide el rendimiento entendiendo diferentes aspectos del conocimiento. Concretamente en esta investigación se utiliza la EGD de 2009, la cual mide el rendimiento de los alumnos de cuarto de Educación Primaria en las competencias básicas de comunicación lingüística, matemática, conocimiento e interacción con el

mundo físico y competencia social-ciudadana. Cada una de estas habilidades se evalúa mediante pruebas divididas en cinco niveles de dificultad y las puntuaciones de cada una de las habilidades se pueden categorizar midiendo el grado de conocimiento que posee el escolar. Como el conocimiento es gradual, aquel escolar que obtenga puntuaciones correspondientes a una determinada categoría, tendrá adquiridas las capacidades y destrezas de la categoría inferior.

1.1. Objetivos. El objetivo principal de esta investigación es proponer una medida fiable del rendimiento académico general, que recoja los puntajes de las cuatro competencias comunicación lingüística, matemática, conocimiento e interacción con el mundo físico y social-ciudadana, para poder clasificar a los escolares de cuarto año de Educación Primaria de España.

De este objetivo general se desprenden los siguientes objetivos específicos:

- Determinar cuantas categorías son necesarias para clasificar a los escolares en función de su rendimiento académico general (propuesta de un modelo).
- Corroborar la viabilidad del modelo propuesto, mediante estrategias de análisis alternativas, para asegurar su fiabilidad.
- Caracterizar cada uno de los grupos que se forman.

2. Material y método

2.1. Material - Instrumento de Medida. La EGD, llevada a cabo por el Ministerio de Educación de España, tiene por objetivo contribuir a la mejora de la calidad de la educación a partir del conocimiento de la situación del sistema educativo, en este caso, a través de la valoración de los aprendizajes de los estudiantes y de los procesos de innovación y mejora de la educación en todo el sistema. La medición del aprendizaje se da mediante exámenes ordinarios para un nivel académico concreto, particularmente, en 2009 se evaluó a cuarto año de Educación Primaria. La prueba se realiza a nivel nacional y se puede contestar en cualquiera de los idiomas oficiales del Estado (castellano, catalán, gallego, valenciano o vasco). La prueba recoge información del escolar, del centro, de los profesores y de los padres.

2.2. Material - Muestra. En este trabajo se analizaron 25295 escolares y las siguientes variables: titularidad del centro al que acude, idioma, sexo y las puntuaciones en las cuatro competencias ya mencionadas (tanto en su versión categórica como en su versión continua).

Para la categorización de las puntuaciones, se calculó la media de las puntuaciones obtenidas para cada una de las competencias y se asignó un nivel de competencia: bajo, medio o alto.

2.3. Método - Clases latentes. El análisis de clases latentes (ACL) (1), (5) es una técnica cuyo objetivo es clasificar, en grupos, a un conjunto de individuos a partir de la determinación de una variable categórica no observable. El algoritmo utiliza como variables de entrada un conjunto de variables categóricas observables (variables manifiestas).

Así, el ACL busca segmentar la tabla/hipercubo de contingencia, creado a partir de las variables observadas/manifiestas, por una variable no observada/latente, asumiendo que las respuestas a todas las variables manifiestas son estadísticamente independientes con respecto a los valores de la variable de clases latentes (independencia local o condicional). El modelo asocia, en términos de probabilidad, cada individuo a una clase latente y permite determinar el valor esperado con el que cada individuo responde a cada variable observada (2).

La estimación de un modelo a través de ACL no establece una cantidad de clases latentes, sin embargo, una de las ventajas de esta técnica, sobre las técnicas de clusterización más comúnmente utilizadas, es la variedad de herramientas existentes para determinar la cuantía de las clases.

Para el caso en estudio, con la finalidad de elegir el *mejor* modelo de clases latentes, se estimaron y compararon modelos de una a tres clases, bajo los siguientes criterios de selección: los estadísticos G², Chi², AIC y BIC.

Los estadísticos G², Chi² y sus p-valores asociados a la prueba de hipótesis que testea:

H_0) El modelo ajustado es el *adecuado*

vs.

H_1) El modelo ajustado NO es el *adecuado*

indican en todos los escenarios planteados que el modelo ajustado es el adecuado.

Por otra parte, teniendo en consideración que: a) el AIC tiende a sobrestimar el número de clases necesarias; b) el BIC tiende a subestimarlos; y c) en el caso de muestras grandes, se puede presentar un sobreajuste, el estadístico que se utilizó fue el BIC, además del nivel de interpretabilidad general de las clases propuestas. Así,

las clases consideradas para caracterizar el rendimiento académico de los escolares de cuarto año de España, fueron 3.

Sin embargo, dada la posible subestimación del modelo, se ocuparon como estrategias de verificación/respaldo de esta decisión, dos alternativas metodológicas: análisis discriminante y análisis CHAID.

2.3.1. Método - Análisis discriminante. El análisis discriminante lineal o linear discriminant analysis (LDA) es un método de clasificación, en el que la categorización en K grupos es previamente conocida, el objetivo principal es catalogar nuevas observaciones, en los grupos ya establecidos (en este caso, grupos provenientes del modelo de clases latentes que queremos verificar).

Existen dos maneras de abordar una aproximación de LDA: la primera, su forma convencional, hace una reducción de dimensionalidad mediante descomposición en valores singulares (4); la segunda, modela la distribución de las variables predictoras X en cada una de las clases de la variable respuesta y , mediante teorema de Bayes, estima $P(Y = k|X = x)$. Cuando se asume una distribución normal, el modelo es similar a la regresión logística, pero con algunas ventajas (8).

En el caso particular de este estudio se utilizó la alternativa conocida como análisis discriminante cuadrático o quadratic discriminant analysis (QDA), que asume un límite de decisión (o de corte para distinguir a los grupos) que, en lugar de ser una línea, es una curva. La elección de QDA tiene que ver con el hecho de no hacer ningún supuesto acerca de la forma que tendrá el límite de decisión (8).

2.3.2. Método - CHAID. Los métodos AID (Automatic Interaction Detection) (11) tienen por objetivo detectar la presencia de interacción en un modelo de predicción, generalmente mediante algoritmos de tipo iterativo. Existen distintos tipos de métodos AID, según la naturaleza de la variable dependiente, dentro de los que se destaca, por su popularidad, el CHAID (9). Este se utiliza, fundamentalmente con fines exploratorios y descriptivos y su principal objetivo es encontrar una clasificación de la población en estudio, en grupos capaces de describir, de la mejor manera, la variable dependiente. El algoritmo CHAID está pensado para variables (dependiente y predictoras) cualitativas y utiliza el contraste chi-cuadrado de independencia en cada una de las fases del proceso iterativo.

Las etapas de implementación de esta metodología

son:

1. **Agrupación de categorías:** se agruparán las categorías de los predictores cuando éstas tengan un perfil similar en la variable dependiente; es decir, las categorías que al aplicar el test chi-cuadrado resulten no significativas.
2. **Selección del mejor predictor:** si existen varios predictores cuyas categorías son significativamente diferentes respecto a la variable dependiente, se selecciona como mejor predictor aquel para el que se obtenga un menor p -valor (o mayor coeficiente de asociación). Así, el mejor predictor será el que mejor discrimina a los sujetos según la variable dependiente.
3. **Segmentación:** si el mejor predictor es significativo, se realiza la segmentación de la población en tantos segmentos como categorías tenga dicho predictor.
4. **Repetición:** para cada segmento se repite el proceso desde el 1er paso. El proceso sigue hasta que no hay predictores significativos; al nodo que no tiene predictores significativos se le conoce con el nombre de *nodo terminal*. La segmentación terminará, por lo tanto, cuando todos los nodos sean terminales.

3. Resultados

Los resultados presentados se dividen en 3 partes. En primer lugar se presentan las variables utilizadas y sus principales características. A continuación, se exponen los principales resultados que dan lugar a la determinación de la cantidad de clases: evaluación de los modelos planteados con ACL, QDA y CHAID. Por último, se presenta la caracterización final del rendimiento de los escolares de cuarto año de España, en función de las variables que lo determinan, así como su asociación con variables sociodemográficas. Los análisis fueron realizados con el software libre R (12), en particular se utilizaron las librerías *MASS* (14) y *poLCA* (3), y SPSS en su versión 25 (6).

3.1. Variables sociodemográficas y de rendimiento.

De la muestra analizada (25295 escolares), el 63.9 % de los estudiantes provienen de una escuela pública y el 36.1 % de escuela privada, el 49.1 % son niñas y el 50.9 % niños. Se registraron pruebas respondidas en cinco lenguas, con la siguiente distribución de frecuencias: 77.5 % de castellano, 11.1 % de catalán, 5.1 % de gallego, 2.8 % vasco y 1.4 % de valenciano. También se registra que el 91.2 % de los participantes no han repetido ningún curso, 3.5 % repitieron en cuarto de primaria, 3.1 % repitieron segundo de primaria y el resto (2.2 %) no contestaron esa pregunta.

Por otra parte, en cuanto a las variables asociadas al rendimiento académico, cabe destacar que la media de las cuatro dimensiones prácticamente coincide en 507 ($sd \simeq 91$).

Tabla 1. Descriptivos variables de rendimiento académico.

	Min.	Q1.	Mediana	Media (sd)	Q3.	Max.
LING	162.7	442.9	506.5	507.4 (91.8)	572.3	822.9
MAT	191.5	442.3	498.5	507.1 (90.5)	567.7	838.3
MFIS	154.6	445.1	509.2	507.2 (90.8)	573.2	815.2
SYC	172.0	446.6	514.1	507.0 (91.4)	573.7	763.5

3.2. Determinación del número de clases latentes.

Tal como fuera especificado en la sección 2.3, el criterio utilizados para determinar el número de clases de la variable no observable (rendimiento académico) fue el de mínimo BIC. Este índice reporta su mínimo valor al estimar un modelo con tres clases latentes. Así, luego de especificado y estimado un modelo con tres clases latentes se procedió a verificar dicha categorización, mediante las estrategias ya mencionadas: análisis discriminante y análisis CHAID.

En cuanto al análisis discriminante, debido a la estructura de los datos y dado que no se cumple el supuesto de normalidad (Lillifors, $p < 0.001$), ni homogeneidad de varianzas (Levene, $p < 0.001$) se decidió utilizar la alternativa no paramétrica del QDA. Se utilizaron las 4 variables continuas originales (puntuaciones en la competencias de comunicación lingüística, matemática, conocimiento e interacción con el mundo físico y social-ciudadana) como variables independientes y la variable latente (con 3 clases) como variable de agrupación.

Los resultados obtenidos muestran que el 90 % de los casos coinciden en ambas clasificaciones (ACL vs QDA). En la Tabla 2 se observa que el grupo más eficientemente clasificado es el grupo número 2 (G2, asociado a la clase latente 2) con un 91.3 % de coincidencias. Sin embargo, es también el grupo que presenta mayor casos de observaciones mal clasificadas. En el otro extremo, el grupo con menor eficiencia en su clasificación es el grupo 3 (G3, asociado a la clase latente 3) con 84.7 % de coincidencia con 15.3 % de casos confusos con el G2. El grupo 1 (G1, asociado a la clase latente 1) presenta un 90.7 % de coincidencia en la clasificación, con el 9.3 % de los casos confusos con la clase latente 2.

A través de un análisis gráfico (ver Figura 1), se puede observar que los casos mal clasificados (en rojo) se encuentran en los límites de las particiones establecidas por el QDA.

En lo que refiere al análisis CHAID, este se realizó tomando como variables explicativas las mismas variables

Tabla 2. Matriz de confusión QDA y clases latentes.

Clasificación QDA	Clase			Total
	c = 1	c = 2	c = 3	
G1	7300	665	0	7965
G2	745	11747	671	13193
G3	0	459	3706	4092
Total	8045	12871	4377	25293

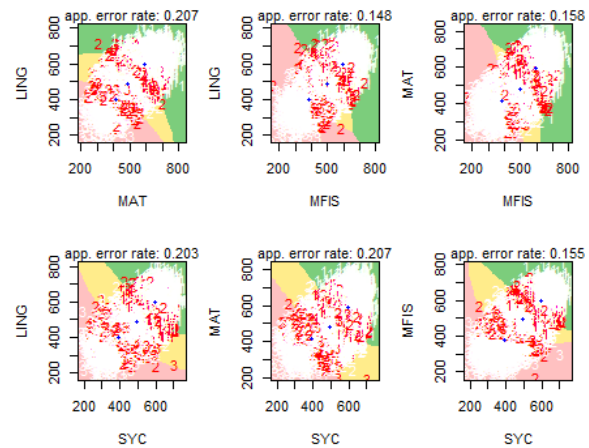


Fig. 1. Partición y clasificación de casos correctos e incorrectos QDA.

que para el ACL (es decir, la versión categorizada de las puntuaciones de cada una de las cuatro competencias). Al igual que para el QDA, la variable dependiente fue la variable de clases latentes (con 3 clases).

Los resultados obtenidos ponen de manifiesto que el 92.7 % de los casos coinciden en ambas clasificaciones (ACL vs CHAID). A partir de los resultados presentados en en la Tabla 3, es posible concluir que el grupo más eficientemente clasificado es el grupo 1 (G1, asociado a la clase latente 1) con un 98.8 % de coincidencias. En el otro extremo, el grupo con menor eficiencia en su clasificación es el grupo 2 (G2, asociado a la clase latente 2) con 88.5 %. La mayor confusión en la clasificación de este grupo se da con la clase 1 (8.1 %). El grupo 3 (G3, asociado a la clase latente 3) presenta un 97.1 % de coincidencia en la clasificación.

Tabla 3. Matriz de confusión CHAID y clases latentes.

Clasificación CHAID	Clase			Total
	c = 1	c = 2	c = 3	
G1	6882	82	0	6964
G2	1161	12674	489	14324
G3	2	115	3888	4005
Total	8045	12871	4377	25293

Así, los resultados obtenidos, respaldan la

determinación de establecer tres clases latentes.

3.3. Caracterización de las clases. En función de lo expuesto previamente se decide estimar un modelo de una variable con tres clases latentes, cuya caracterización se presenta a continuación.

Tal como se observa en la Tabla 4, la probabilidad de pertenecer a la clase 2 es la mayor, mientras que la menor corresponde a la clase 3

Tabla 4. Probabilidad de cada una de las clases.

Clase	1	2	3
P(c)	0.3	0.5	0.2

La caracterización de cada una de las clases se realiza en función de la probabilidad condicional, de cada una de las categorías de cada variable manifiesta, dada la clase. Tomando como referencia los resultados presentados en la Tabla 8, la caracterización de las clases en las que se agrupan a los 25295 niños y niñas en estudio es la siguiente:

Clase Latente 1

Los escolares de esta clase se caracterizan por presentar valores mayoritariamente altos en las 4 variables observadas. Se detectan, también en esta clase, escolares con valores medios en lingüística y matemática. Cabe destacar que la probabilidad de que un escolar que pertenece a esta clase, tenga niveles bajos en cualquiera de las variables observadas, es prácticamente 0. En función de la descripción hecha, se entiende que los patrones característicos de esta clase son los que se presentan en la Tabla 5.

Tabla 5. Patrones - Clase 1.

Variable observada	Categoría(s)
CatLing_R	3/2
CatMat_R	3/2
CatMfis_R	3
CatSys_R	3

Clase Latente 2

Los escolares que se encuentran en la clase latente 2, presentan un nivel medio de todas las variables observadas. Cabe destacar, además, el hecho de que la probabilidad de que un niño o niña que pertenece a esta clase, tenga tanto niveles bajos como altos en alguna de las variables observadas es menor a 0.1 y 0.2 respectivamente. El patrón específico de esta clase es se presenta en la Tabla 6.

Clase Latente 3

Por último, los escolares de esta clase presentan valores

Tabla 6. Patrones - Clase 2.

Variable observada	Categoría(s)
CatLing_R	2
CatMat_R	2
CatMfis_R	2
CatSys_R	2

bajos en las competencias de lingüística, mundo físico y social-ciudadana. Además se caracterizan por tener registros medios en matemática. La caracterización de estos escolares se complementa destacando que la probabilidad de que un escolar que pertenece a esta clase, tenga niveles altos en cualquiera de las variables observadas, es prácticamente 0. Así, los patrones característicos de esta clase son los que se presentan en la Tabla 7.

Tabla 7. Patrones - Clase 3.

Variable observada	Categoría(s)
CatLing_R	1/2
CatMat_R	2/1
CatMfis_R	1/2
CatSys_R	1/2

Tabla 8. Probabilidades condicionales $P(Y_i/c)$.

<i>CatLing_R</i>	1	2	3
c = 1	0.0002	0.3338	0.6661
c = 2	0.0557	0.8658	0.0785
c = 3	0.5715	0.4281	0.0004
<i>CatMat_R</i>	1	2	3
c = 1	0.0034	0.3525	0.6442
c = 2	0.0940	0.7958	0.1103
c = 3	0.4124	0.5840	0.0036
<i>CatMfis_R</i>	1	2	3
c = 1	0.0003	0.2362	0.7635
c = 2	0.0363	0.8408	0.1229
c = 3	0.5876	0.4123	0.0001
<i>CatSys_R</i>	1	2	3
c = 1	0.0000	0.2169	0.7831
c = 2	0.0487	0.7725	0.1788
c = 3	0.5651	0.4286	0.0063

En cuanto a la composición de cada una de las clases, en función de las probabilidades a posteriori de cada una de ellas, el total de escolares de cuarto año queda distribuido en un 32 %, 51 %, y 17 % respectivamente.

Para finalizar, en lo que refiere al *rendimiento académico*, las clases se podrían categorizar como:

- **c = 1:** Niños y niñas cuyo *rendimiento académico en 4to año es alto*.

- **c = 2:** Niños y niñas cuyo *rendimiento académico en 4to año es medio*.

- **c = 3:** Niños y niñas cuyo *rendimiento académico en 4to año es bajo*.

Esta estructura se verifica tanto al analizar el comportamiento por clase, de cada una de las variables originales, utilizadas para el QDA (ver Tabla 10) como al analizar el árbol resultante del CHAID (ver Figura 2).

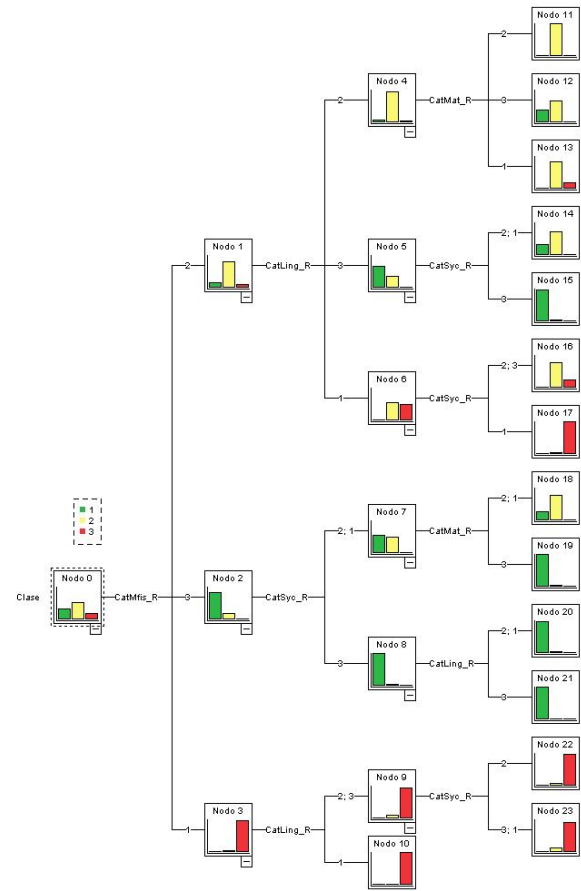


Fig. 2. Resultado CHAID.

Por último, se presenta una caracterización de los escolares por clase latente, en función de su sexo, lengua en la que realizaron la prueba, titularidad de la escuela (Pub./Priv.) y repetición escolar. En cuanto al sexo de los escolares se destaca que la mayor diferencia se da en la clase 3 (46 % niñas y 54 % niños). Al analizar la composición de las clases teniendo en consideración la lengua en la que fue realizada la prueba, se observa que la clase 1 es la que presenta la mayor proporción de pruebas en castellano y la menor proporción de pruebas realizadas en vasco y valenciano. Por otra parte, la clase 2 es la que presenta la mayor proporción de pruebas realizadas en gallego, mientras que en el caso de la clase 3, se registran las mayores proporciones de pruebas en realizadas en catalán, valenciano y vasco.

En lo que refiere a la titularidad de la escuela, se observa que la mayor paridad entre escuelas públicas y privadas se da en la clase 1, mientras que la mayor diferencia de composición (menor proporción de escuelas privadas) se da en la clase 3.

Para finalizar, en lo que refiere a la repetición escolar,

Tabla 9. Media (sd) de las variables originales por clase.

Clase	c = 1	c = 2	c = 3
LING	595.6 (63.2)	489.4 (59.4)	398.3 (61.8)
MAT	592.3 (73.3)	484.2 (61.9)	417.5 (58.8)
MFIS	595.1 (56.3)	495.7 (52.1)	379.8 (54.9)
SYC	591.6 (52.1)	493.1 (62.5)	392.3 (68.7)

Tabla 10. Distribución por clase, según Lengua de la prueba.

Clase/Lengua	Cas.	Cat.	Gal.	Val.	Vas.
c = 1	83.5	8.0	4.8	0.9	2.8
c = 2	75.8	11.5	5.7	1.6	5.4
c = 3	71.6	15.5	4.1	1.9	6.8

Tabla 11. Distribución por clase, según Titularidad.

Clase/Titularidad	Púb.	Priv.
c = 1	0.57	0.43
c = 2	0.65	0.35
c = 3	0.75	0.25

y su distribución por clase latente, se destaca que la clase 1 es la que reporta mayor cantidad de escolares que no han repetido, mientras que la clase 3 es la que registra mayor proporción de escolares que si lo han hecho.

Tabla 12. Distribución por clase, según Repetición.

Clase/Rep.	No	Rep.2do	Rep.4to
c = 1	0.98	0.01	0.01
c = 2	0.93	0.03	0.04
c = 3	0.83	0.09	0.07

4. Conclusiones

Los resultados obtenidos y presentados en este trabajo, nos permiten obtener las siguientes conclusiones:

Existe una categorización de 3 clases latentes que permite clasificar a los escolares según su rendimiento académico, competencias comunicación lingüística, matemática, conocimiento e interacción con el mundo físico y social-ciudadana. La categorización realizada por el ACL es fiable, dado que fue verificada a partir de dos técnicas más: QDA y CHAID.

Con los resultados de esta clasificación del rendimiento académico final, se puede distinguir que un 32 % de los estudiantes de cuarto año de primaria en España cuenta con un rendimiento académico alto, un 51 % con un rendimiento académico medio y 17 % con un rendimiento académico bajo.

Un rendimiento académico alto, se caracteriza por presentar valores mayoritariamente altos en las 4 competencias, aunque se detectan también escolares con valores medios en comunicación lingüística y matemática. Un rendimiento académico medio, se caracteriza por presentar un nivel medio en las cuatro competencias. Por último, un rendimiento académico

bajo, se caracteriza por presentar valores medios en matemáticas y bajos en las otras tres competencias.

La mayor proporción de escolares en escuelas privadas reportan un nivel de rendimiento alto, mientras que la mayor proporción de escolares en escuelas públicas reportan un rendimiento académico medio.

En cuanto al hecho de ser repetidores o no, se detectó que la mayor proporción de niños no repetidores cuenta con un rendimiento académico alto, mientras que la mayor proporción de niños repetidores reportan un rendimiento académico bajo.

Referencias

- [1] Agresti, A. (2013). *Categorical data analysis*. Wiley-Interscience, Hoboken, N.J.
- [2] Álvarez Vaz, R. & Vernazza, E. (2019). *Probabilidad, Estadística y sus Aplicaciones Benemérita Universidad de Puebla, México. Capítulo 3. Primera Edición 2019. ISBN: 978-607-525-589-7.*
- [3] Drew A. Linzer, Jeffrey B. Lewis (2011). *poLCA: An R Package for Polytomous Variable Latent Class Analysis*. *Journal of Statistical Software*, 42(10), 1-29. URL <http://www.jstatsoft.org/v42/i10/>.
- [4] Ghassabeth, A.G., Rudzick, F. & Moghaddam, H.A. (2015). *Fast incremental LDA feature extraction*. *Pattern recognition*, 48(6), 1999-2012. doi: 10.1016/j.patcog.2014.12.012.
- [5] Hagenaars, J. (2002). *Applied latent class analysis*. Cambridge University Press, Cambridge New York.
- [6] IBM Corp. Released 2017. *IBM SPSS Statistics for Windows, Version 25.0*. Armonk, NY: IBM Corp.
- [7] Instituto de Evaluación (España).(2010). *Evaluación general de diagnóstico 2009*.
- [8] James, G., Witten, D. & Tibshirani, T.H.R (2013). *An Introduction to Statistical Learning with Applications in R*. En Allen, G., De Veaux, R., Nugent, R. (Ed. De la serie), *Springer Text in Statistics: Cap.4. Clasificación* (pp.138-151). doi: 10.1007/978-1-4614-7138-7.
- [9] Kass, G.V. (1980). *An Exploratory Technique for Investigating Large Quantities of Categorical Data*. *Applied Statistics*, Vol. 29, No. 2., pp. 119-127.
- [10] Montes, J. F. C., Castro, B. E. P., & González, M. G. M. (2007). *Niveles de estrés y rendimiento*



académico en estudiantes de la carrera de Psicología del Centro Universitario de Los Altos. *Revista de Educación y Desarrollo*, 77-82.

- [11] Morgan, J.N. & Sonquist, J. A. (1963). Problems in the Analysis of Survey Data, and a Proposal, *Journal of the American Statistical Association*, 58:302, 415-434, DOI: 10.1080/01621459.1963.10500855.
- [12] R Core Team (2019). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.
- [13] Urruticoechea, A., Vernazza E., del Callejo Canal, D., Canal Martínez, M. & Álvarez Vaz, R. (2020). Estudio comparativo del rendimiento de niños repetidores y no repetidores. *International Journal of Developmental and Educational Psychology INFAD Revista de Psicología*, 1(1) 53-62. ISSN: 0214-9877.
- [14] Venables, W. N. Ripley, B. D. (2002) *Modern Applied Statistics with S*. Fourth Edition. Springer, New York. ISBN 0-387-95457-0