

La predicción del rendimiento académico: regresión lineal versus regresión logística

M^a Visitación García Jiménez, Jesús M^a Alvarado Izquierdo y Amelia Jiménez Blanco
Universidad Complutense de Madrid

El objetivo de este estudio es evaluar la capacidad de la regresión lineal y de la regresión logística en la predicción del rendimiento y del éxito/fracaso académico, partiendo de variables, como la asistencia y la participación en clase, cuya relevancia ya ha sido puesta de manifiesto en anteriores trabajos de nuestro equipo (Alvarado y García Jiménez, 1997). La muestra la constituyeron 175 universitarios de primero de psicología, tomándose los datos en la asignatura de «Métodos y Diseños de Investigación en Psicología I», del área de Metodología. Las conclusiones de este estudio son que (a) el rendimiento previo es un buen predictor del rendimiento futuro y (b) la asistencia y sobre todo la participación son variables con un peso importante en la predicción del rendimiento.

Predicting Academic Achievement: Linear Regression versus Logistic Regression. The main goal of this work was to evaluate the capability of linear regression and logistic regression in predicting academic achievement and academic success. Our study pays special attention to attendance and participation, whose variables that are associated to academic achievement (Alvarado and García Jiménez, 1997). We analyzed data from 175 undergraduates enrolled on their first year of psychology course in the subject of «Methods and Designs of Investigation in Psychology I», in the area of Methodology. The conclusions of this study are that (a) prior academic achievement is a good predictor of the future academic achievement and (b) the attendance and mainly the participation are variables with an important emphasis on the prediction of academic achievement.

El rendimiento académico se concibe como un constructo en el que no sólo se contemplan las aptitudes y la motivación del alumno sino también otras variables intervinientes como los aspectos docentes, la relación profesor-alumno, el entorno familiar, etc. Desde este enfoque multidisciplinar, Forteza (1975) define el rendimiento académico como «*la productividad del sujeto, el producto final de la aplicación de su esfuerzo, matizado por sus actividades, rasgos y la percepción más o menos correcta de los cometidos asignados*». No obstante, a la hora de operativizar el rendimiento, tal como apunta González (1988) se tiende al reduccionismo, así, en la bibliografía observamos que la mayor parte de las investigaciones toman de él dos tipos de medidas: «*las pruebas objetivas*» y «*las calificaciones del profesorado*» que, como muestran Marrero y Espino (1988) son entre sí medidas complementarias: «*Así, mientras que las notas recogen variables importantes referidas al individuo, a su contexto y a la interacción entre ambas, las pruebas objetivas miden el conocimiento adquirido sin considerar especialmente otras variables importantes, pero de una forma más objetiva*». Estos mismos autores analizaron el poder predictivo de las distintas aptitudes, mediante regresión múltiple, concluyendo que la más importante predictora del rendimiento académico es la

verbal, seguida de la aptitud numérica y del razonamiento, ocupando los últimos lugares de la jerarquía las aptitudes mecánica y espacial, todo lo cual coincidiría con los resultados encontrados por la investigación empírica pertinente, aunque el orden de la jerarquía varía según se trate de predecir áreas específicas de rendimiento académico.

Marcelo, Villarín y Bermejo (1987) tras un pormenorizado estudio de los predictores del rendimiento académico seleccionaron distintas escalas y subescalas para la medida de la personalidad, inteligencia, percepción del ambiente de clase, habilidades específicas y destrezas en técnicas de trabajo intelectual. Encontraron que las variables «*planificación del estudio*» (I.H.E. 2), «*inteligencia (O.T.I.S)*», «*apoyo del profesor*» (E.A.C. 3), «*estudio*» (ILLU 5), «*tiempo*» (ILLU 2), «*condiciones ambientales de estudio*» (I.H.E. 1) e «*implicación*» (E.A.C. 1) entraban a formar parte de la ecuación de predicción de regresión múltiple, explicando un 25,70 % de la varianza del rendimiento escolar (promedio de las calificaciones) en cursos de bachillerato; siendo la variable «*planificación del estudio*» con un coeficiente de correlación parcial de 0.29, la que mayor contribución presentaba a la explicación del rendimiento académico. Marcelo y otros (1987) concluyen que las variables predictoras de mayor importancia son aquellas que se relacionan con las técnicas de trabajo intelectual, por lo que recomiendan para la reducción del fracaso escolar el entrenamiento de los alumnos mediante el diseño de un Taller de Técnicas de Trabajo Intelectual. En esta línea, Castejón, Navas y Sampascual (1993) utilizando las técnicas de regresión múltiple y análisis causal llegan a la conclusión de que los factores individuales de «ren-

dimiento inicial», «nivel socioeconómico» y «autoconcepto» son los que tienen un mayor peso a la hora de explicar el rendimiento medio del alumno en educación secundaria.

Jiménez (1987) estudia la problemática del rendimiento escolar en un entorno universitario peculiar como es la universidad a distancia (UNED), caracterizado por un bajo rendimiento escolar y un elevado índice de abandono. Los datos con los que realiza su investigación «ex-post-facto» son de tipo académico (curso, número de asignaturas matriculadas, carrera, calificación obtenida, etc.) y de tipo sociopersonal (edad, sexo, desempeño o no de actividad profesional, tipo de actividad desempeñada en su caso, etc.). La técnica que utiliza es el análisis discriminante para conocer que variables son las que mejor definen a los grupos con buen o mal rendimiento (mediana de la distribución de las notas) y obtener los coeficientes discriminantes que proporcionan el poder predictivo de las variables. De las 15 variables con poder discriminante para el conjunto de las muestras una apareció en las cuatro muestras, «la primera asignatura matriculada». En cuanto a los alumnos presentados a examen de primer curso en la UNED el buen/mal rendimiento estuvo significativamente asociado al hecho de trabajar o no y con la edad del alumno, y al estudiar el rendimiento a lo largo de los distintos cursos se comprobó que el rendimiento anterior es un buen predictor del rendimiento futuro. Otro estudio posterior en la UNED, utilizando la técnica del análisis discriminante fue realizado por García Llamas (1986), en el cual se pasó un cuestionario a los alumnos de la asignatura «Ciencias de la Educación», encontrándose que los mejores predictores del rendimiento fueron las variables: «autovaloración que realiza el alumno de su propio rendimiento en función de su situación personal», «tiempo semanal dedicado al estudio» y la «frecuencia de las consultas al profesor-tutor».

Otros estudios se han centrado en la predicción del rendimiento académico con el objetivo de mejorar la selectividad o de encontrar una mejor alternativa para seleccionar a los alumnos que con mayor éxito puedan realizar estudios universitarios (por ej. Cuxart, Martí, Ferrer, 1997; Goberna, Lopez y Pastor, 1987). Goberna y otros (1987) buscan un criterio para la selección de los alumnos que deben acceder a la universidad, partiendo de lo inadecuado de utilizar la nota de selectividad o pruebas de aptitud de acceso a la universidad (PAAU) como criterio de acceso, puesto que la nota PAAU sin promediar con el expediente de bachillerato sólo explicaría el 6% del rendimiento posterior. Utilizando como índice de rendimiento académico la media ponderada del expediente, Goberna y otros (1987) inician la búsqueda, utilizando la regresión múltiple, de las asignaturas de bachillerato que resultan ser los mejores predictores del rendimiento en las distintas facultades, encontrando que los predictores y su ponderación varía en función de las distintas titulaciones ofertadas, concluyendo que lo idóneo sería utilizar baremos específicos para el acceso a las distintas titulaciones.

El problema de encontrar buenos predictores del rendimiento futuro de manera que se reduzca el fracaso académico en los programas de postgrado ha recibido una especial atención en EEUU como nos indican Wilson y Hardgrave (1995), en un estudio en el que intentan encontrar la mejor técnica para la predicción del rendimiento académico, con el fin de aumentar la tasa de éxito en un programa de postgrado MBA («master's-level business administration»). En la investigación de Wilson y Hardgrave (1995) se utilizaron como predictores del rendimiento académico distintos factores como «la nota media durante la carrera», la «puntuación en

el test de admisión al postgrado», las «cartas de recomendación», «experiencia profesional», etc. Encontraron que las técnicas de clasificación como el análisis discriminante o la regresión logística son más adecuadas que la regresión lineal múltiple a la hora de predecir el éxito/fracaso académico, puesto que la regresión múltiple tiende a ignorar los casos extremos en rendimiento

En definitiva, la mayoría de las investigaciones han utilizado la técnica de la regresión lineal múltiple para encontrar los mejores predictores del rendimiento académico. No obstante, el valor explicativo de la regresión múltiple parece escaso, puesto que, salvo excepciones, no suele superar el valor $R^2=0.19$, lo cual ha llevado a la búsqueda de técnicas alternativas, que al menos nos permitan pronosticar el éxito/fracaso académico, como el análisis discriminante. El análisis discriminante permite hacer una clasificación correcta (éxito/fracaso académico) en torno al 64% de los casos (Remus y Wong, 1982).

A pesar de que la principal técnica de clasificación utilizada ha sido el análisis discriminante, Wilson y Hardgrave (1995) defienden la utilización de la regresión logística como alternativa al análisis discriminante, escasamente utilizada hasta la fecha. La regresión logística es una técnica adecuada cuando se pretende hacer una clasificación basada en las características de los datos. Una ventaja adicional de esta técnica es que no requiere la normalidad estricta de los datos y además un gran número de investigaciones han evidenciado que hay otras muchas características deseables que hacen de la regresión logística una mejor herramienta para la categorización que el análisis discriminante (ver Aldrich y Nelson, 1984; Press y Wilson, 1978).

Para la realización de este trabajo se elaboró un cuestionario «ad hoc» que contemplara la mayor parte de las variables que han mostrado tener influencia en el rendimiento académico. A partir de este cuestionario y de los resultados de la investigación de Alvarado y García Jiménez (1997), en la que se mostraba la importancia de la asistencia y la participación en clase como factores explicativos del rendimiento académico, pretendemos conocer qué variables son las mejores predictoras del éxito/fracaso en las asignaturas del área de Metodología.

Método

Muestra

La muestra estaba constituida por 175 estudiantes (140 mujeres y 35 hombres) de primer curso de Psicología de la UCM. Los datos fueron tomados durante el curso académico 1997/98 en la asignatura cuatrimestral de «Métodos y Diseños de Investigación en Psicología I».

Variables

Las variables estudiadas fueron las siguientes:

- Rendimiento académico: calificación obtenida en el examen final de febrero en la asignatura estudiada.
- Asistencia a clase: número de veces que asistieron los alumnos a clase en los días en que aleatoriamente se controló su presencia.
- Participación en clase: número de veces que los alumnos salían voluntariamente a corregir los ejercicios de prácticas. Cada vez que los alumnos salían voluntariamente y realizaban correctamente todos los ejercicios que les eran solicitados recibían una calificación de 0,25 puntos que sería posteriormente sumada a la ca-

lificación final que en este trabajo es considerada como rendimiento académico.

d) Cuestionario: las variables que éste contemplaba pretendían medir, en el primer día de clase, aquellos aspectos bibliográficos, motivacionales y actitudinales que han sido documentados como relevantes en la literatura revisada.

1. Calificación media del bachillerato (hasta 7 puntos/mayor de 7 puntos).
2. Nota de acceso a la universidad.
3. Número de opción en que se eligió la carrera de psicología.
4. Supuesta elección de otra carrera de haber tenido más nota. (si/no)
5. Grado de ilusión por la carrera en su comienzo (ilusionado / indiferente/ desilusionado).
6. Supuesta relación del grado de ilusión por la carrera con el rendimiento (si/no)
7. Previsión de las asignaturas de Metodología al elegir Psicología (si/no)
8. Elección de la carrera en caso de haber tenido conocimiento de las asignaturas de Metodología integradas en el Plan de Estudios vigente. (si/no)
9. Estimación del agrado por las asignaturas de Metodología (agrado / no agrado /desagrado).
10. Previsión de dificultades con las asignaturas de Metodología (si/no)
11. Estimación de la motivación al empezar la carrera (mucho/regular/poca)
12. Estimación de la vocación de Psicólogo/a (alta/media/baja)
13. Tipo de interés al elegir la carrera (personal/profesional)
14. Especialidad más atractiva (clínica/educativa/industrial/social/metodología)
15. Personalidad (introvertida/extravertida).
16. Dificultad en salir a la pizarra (mucho/poca/ninguna)
17. Previsión de asistencia a clase (si/no)
18. Ordenación de factores supuestamente influyentes para el buen rendimiento académico: competencia del profesor, esfuerzo personal, capacidad intelectual, interés por las materias, motivación por aprender
19. Influencia de la existencia de un libro de texto para la asistencia en la asistencia a clase (si/no)
20. Estimación actual de asistencia futura a clase (si/no)
21. Enumeración por orden de preferencia de cinco asignaturas de la carrera.

Técnicas de análisis

Las técnicas de análisis de datos empleadas fueron la regresión lineal múltiple para corroborar el efecto de la asistencia y la participación real en la predicción del rendimiento académico, así como la regresión logística para detectar las variables que mejor permiten predecir el éxito/fracaso académico.

Resultados

A) Regresión lineal múltiple:

Mediante el procedimiento de regresión por pasos encontramos el mejor modelo predictor del rendimiento académico ($F_{3,171} = 12.14212$, $p < 0.01$) obteniéndose un valor de Regresión Múltiple igual a 0,41906, siendo las variables predictivas que entran a for-

mar parte de la ecuación: calificación durante el bachillerato «CB» ($t = 3,316$, $p < 0.01$), participación «PAR» ($t = 2,748$, $p < 0.01$) y asistencia «ASIS» ($t = 2,260$, $p = 0,0251$):

$$\text{Rendimiento} = 2.522246 + 0.814101CB + 1.125477PAR + 0.083810ASIS$$

Confirmamos así la importancia de las variables participación y asistencia como variables predictoras sobre el rendimiento (Alvarado y García, 1997) y que a diferencia del rendimiento parecen ser variables sobre las que, en principio, es posible actuar, para lo cual sería necesario conocer cuales son los factores que facilitan una mayor asistencia y participación. Para este fin se realizó un análisis de regresión múltiple que nos permitiera conocer cuales son las variables predictoras de la participación y de la asistencia:

Para la participación, encontramos cuatro variables, que proporcionaban un valor de Regresión Múltiple igual a 0,41376, para las variables calificación durante el bachillerato «CB» ($t = 3,934$, $p < 0.01$), motivación al comenzar la carrera «MOT» ($t = 2.456$, $p = 0.0151$), «timidez» o dificultad en salir a la pizarra «PIZ» ($t = 2.453$, $p = 0.0151$) e importancia que confieren al esfuerzo personal para superar la materia «ESF» ($t = 2.081$, $p = 0.0390$):

$$PAR = 0.172095CB + 0.112153MOT - 0.079409PIZ + 0.038786ESF$$

En cuanto a la asistencia, encontramos seis variables para un valor de Regresión Múltiple igual a 0,50021, siendo las variables predictoras: agrado de la asignatura «AGR» ($t = 3,978$, $p < 0.01$), personalidad introvertida «PER» ($t = 3.372$, $p < 0.01$), previsión de asistencia «PAS» ($t = 2,874$, $p < 0.01$), especialidad metodología «ESP» ($t = 2.551$, $p = 0.0116$), sentirse motivado al iniciar la carrera «MOT» ($t = 2.452$, $p = 0.0152$) y atribuir a la competencia del profesor éxito académico «PROF» ($t = 2,028$, $p = 0,0441$):

$$ASIS = 19.174 + 3.288AGR + 0.769PER + 3.020PAS + 2.021ESP + 1.372MOT + 0.310PROF$$

B) Regresión logística

Para la predicción del éxito/fracaso académico utilizamos la regresión logística categorizando la variable dependiente en aprobados (90 alumnos) y suspensos (85 alumnos). El modelo permite hacer una estimación correcta ($\chi^2_4 = 28.424$, $p < 0.01$) del 68,57% de los casos (ver Tabla 1), entrando a formar parte de la ecuación como predictores las variables: participación en clase «PAR» (Wald= 7,7636, $p < 0.01$), considerar la especialidad de cognitiva-metodología como la más atractiva «ESP» (Wald= 6,0533, $p = 0.0139$), calificación durante el bachillerato «CB» (Wald= 5.9532, $p = 0.0147$) y el estar motivado al iniciar la carrera «MOT» (Wald= 4.2451, $p = 0.0394$):

$$\text{Rendimiento} = \frac{1}{1 + e^{-n}}; H = 1.9085PAR + 0.8534ESP + 0.4081CB + 0.3617MO;$$

En la figura 1 se observa que todos los casos con probabilidad superior a 0,75 fueron correctamente clasificados como aprobados, mientras que la mayor parte de los errores en la clasificación fue para los sujetos a los que se predecía una probabilidad entre 0.25 y 0.50.

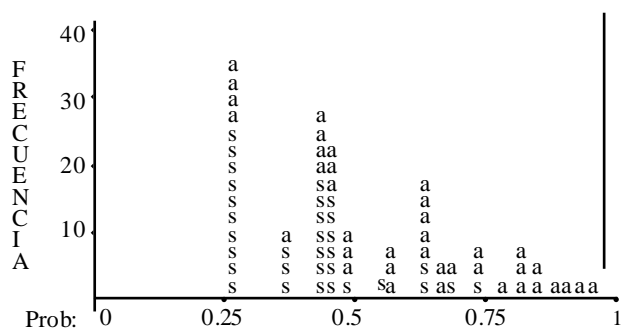


Figura 1. Frecuencia de casos observados en función de las probabilidades predichas. Cada símbolo representa 2,5 casos, habiéndose utilizado el símbolo «s» para indicar los suspensos observados y el símbolo «a» para los aprobados observados. Se predice el aprobado para probabilidades superiores a 0.50, destacándose los casos mal clasificados

También se realizó un análisis, mediante regresión logística, para conocer que aspectos predicen la participación y la asistencia.

En el caso de la participación, se dicotomizó la VD en los alumnos que en alguna ocasión habían salido a corregir problemas en la clase práctica (63 alumnos) y los que no habían salido en ninguna ocasión (112). El modelo permite hacer una estimación correcta ($\chi^2_4 = 25.794, p < 0.01$) del 68% de los casos (ver Tabla 2), entrando a formar parte de la ecuación como predictores

Tabla 1
Número y porcentaje de casos correctamente clasificados en la predicción del éxito/fracaso académico

OBSERVADO	PRONOSTICADO		% casos acertados
	Suspensos	Aprobados	
Suspensos	67	18	78,82
Aprobados	37	53	58,89
% casos acertados	64,42	74,65	68,57

Tabla 2
Número y porcentaje de casos correctamente clasificados en la predicción de la participación en clase

OBSERVADO	PRONOSTICADO		% casos acertados
	Participa	No participa	
Participa	102	10	91,07
No participa	46	17	26,98
% casos acertados	68,92	62,96	68,00

Tabla 3
Número y porcentaje de casos correctamente clasificados en la predicción de una alta asistencia a clase

OBSERVADO	PRONOSTICADO		% casos acertados
	Asiduo	No asiduo	
Asiduo	34	58	36,96
No asiduo	10	73	87,95
% casos acertados	77,27	55,73	61,14

las variables: grado de ilusión al comenzar la carrera «ILU» (Wald= 6.2253, p= 0.0126), calificación durante el bachillerato «CB» (Wald= 6.5224, p= 0.0126), «timidez» o dificultad en salir a la pizarra «PIZ» (Wald= 4,8126, p= 0.0283) y nivel de agrado de las asignaturas de metodología «AGR» (Wald= 3,6271, p= 0.0568):

$$PAR = \frac{1}{1 + e^{-H}}; H = -1.9064 + 0.6444ILU + 0.4253CB - 0.3771PIZ + 1.0291AGR$$

En cuanto a la asistencia, se dicotomizó la variable en los alumnos que firmaron más del 85% de los partes de asistencia (83 alumnos) y los que no habían asistido con tanta asiduidad (92 alumnos). El modelo permite hacer una estimación correcta ($\chi^2_4 = 24.056, p < 0.01$) del 61,14% de los casos (ver Tabla 3), entrando a formar parte de la ecuación como único predictor el sentirse ilusionado al inicio de la carrera «SEN» (Wald= 4.7105, p= 0.030):

$$ASIS = \frac{1}{1 + e^{-0.4616SEN}}$$

Conclusiones

El procedimiento de regresión múltiple no nos permite hacer un buen pronóstico del rendimiento académico ($R^2 = 0.17$), mientras que la regresión logística si parece ser un instrumento idóneo para hacer una buena predicción del éxito/fracaso académico como muestra el casi 70% de casos bien clasificados que hemos obtenido (véase Remus y Wong, 1982).

Se confirma que el mejor predictor del rendimiento académico futuro es el rendimiento anterior, como han puesto en evidencia múltiples estudios (Goberna y otros, 1987; House, Hurst, Keely 1996; Jiménez, 1987; Wilson y Hardgrave, 1995).

Los resultados de nuestro trabajo confirman el criterio de Goberna y otros (1987), puesto que se muestra el menor valor pronóstico de la nota de selectividad respecto al rendimiento continuado (rendimiento medio durante el bachillerato) en cuanto a la predicción del rendimiento académico en la universidad. Tanto si se utiliza la regresión logística como si se utiliza la regresión múltiple, los mejores modelos contemplan la variable rendimiento medio durante el bachillerato en vez de la nota en selectividad.

Además del rendimiento académico anterior, se evidencia la importancia de las variables participación y asistencia (Alvarado y García Jiménez, 1997) en la predicción del rendimiento. Esto tiene una especial transcendencia, al tratarse de variables sobre las que es posible actuar. Para la participación en clase serían buenos predictores, la elevada motivación al iniciar la carrera, la menor dificultad para salir a la pizarra y el alto nivel de agrado por la asignatura impartida. Mientras que para la asistencia aparece como factor clave la ilusión/motivación ante el comienzo de la carrera, aunque aspectos como la competencia del profesor y el agrado de la asignatura también tienen una incidencia positiva sobre la asistencia.

Nos parecen suficientemente sugestivas estas conclusiones, para seguir indagando nuevos predictores y perfilando nuevos estudios, a partir de la regresión logística, dado el alto porcentaje de clasificaciones correctas obtenidos en este estudio.

Referencias

- Aldrich, J.H., y Nelson, F.D. (1984). *Linear probability, logit, and probit models*. Beverly Hills, CA: Sage.
- Alvarado, J. y García Jiménez, M.V. (1997). Incidencia de la asistencia a clase, del trabajo efectivo y de factores individuales sobre el rendimiento académico. *V Congreso de Metodología*. Sevilla.
- Castejón, J.L., Navas, L. y Sampascual, G. (1993). Investigación sobre la eficacia de centros de enseñanza secundaria. Un modelo de identificación y funcionamiento. *Revista de Educación*, 301, 221-244.
- Cuxart Jardí, A., Martí Recober, M. y Ferrer Julià (1997). Algunos factores que inciden en el rendimiento y la evaluación en los alumnos de las pruebas de acceso a la universidad. *Revista de Educación*, 314, 63-88.
- Forteza, J. (1975). Modelo instrumental de las relaciones entre variables motivacionales y rendimiento. *Revista de Psicología General y Aplicada*, 132, 75-91.
- García Llamas, J.L. (1986). El análisis discriminante y su utilización en la predicción del rendimiento académico. *Revista de Educación*, 280, 229-252.
- Goberna, M.A., López M.A. y Pastor J.T. (1987). La predicción del rendimiento como criterio para el ingreso en la universidad. *Revista de Educación*, 283, 235-248.
- González, A.J. (1988). Indicadores del rendimiento escolar: relación entre pruebas objetivas y calificaciones. *Revista de Educación*, 287, 31-54.
- House, J. D., Hurst, R.S., Keely, E.J., (1996). Relationship between learner attitudes, prior achievement, and performance in a General Education Course: A multi-Institutional Study. *International Journal of Instructional Media*, 23, 257-271.
- Jiménez Fernández, C. (1987). Rendimiento académico en la universidad a distancia. Un estudio empírico sobre su evolución y predicción (II). *Revista de Educación*, 284, 317-347.
- Marcelo García, C., Villarín Martínez, M. y Bermejo Campos, B. (1987). Contextualización del rendimiento en bachillerato. *Revista de Educación*, 282, 267-283.
- Marreno Hernández, H. y Orlando Espino, M. (1988). Evaluación comparativa del poder predictor de las aptitudes sobre notas escolares y pruebas objetivas. *Revista de Educación*, 287, 97-112.
- Press, S.J. y Wilson, S. (1978). Choosing between logistic regression and discriminant analysis. *Journal of American Statistical Association*, 73, 699-705.
- Remus, W., y Wong, C. (1982). An evaluation of five models for the admission decision. *College Student Journal*, 16, 53-59.
- Wilson, R.L., Hardgrave, B.C. (1995). Predicting graduate student success in an MBA program: Regression versus classification. *Educational and Psychological Measurement*, 55, 186-195.