

ACERCA DE LA MEDICIÓN DE LA FIRMEZA FRENTE AL DELITO MEDIANTE PREGUNTAS ÚNICAS. UN ANÁLISIS DE CLASES LATENTES*

ALFONSO SERRANO MAÍLLO

Departamento de Derecho penal y Criminología, UNED

Resumen: La firmeza frente al delito o punitividad se ha convertido en uno de los principales objetos de estudio de la Criminología. Sin embargo, su concepto, naturaleza y medición son cuestiones muy complejas. Una prueba de ello son las dudas que encierra la dimensionalidad.

Utilizando datos de una encuesta telefónica de ámbito nacional que incluye algunas preguntas sobre la materia, en el presente trabajo se trata de testar la hipótesis de que, a nivel individual, *existe una variable latente de firmeza frente al delito con indicadores categóricos de actitudes ciudadanas*. Como novedad, se incluye una variable observada de carácter procesal-penal.

A tal fin, se recurre aquí a una herramienta estadística poderosa y cada vez más importante en ciencias humanas y sociales, sobre todo en el caso de indicadores categóricos u ordinales, como son los análisis de clases latentes. Esta estrategia ofrece la oportunidad de testar la hipótesis de unidimensionalidad, a la vez que también permite una aproximación al error de medición. Entre sus dificultades se encuentran sus relativamente fuertes asunciones, cuya comprobación debe asegurarse. En el presente trabajo también se testa la existencia

* Este trabajo se enmarca en el Proyecto de Investigación SEJ-2007/G7312/JURI, financiado por el Ministerio de Educación y Ciencia, cuyo Investigador Principal he sido yo mismo.

de una clase latente de «ideólogos» y se valora la influencia de los datos perdidos en nuestros modelos.

Palabras clave: firmeza frente al delito, punitividad, preguntas únicas, análisis de clases latentes, error de medición.

Abstract: Firmness towards crime or punitivity has become one of the most important topics of study in Criminology. Nevertheless, its concept, nature and measurement are extremely complex issues. The doubts in the area of dimensionality are just a good proof of this complexity.

With data from a national telephone survey that included some questions about this issue, we test the hypothesis that there exists, at the individual level, a latent variable of firmness towards crime with categorical indicators of individual attitudes. It is worth mentioning that the data include an observed variable of a procedural-law nature.

To fulfill this end, we use here a powerful statistical tool, which is becoming more and more important in human and social sciences, especially in the case of categorical and ordinal indicators, as is the case of latent class analysis. This opens the door to test the unidimensionality hypothesis, as well as an approximation to measurement error. Its main difficulties include that assumptions are rather strong, so that they must be carefully checked. The potential presence of a latent class of «ideologues» is also tested, as well as the impact of missing data in our models.

Key words: firmness towards crime, punitivity, single questions, latent class analysis, measurement error.

1. INTRODUCCIÓN

Importantes cambios que se vienen observando desde hace algunas pocas décadas en las legislaciones jurídico-penales y sancionadoras en general, en las tasas de personas privadas de libertad, en los derechos fundamentales y las garantías individuales, en los Sistemas de Administración de Justicia y su actuación, etc., se han convertido en uno de los objetos de estudio más trascendentales de la Criminología contemporánea¹. Para referirse a estos desarrollos, la literatura utiliza términos tales como *punitivismo*, *punitividad*, *firmeza frente al delito* y algunos otros. Aunque algunos de ellos claramente tienen significados específicos, como *populismo punitivo*, en ocasiones se utilizan todos de modo intercambiable. Aquí utilizaré *firmeza frente al delito* —y su

¹ SERRANO MAÍLLO, 2006: 245-252.

opuesto *benevolencia frente al delito*— para referirme a desarrollos sistémicos como los descritos más arriba, así como a las actitudes ciudadanas sobre las mismas o semejantes cuestiones. El motivo reside en que este sustantivo tiene una menor carga ideológica y valorativa, siendo el caso que las ciencias humanas y sociales deben, como regla, tratar sus objetos de estudio con neutralidad.

La firmeza frente al delito es un concepto muy complejo. Además, no sólo se sabe relativamente poco sobre sus correlatos, causas y explicación y las formas de contrarrestarla², sino que ya su propia definición, naturaleza y la forma en que puede ser aproximada científicamente presenta serias dudas³. Ante tal complejidad, no es extraño que algunos ilustres criminólogos como nada más y nada menos que Helmut KURY hayan llegado incluso a mostrar un sano escepticismo sobre las posibilidades de encontrar en la firmeza frente al delito un objeto susceptible de estudio científico⁴.

Así las cosas, no es extraño que, desde esta perspectiva, muchos de los enfoques de estudio actuales puedan calificarse de limitados⁵. Ello se aprecia fácilmente, verbigracia, ya en lo que constituye el objeto principal de la presente investigación: la *medición* de la firmeza frente al delito o de la punitividad *a nivel individual*. KURY y sus colegas, por ejemplo, afirman que «Los resultados de los estudios de opinión pública en lo que se refiere a las actitudes punitivas [...] sugieren que la metodología necesita refinarse», así como que «Los métodos más habitualmente utilizados para evaluar (*assessing*) la punitividad son por lo tanto incompletos»⁶. En especial, éste es el caso de medidas basadas en preguntas únicas y a menudo limitadas al castigo o a alguno de sus aspectos⁷ —que prácticamente ha sido la

² KUHN, 1993: 271; KURY, OBERGFELL-FUCHS y WÜRGER, 2002: 3; KURY et al., 2004: 51-52.

³ KURY et al., 2004: 52-64; SERRANO MAÍLLO y KURY, 2008: 342; STALANS, 2002: 15-29; VIKI y BOHNER, 2009: 96-115.

⁴ KURY y PUTKARADZE, 2009.

⁵ MAYHEW y VAN KESTEREN, 2002: 66-67; TONRY y FARRINGTON, 2005: 31.

⁶ KURY et al., 2008: 130 y 133.

⁷ CULLEN et al., 2000: 7; CULLEN et al., 2009: 77; HARTNAGEL y TEMPLETON, 2008: 359; KEIL y VITO, 1991: 455; KURY, OBERGFELL-FUCHS y WÜRGER, 2002: 4 y 170; KURY et al., 2004: 97; MAYHEW y VAN KESTEREN, 2002: 67; NELLIS y LYNCH, 2008: 34, 37, 41 y 46; SPOTT, 1999b: 468 y 473. Aunque aquí nos referiremos al nivel de análisis individual, estas mismas observaciones son aplicables a la práctica agregada, que también abusa de medidas únicas, por ejemplo sobre tasas de encarcelamiento; vid. un enfoque alternativo, con varias mediciones independientes, en BLUMSTEIN et al., 2005: 349 sobre todo. Consecuentemente, aunque aquí se utiliza el término «preguntas» porque lo que nos interesa es el nivel individual, el texto es extensible a ítems, variables observadas, indicadores, etc.

regla tradicionalmente—. En otras ocasiones se utiliza una pluralidad de preguntas, pero ello no siempre asegura una medición óptima si las mismas no se combinan de modo correcto —cuando ello es posible—. Aunque todo ello es bien conocido, menos abundantes son los esfuerzos por justificar empíricamente esta apreciación —más allá, por supuesto, del caso general, aunque matizable, de que más preguntas válidas ofrecen mayor información, por ejemplo en el marco de la teoría clásica de los tests—.

En este trabajo, pues, se ofrecerán algunas reflexiones sobre la medición de la firmeza frente al delito a nivel individual tal y como habitualmente se hace en la práctica contemporánea⁸. Del mismo modo, nos preguntaremos por la relativamente poco investigada cuestión de su dimensionalidad⁹. Preocupaciones como las precedentes recomiendan, desde un punto de vista metodológico, enfoques más bien sofisticados como el de variables latentes¹⁰, en particular para testar la hipótesis de que existe un factor o variable subyacente —o varios—, no manifiesto del que se observan ciertos indicadores¹¹. Permítame el amable y crítico lector desarrollar un poco más esta idea con la ayuda de un ejemplo. SPROTT, en un artículo no tan antiguo subtítulo «Las dimensiones de la “punitividad” pública», insiste en que «Los resultados presentados [...] contradicen la idea de que la punitividad es un concepto simple, unidimensional [...] la punitividad, como concepto, es complejo»; a la vez que «rechaza el uso de medidas simples, únicas para valorar la punitividad»¹². Sería ingenuo ignorar que esta postura —la del primer inciso— es probablemente la más plausible. Sin embargo, esta autora se limita a comprobar que, para las diversas preguntas que incluye en su trabajo, y que siempre utiliza de modo aislado, no siempre el sexo es una variable relevante. Dicho de otro modo, mujeres y hombres parecen igual de punitivos para algunas preguntas y en otras las mujeres sólo son más benévolas bajo ciertas circunstancias, por ejemplo cuando se trata de castigar a jóvenes¹³. En otra publicación con un subtítulo casi idéntico, la misma autora insiste en su rechazo a la opción unidimensional y en favor de una bidimensional con actitudes sobre una punitividad amplia y otra para casos específicos¹⁴. De nuevo, no existen esfuerzos por explorar empíricamente si alguna variable latente

⁸ KURY y OBERGFELL-FUCHS, 2008: 277-302.

⁹ En este mismo sentido, SPROTT, 1999a: 10.

¹⁰ BIEMER, 2011: 116-119; BOLLEN, 1989: 11-20; POWERS y XIE, 2000: 9-11.

¹¹ AGRESTI, 2002: 441 y 539.

¹² SPROTT, 1999b: 472-473.

¹³ SPROTT, 1999b: 468-473; vid. también la misma 1999a: 11, 69-97 y 199-200.

¹⁴ SPROTT, 1999a: 9-10, 14, 46-48 y 199-200.

subyace a las observaciones, sino que simplemente se abunda en que los factores explicativos (o factores predictivos) de una y otra dimensión serían en buena medida distintos¹⁵. Sin embargo, nada en los estudios de Sprott permite excluir efectos, por ejemplo, de la muestra y su tamaño o de errores serios de medición; ni mucho menos excluye, en sentido estricto, que una solución unifactorial de segundo grado subyaga a los items con que cuenta; o que incluso los items compartan una causa común. A tal fin se requieren enfoques distintos, como el de variables latentes. Esta será, en efecto, la estrategia analítica que seguiremos en la presente investigación.

En la presente publicación, como se dijo, nos ubicaremos en el nivel individual de análisis y en aspectos relacionados con el castigo y nos centraremos en la *identificación* del (potencial) constructo subyacente que denominamos *firmeza frente al delito* a partir de una serie de variables categóricas —algunas de ellas en realidad ordinales— que se hipotetiza están causadas por aquélla, esto es que son indicadores suyos. Dicho con otras palabras, nuestra (primera) hipótesis sustantiva nuclear es la existencia de una variable latente categórica de *firmeza frente al delito* —cuyos indicadores, como acaba de señalarse, son también categóricos (H_1)—. Como también se verá, nuestro test de H_1 mediante análisis de clases latentes abre las puertas, dentro de unos márgenes, a una valoración del *error de medición* —algo sobre lo que nos extenderemos más abajo con un cierto detalle—.

Nuestro estudio, por otro lado, tiene la particularidad de incluir, junto a preguntas habituales sobre el castigo, una de naturaleza procesal-penal¹⁶. Puesto que la firmeza frente al castigo puede guardar alguna relación *indirecta* con aspectos procesales del mismo en cuanto que, de modo mediato, los requisitos de la actuación policial, los medios de prueba admitidos, el sistema de recursos, etc. deben influir, en igualdad de condiciones, en el castigo y su probabilidad; merece la pena tratar de profundizar en la naturaleza de la firmeza frente al delito desde un punto de vista empírico, en particular si nos encontramos ante un constructo uni- o pluridimensional —aunque intuitivamente esta última parezca la hipótesis más plausible—. Algún estudio sobre la punitividad ha incluido interpelaciones de esta naturaleza, al menos en parte, procesal-penal. Este es el caso de SECRET y JOHNSON, quienes incluyen la siguiente pregunta: «¿Considerando todo en conjunto, diría que, en general, aprueba Vd. las es-

¹⁵ SPOTT, 1999a: 17-45 y 199; también es importante reseñar que la autora concede un cierto grado de relación entre ambas dimensiones, 199.

¹⁶ En alguna investigación se han realizado preguntas semejantes, vid. ROHNE, 2008: 169.

cuchas (*wiretapping*)?»¹⁷; a la par que existen pruebas empíricas de la existencia de un constructo (latente) de punitividad con indicadores de actitudes ciudadanas sobre derechos fundamentales procesal-penales, si bien su relación con actitudes directamente conectadas con el castigo no es conocida¹⁸. Así, nuestra segunda hipótesis, que añade las anteriores consideraciones, es que *existe una variable latente categórica unidimensional de firmeza frente al delito con indicadores también categóricos de actitudes ciudadanas directas e indirectas sobre el castigo* (H₂). Puesto que el concepto *unidimensionalidad* aparece en ocasiones con significados distintos, aquí la entendemos como «la existencia de un rasgo [una variable] latente que subyace a los datos»¹⁹.

Una vez valoradas las primeras, testaremos la hipótesis de que nuestra (potencial) variable latente tiene una naturaleza continua (H₄). A mayor abundamiento, la literatura ha sugerido la existencia de una clase latente específica de individuos con opiniones muy consistentes en determinadas materias, a los cuales denomina «ideólogos»²⁰ —una hipótesis (H₃), pues, que también testaremos—. Nótese que, independientemente de nuestras conclusiones sobre todo para nuestras dos primeras hipótesis, partimos de la base de que items observados como los nuestros pueden perfectamente formar parte de un estudio empírico tal y como se lleva a la práctica en la investigación contemporánea.

2. EL PRESENTE ESTUDIO

2.1. Hipótesis

Siguiendo la lógica descrita en el apartado precedente y siempre a nivel de análisis individual, nuestras hipótesis sustantivas son las siguientes:

H₁. Existe una variable latente categórica de firmeza frente al delito con indicadores también categóricos de actitudes ciudadanas sobre el castigo.

H₂. Existe una variable latente categórica unidimensional de firmeza frente al delito con indicadores también categóricos de actitudes ciudadanas directas e indirectas sobre el castigo.

¹⁷ SECRET y JOHNSON, 1989: 365; vid. también, con ulteriores ejemplos, BROWN, 2006: 306; y SPROTT, 1999a: 17.

¹⁸ SERRANO MAÍLLO, 2011: 12-15 y 29.

¹⁹ HATTIE, 1985: 139.

²⁰ McCUTCHEON y MILLS, 1998: 87.

H_3 . Existe en la variable latente categórica firmeza frente al delito una clase latente de «ideólogos».

H_4 . La variable latente firmeza frente al delito tiene una naturaleza continua.

2.2. Datos

Para testar nuestras hipótesis se llevó a cabo, en el marco de una investigación más amplia, una encuesta a una muestra de 953 individuos residentes en España mayores de 18 años. Sólo se consideraron teléfonos fijos, dejando fuera los móviles²¹. El muestreo siguió una afijación proporcional según el peso poblacional del cruce del tamaño de hábitat y Comunidad Autónoma. Se recurrió a la introducción de cuotas proporcionales para sexo y edad. Se utilizó un marco muestral de números telefónicos, lo cual es habitual en la investigación comercial. El modo utilizado, entonces, fue el telefónico²², con asistencia de ordenadores (CATI)²³. El trabajo de campo de la encuesta tuvo lugar en la primavera del pasado año. Debido a desviaciones de los objetivos muestrales, se recurre a ponderaciones por sexo, edad, Comunidad Autónoma y tamaño del hábitat²⁴.

Como puede comprobarse, pues, se trata de un diseño muestral no probabilístico. Aunque en sus primeros pasos el mismo es muy parecido al de estratificación, con la introducción de cuotas y la selección deliberada del entrevistado por parte del encuestador se quiebra de modo definitivo e irreversible el enfoque probabilístico y la intro-

²¹ Los muestreos que incluyen móviles no son muy habituales, por no mencionar que encierran serias complicaciones, cuyas vías de solución, por si fuera poco, no son bien conocidas, KUUSELA et al., 2008: 111-112; STEEH, 2008: 233-234. Una razón que fue definitiva en nuestro estudio, en el marco de la *perspectiva total* (1), fue el coste. En nuestro caso particular parecía injustificado cuando no había visos de alcanzar representatividad en nuestra muestra. (1) FOWLER, 2002: 7-8, 159-160 y 164-165.

²² Hasta hace poco, prácticamente han representado el modo de recogida de datos por defecto. Desde hace algún tiempo, sin embargo, han ido apareciendo serias dificultades que están obligando a buscar nuevas soluciones y adaptaciones, vid. así TUCKER y LEPKOWSKI, 2008: 3-26.

²³ GROVES et al., 2004: 139; LAVRAKAS, 1993: 17-18 y 96.

²⁴ Es una práctica habitual, STEEH, 2008: 230. Sobre la introducción de ponderaciones en estudios como el presente, vid. BIEMER y CHRIST, 2008: 317-341; KALSBECK y AGANS, 2008: 44-52.

Debido a este procedimiento y a los consiguientes redondeos, el número total se ha visto elevado a 955 —en realidad 954,529, lo cual representa un aumento del 0,16% respecto al número originario y carece de relevancia alguna para nuestros análisis—, que pasará a ser nuestro tamaño muestral final.

ducción de sesgos se vuelve probable. También debe mencionarse entre paréntesis que la utilización de un marco muestral de teléfonos, que como acaba de decirse es una práctica habitual en las encuestas telefónicas españolas, representa en realidad una caja negra de muy difícil valoración²⁵. Las consecuencias, así, son dramáticas para funciones habituales de los estudios de encuesta como las descripciones generales o algunos tests de hipótesis. Aunque se han propuesto formas de mejorar este diseño, se requiere un buen modelo cuyos planteamientos rara vez se cumplen en la práctica²⁶.

Con algo más de detalle sobre las problemáticas cuotas, puede señalarse que, para empezar, las mismas en realidad no suelen tener un significado unívoco y a menudo su definición es arbitraria. La elección de los entrevistados, aunque las cuotas imponen limitaciones, deja un importante grado de libertad a los entrevistadores. No sólo eso, igual de importante es que la selección de unos participantes no es independiente de la de los otros, lo cual viola una de las claves de la probabilidad. Como consecuencia, cuando se introducen cuotas no es posible obtener errores muestrales. Así las cosas, el muestreo por cuotas es claramente no probabilístico y sus datos no pueden emplearse para ningún tipo de inferencia, el control es bajo y no existen formas de valorar objetivamente la calidad²⁷.

Aunque durante el diseño del muestreo se consideró un planteamiento probabilístico y evitar, por lo tanto, el uso de cuotas, esta opción fue finalmente desestimada. Varios motivos concurren a explicar esta decisión, tales como, sin ánimo de exhaustividad, los costes que implicaba añadir preguntas para seleccionar a los entrevistados de modo más puramente aleatorio; la falta de familiaridad con el procedimiento de la empresa que se ocupó del trabajo de campo y de sus empleados; o las desviaciones que se anticipaban por razones tales como la no respuesta, cada uno de los cuales hubiera probablemente hecho estériles los esfuerzos²⁸. Sin embargo, para el test de determinadas hipótesis —y ésta es la orientación de este trabajo— a menudo no se requieren muestras representativas —aunque, por supuesto, ello depende de la propia hipótesis—. Por lo tanto, estas reflexiones no deberían ser una preocupación para nosotros aquí.

²⁵ Vid. GONZÁLEZ SÁNCHEZ, 2010: 257-258; KALSBECK y AGANS, 2008: 29-32.

²⁶ Vid. una interesante propuesta de modelo en DEVILLE, 1991: 163-180.

²⁷ Sobre todo lo anterior, vid. COCHRAN, 1977: 135; DEMING, [1960]: 131; DOHERTY, 1994: 21-28; GANNINGER, 2011.

²⁸ Por ejemplo, para el primer punto y sus poco despreciables dificultades, vid. KALSBECK y AGANS, 2008: 39-42.

Habiendo dicho lo anterior, no está de más insistir en que es lamentable que nuestra tradición siga utilizando cuotas en los estudios de encuesta. En efecto, esta es la práctica del Centro de Investigaciones Sociológicas, de la inmensa mayoría de las empresas comerciales y de numerosos estudios de investigación científica, desde luego en el ámbito de nuestra disciplina. Es importante ser consciente de las tremendas limitaciones de los diseños no probabilísticos para funciones descriptivas e incluso de test de ciertas hipótesis que exigen inferencias estadísticas que, debe insistirse, los métodos por cuotas no permiten. A menudo es el caso que el coste es menor, pero, como señala DEMING, el coste carece de significado sin una medida de la calidad de la encuesta²⁹. El cambio de esta orientación debe fijarse entre los principales objetivos de nuestras ciencias humanas y sociales.

La literatura insiste en la importancia del entrenamiento de los entrevistadores³⁰. Con ello se aspira a controlar una importante fuente de error no muestral y a reducir la varianza³¹. Aunque el estilo de entrenamiento flexible o conversacional cuenta con partidarios, un estilo más estandarizado favorece los objetivos recién mencionados. El *entrenamiento estandarizado* idealmente incluye, siguiendo a la doctrina, ejercitar a los entrevistadores para que tengan cuidado en evitar cualesquiera actos, expresiones o comentarios que puedan implicar crítica, sorpresa, aprobación o desaprobación; para que formulen las preguntas tal y como han sido redactadas; y lentamente, para que se entiendan perfectamente; para que realicen las interrogaciones siguiendo el orden del cuestionario o, en su caso, que les proponga el sistema de asistencia por ordenador; por supuesto sin saltarse ninguna, evitando asumir que ya ha sido respondida por comentarios previos; repetir las preguntas cuando pueda haberse ocasionado algún malentendido o alguna mala interpretación; moldear al entrevistado para que se ajuste a la estructura y naturaleza del cuestionario; y tener mucho cuidado con las indicaciones o comentarios informales que se ofrecen³². Los procesos de entrenamiento, entonces, pueden ser muy exigentes y se recomienda una extensión de al menos un día en un lugar aislado y confortable³³. Ello, así las cosas, quedaba muy lejos de nuestras posibilidades.

²⁹ DEMING, [1960]: 131.

³⁰ FOWLER, 2002: 124-126; LAVRAKAS, 1993: 130-143; TARNAI y MOORE, 2008: 378-382 y 384; y en especial sobre el impacto en la calidad de los datos, LESSLER et al., 2008: 446-452.

³¹ BILLIET y LOOSVELDT, 1988: 190-209; FOWLER y MANGIONE, 1990: 106-117 y 139-140.

³² GROVES y COUPER, 1998: 19-20, 45, 51, 188-189, 195-196, 209-211, 233, 265-267 y 306-307; GROVES et al., 2004: 294-295; LESSLER et al., 2008: 443-446.

³³ FOWLER y MANGIONE, 1990: 115 sobre todo.

El entrenamiento específico para nuestra encuesta a que se pudo someter a los entrevistadores fue tan modesto que no merece dicho nombre. En una hora el primer día y con breves sesiones de seguimiento los siguientes, la preparación se concentró en tres cuestiones básicas: que se tratara de convencer a los individuos contactados para que participaran en la entrevista, facilitando un protocolo de actuación ante negativas incluso insistentes; que se abstuvieran de ofrecer *feedback* a los entrevistados; y que se leyeran todas las preguntas tal y como habían sido redactadas. En las reuniones se realizaron también experiencias prácticas por parte de los encuestadores.

Siguiendo también a la literatura y a la investigación empírica, durante el trabajo de campo se trató de vigilar y supervisar a los entrevistadores por parte de este investigador mediante la escucha e incluso la observación, tratando de corregir los problemas que se fueron detectando en tiempo real cuando era posible y en las sesiones diarias de *debriefing*³⁴. Nuestra experiencia coincide con Biemer cuando insiste en la importancia de un buen entrenamiento, así como en sus serias dificultades; y, sobre todo, cuando añade que los entrevistadores a menudo no tienen ni los conocimientos ni las ganas que se les suelen atribuir³⁵. Así las cosas, y aunque hasta ahora no ha sido posible realizar análisis más formales por constreñimientos temporales, la impresión general fue la siguiente.

1. Se detectó una gran heterogeneidad entre los entrevistadores. Se observó desde los preparados y altamente motivados hasta algún caso de actitudes prácticamente desafiantes hacia las instrucciones que recibían —en estos casos, se procuraba que abandonaran inmediatamente el proyecto—. Esta es una de las advertencias más importantes que deben hacerse, así como que sus consecuencias para la varianza y el error son potencialmente dramáticas.

2. En muchos casos, los entrevistadores no trataban de convencer a los contactados para que participaran ni, una vez comenzada la entrevista, retenerlos. Lo que se comprobó, de hecho, fue en ocasiones ¡más bien lo contrario! Algunos entrevistadores tendían a terminar las entrevistas cuando comprobaban que éstas se alargaban demasiado o los entrevistados no contestaban de modo fluido.

3. Tampoco fue inhabitual que se ofrecieran aclaraciones o *feedback* a los entrevistados, incluso cuando no lo solicitaban.

³⁴ BIEMER, 2001; FOWLER, 2002: 126-129; FOWLER y MANGIONE, 1990: 120-133 y 141; STEVE et al., 2008: 401-408.

³⁵ BIEMER, 2001.

4. Algunas de las preguntas o, mejor dicho, de las palabras no eran léidas al pie de la letra por algunos entrevistadores.

No es preciso añadir mucho más para que este escenario sugiera una notable variabilidad y potencialmente serias fuentes de error no muestral. Ello debe mantenerse en mente. Nótese en todo caso que me he centrado en mi exposición justamente en los puntos en los que más se insistió en las reuniones con los entrevistadores. Nuestra experiencia, por lo tanto, coincide con la literatura en la necesidad del entrenamiento y supervisión continuados de los entrevistadores, tratando de no perder el control de esta parte del trabajo de campo. En particular, es importante insistir en que el modo telefónico precisamente tiene entre sus principales ventajas, si no es ya la más importante, el control que permite del proceso³⁶.

El cuestionario incluía una parte específicamente destinada al estudio de la firmeza frente al delito. El mismo fue sometido a un proceso de pretest, comenzando por la solicitud de revisión por parte de compañeros y un estudio con una pequeña muestra de estudiantes muy heterogéneos entre sí³⁷. A los primeros se les solicitó, en su calidad de investigadores, que ofrecieran comentarios sobre las preguntas y su redacción. A los segundos se les administró una versión mejorada del test y luego se aplicaron diversas técnicas de naturaleza cualitativa para valorar distintos aspectos del instrumento. De este modo se elaboró una ulterior versión mejorada del cuestionario que fue sometida a la revisión por parte de especialistas de la empresa que se encargó del trabajo de campo. Ello produjo una nueva versión con la que se llevó a cabo un estudio piloto con una muestra de pequeño tamaño (N=50)³⁸. En propiedad, un estudio piloto se

³⁶ LAVRAKAS, 1993: 125; STEEH, 2008: 237; STEVE et al., 2008: 401.

³⁷ FOWLER, 1995: 115-129; el mismo, 2002: 108-115.

Es una crítica habitual que algunos estudios utilizan estudiantes universitarios e incluso de niveles de formación inferior para estudiar cuestiones generales de una sociedad, WRIGHT et al., 2004: 181 y 189. El perfil de los alumnos de una Universidad a distancia es marcadamente distinto de los de la presencial. En particular, son mucho más heterogéneos entre sí y muchas de sus características socio-demográficas guardan una cierta similitud con la población general. Por este motivo, presentan claras ventajas frente a las muestras habituales de estudiantes, sin que pueda tampoco pretenderse que conformen una muestra representativa, ni mucho menos. Pero para evaluaciones de cuestionarios, como en el presente caso, pueden resultar de una gran utilidad.

³⁸ La conducción de estudios piloto, en el sentido reflejado en el texto, es sin duda especialmente aconsejable. Diversos colegas desaconsejaron su realización debido a que, al sacrificar un pequeño número de entrevistas, se perdía el supuesto número mágico de «1000». Algunos en particular adujeron ventajas a la hora de la potencial publicación de los resultados. Desde mi punto de vista, las ventajas de la

caracteriza por ser una versión preliminar y a pequeña escala del estudio principal. Por este motivo, verbigracia, utilizamos por supuesto el mismo modo, (algunos de) los mismos encuestadores y un mismo proceso de selección muestral. Se escuchó una amplia muestra aleatoria de estas entrevistas telefónicas, con una vocación evaluadora del instrumento y su administración, por este investigador. La matriz de datos que produjo el pilotaje se analizó buscando ulteriores problemas de ítems particulares y, en especial, comprobando su correcto comportamiento en análisis estadísticos de diversa naturaleza y asegurando una variabilidad mínima. De este modo se llegó a la versión definitiva del cuestionario³⁹. El resultado de todo este proceso relativamente sistemático y exhaustivo fue la detección de carencias en las diversas versiones preliminares del cuestionario, así como puntos débiles que pudieron ser objeto de mejoras significativas. Se asume, por lo tanto, que la última versión es una adaptación mejorada de la inicial.

Merece la pena señalar que en todo el proceso se prestó una gran atención a que la carga que se imponía a los entrevistados fuera asumible⁴⁰, ya que existen pruebas de que la misma afecta a la calidad de la encuesta⁴¹. Así, las entrevistas duraban de media alrededor de 10 minutos y, salvo alguna excepción sobre la que habrá que volver y que no fue posible obviar dada la orientación más amplia de la encuesta, la complejidad de las preguntas y sus categorías de respuesta se redujo tanto como fue posible. También aquí fue de gran utilidad el estudio piloto.

Las preguntas de nuestro cuestionario aquí de interés son las número 1, 2, 5 y 7. La Tabla 1 ofrece el texto de las mismas y sus categorías originarias de contestación, así como la distribución de frecuencias y porcentajes de respuesta. El orden de las categorías de respuesta de cada pregunta cambiaba de modo aleatorio gracias a la asistencia por ordenador para cada entrevista con el fin de evitar potenciales sesgos. Esto último es importante ya que existen pruebas empíricas sobre un efecto del orden de las respuestas posibles en estudios de encuesta general⁴² y para la punitividad en particular⁴³.

estrategia seguida en este trabajo son inconmensurablemente superiores a éste y otros costes reales o presuntos.

³⁹ Sobre el proceso de pretest y estudio piloto, vid. CAMPANELLI, 2008: 176-198; FORSYTH et al., 2004: 525-546; HANSEN et al., 2004: 337-360.

⁴⁰ FOWLER, 2002: 115; LAVRAKAS, 1993: 6.

⁴¹ TUCKER y LEPKOWSKI, 2008: 14.

⁴² GROVES et al., 2004: 159.

⁴³ Así, KURY y OBERGFELL-FUCHS, 2008: 286-287.

Tabla 1. Texto, categorías originarias de contestación y distribución de frecuencias y porcentajes de respuesta de las preguntas 1, 2, 5 y 7

P1. Hablemos de seguridad ciudadana.

Las personas tienen distintas opiniones acerca de las penas o sanciones que deberían imponerse a los delincuentes. Tomemos por ejemplo el caso de un hombre de 21 años a quien se le encuentra culpable de robar en una casa por segunda vez. Esta vez se llevó un televisor a color. ¿Cuál de las siguientes penas considera usted que es la más apropiada para un caso así?

	Frecuencia	Porcentaje
1. Multa	125	13,1
2. Prisión	234	24,6
3. Trabajo comunitario	523	54,9
4. Que se le condene, pero que se suspenda la condena	18	1,9
5. Otra pena	37	3,9
8. No sabe	10	1
9. No contesta	6	0,6
<i>Total</i>	953	100

P2. Aunque en España no existe la pena de muerte, me gustaría plantearle una pregunta sobre la misma. Con independencia de lo que usted personalmente haría, quisiera que me dijera si está a favor o en contra de aplicar la pena de muerte a personas con delitos muy graves.

	Frecuencia	Porcentaje
1. En contra	589	61,8
2. A favor	335	35,2
8. No sabe	19	2
9. No contesta	10	1
<i>Total</i>	953	100

P5. Dígame, para estos casos de la ablación o mutilación genital femenina ¿cuál de las siguientes penas considera usted que es la más apropiada para el culpable o ejecutor en un caso así?*

* Preguntas previas introducían el concepto de mutilación genital femenina y, en particular, interrogaban sobre si era conocido o no por el entrevistado (P3). Un 4,2% contestó que «No, no le suena» a la pregunta de si había oído hablar de este tipo de mutilaciones, mientras que un 95,8% contestó que «Sí, con seguridad» o «Sí, he oído algo». Independientemente de su grado de conocimiento, los entrevistados eran informados, mediante una entradilla en el cuestionario, de la definición del concepto de modo sencillo.

(continúa)

(continuación)

	Frecuencia	Porcentaje
1. Multa	14	1,5
2. Prisión	659	69,2
3. Trabajo comunitario	61	6,4
4. Que se le condene, pero que se suspenda la condena	14	1,5
5. Otra pena	174	18,3
8. No sabe	17	1,8
9. No contesta	14	1,5
<i>Total</i>	<i>953</i>	<i>100</i>

- P7. También existe la posibilidad legal de juzgar y condenar a las personas que sean culpables de estos delitos, incluso aunque lo hayan realizado fuera de España. En el caso de que esta persona viniese a España, usted personalmente, ¿está de acuerdo o en desacuerdo con esta posibilidad de que sean juzgados en España?

	Frecuencia	Porcentaje
1. Muy en desacuerdo	109	11,4
2. Más bien en desacuerdo	104	10,9
3. Ni de acuerdo ni en desacuerdo	38	4
4. Más bien de acuerdo	221	23,2
5. Muy de acuerdo	441	45,9
8. No sabe	41	4,3
9. No contesta	3	0,3
<i>Total</i>	<i>953</i>	<i>100</i>

La primera de nuestras cuestiones que aparece en nuestro cuestionario, la P1, la hemos denominado ROBO (A) y es muy bien conocida por ser utilizada habitualmente en muchas investigaciones comparadas y nacionales⁴⁴. La misma cuenta con varias categorías de respuesta sustantiva, a saber: «Multa», «Prisión», «Trabajo comunitario», «Que se le condene, pero que se suspenda la condena» y «Otra pena»⁴⁵. Es claro que se trata de una variable medida a nivel categó-

⁴⁴ Ello no excluye que la misma haya sido objeto de atendibles críticas, vid. en este sentido WALKER et al., 1988: 181.

⁴⁵ En particular la traducción de algunas de las categorías de respuesta no coincide con otras propuestas en la Criminología española. Aquí se ha procurado en particular evitar un potencial sesgo de deseabilidad social. La desventaja de nuestra decisión es que complica las posibles comparaciones —incluso aunque éstas son ya de por

rico. El mismo esquema sigue nuestra pregunta P5, solo que ahora el delito de referencia es la ablación, circuncisión femenina o mutilación genital femenina. Ha sido denominada aquí MGF (C). La justificación que subyace a la selección de este comportamiento delictivo es que se hipotetiza que, respecto al mismo y debido al enorme rechazo social que le caracteriza, la ambivalencia y deseabilidad social deberían desempeñar un rol menor en las respuestas de los encuestados que en otros casos. Las categorías de respuesta son las mismas que en P1. Otra pregunta habitual en distintos estudios comparados es la referida a la pena de muerte. Aquí la hemos incluido también como P2 y bajo el nombre de PENA_CAPITAL (B), utilizando la redacción del Centro de Investigaciones Sociológicas por sus apreciables características psicométricas —aunque será difícil llevar a cabo comparaciones debido a las grandes diferencias entre nuestra práctica y la del mencionado instituto⁴⁶—. Es una pregunta dicotómica,

sí complicadas por diversas ulteriores razones— dentro de nuestro país. Sin embargo, varias de las propuestas de otros autores introducen alteraciones respecto de las redacciones originarias como la edad del protagonista del escenario, con lo cual las posibles comparaciones quedan ya descartadas de antemano. La traducción que proponemos en nuestro cuestionario, entonces, aspira a ser fiel a la redacción original y a minimizar potenciales sesgos de deseabilidad social. Sobre la traducción en instrumentos, vid. HARKNESS et al., 2004: 453-466 y 472-473.

⁴⁶ Como debería quedar claro, este trabajo no avala la investigación de la firmeza frente al delito con preguntas únicas, tampoco con la referida a la pena de muerte. Ello debería quedar claro, aunque tampoco es tan dogmático como para excluir estudios de este tipo, sobre todo cuando no se toman de modo aislado. En efecto, a menudo un investigador ha de utilizar lo que tiene a su disposición —sin mencionar que otro debate es o debería ser en qué debe invertirse el dinero público para la investigación—.

Habiendo dicho esto, es pertinente mencionar aquí las críticas que VARONA GÓMEZ, 2009 ha vertido sobre algunas formas particulares de medición de la punitividad a nivel individual. En su trabajo distingue «Malas fuentes» y «Otras fuentes», aunque no alcanzo a ver con claridad la diferencia entre ambas —aparte de que las primeras parecen ofrecer una imagen punitiva de los ciudadanos y las segundas no—. Lo importante a nuestros intereses, sin embargo, es que su crítica alcanza también a preguntas utilizadas por el Centro de Investigaciones Sociológicas y, así, mantiene que «encuestas como las realizadas por el CIS han de interpretarse con mucho cuidado, pues más que medir los “hechos objetivos” [...] lo que fundamentalmente documentan [...] es la atención mediática que reciben determinados temas», 14. A mi modo de ver, ello tiene hondas consecuencias epistemológicas en las que no podemos detenernos aquí. Una más llana, aunque también más práctica, es que con este punto de vista podrían ponerse en duda casi cualesquiera encuestas y casi cualesquiera preguntas del CIS y otros institutos de investigación, cuando es el caso que muchos investigadores recurrimos en ocasiones a unas y otras; vid. 18-20. En particular, entre las «Malas fuentes», VARONA GÓMEZ incluye las que denomina «preguntas-trampa», que define como «cuestiones formuladas de una manera demasiado simple como para captar adecuadamente dichas actitudes punitivas», 10. Quizá se podría considerar la pregunta sobre la pena de muerte en la redacción aquí seguida una «pregunta-trampa».

ordinal desde nuestro punto de vista teórico. La cuestión P7 de nuestro cuestionario es quizá la más novedosa respecto a la investigación comparada. Interroga sobre el grado de acuerdo (cinco categorías de respuesta) con la posibilidad que ofrece nuestra legislación de juzgar a personas acusadas de mutilación genital femenina cuando hayan realizado el hecho fuera de España y se encuentren a disposición de las autoridades españolas. Hemos reservado para ella el término JUZGAR (D). Esta pregunta tiene la particularidad de referirse de modo inmediato a una cuestión de competencia y de ámbito de aplicación de las leyes penales. Desde este punto de vista se refiere a una cuestión de naturaleza eminentemente jurídico-procesal —aunque también de modo indirecto debe referirse al castigo, en cuanto que de un tal enjuiciamiento se puede derivar, lógicamente, una condena y una sanción—. Aunque intuitivamente es plausible que la firmeza frente al delito tenga un carácter marcadamente heterogéneo, esto es pluridimensional; pienso que no existen pruebas empíricas definitivas de ello⁴⁷. Como digo, parece también intuitivamente plausible que la firmeza frente al castigo debe guardar alguna relación más o menos íntima con aspectos jurídico-procesales, en particular procesal-penales. Esta relación puede hipotetizarse *indirecta* en el bien entendido que normas procesales tales como los requisitos de la actuación policial, los medios de prueba admitidos o el sistema de recursos deben influir, en igualdad de condiciones, en el castigo. La inclusión de JUZGAR (D) en nuestro cuestionario, así las cosas, permite explorar, aunque limitadamente, esta cuestión. Sin embargo, aparte de la escucha de las entrevistas⁴⁸, una somera lectura deja claro que se trata de un ítem de relativamente difícil comprensión para un entrevistado medio. Adviértase en todo caso que el marco general de la encuesta en la que se basaba nuestro estudio no nos dejaba muchas alternativas.

Para nuestro estudio se han eliminado las categorías de respuesta que carecían de un contenido sustantivo de interés para nuestra misión, las cuales coincidían en los cuatro casos: «No sabe» y «No con-

A mi juicio, no queda claro ni es sencillo decidir cuándo una pregunta ha sido formulada de modo *demasiado simple* —como tampoco lo es saber cuándo un encuestado tiene suficiente información o ha deliberado lo suficiente—; a la vez que es también difícil de decidir si un determinado porcentaje de contestaciones favorables a la pena de muerte es elevado o bajo desde el punto de vista de la punitividad si es tomado en términos absolutos, sin comparaciones con otros datos semejantes. Así las cosas, a mi modo de ver nuestra pregunta sobre la pena de muerte es muy problemática —de hecho es en parte lo que deseo mostrar empíricamente aquí—, pero no es una trampa.

⁴⁷ SPROTT, 1999a: 10.

⁴⁸ Sobre la detección de preguntas difíciles, vid. CONRAD et al., 2008: 212-230.

testa». Las mismas, pues, constituyen los datos perdidos para cada una de las cuatro cuestiones utilizadas en los presentes análisis. En particular, los datos perdidos para cada uno de los ítems utilizados aquí es el siguiente: A=17 (1,8%); B=27 (2,8%); C=31 (3,2%); D=42 (4,2%). Cuando se combinan en una tabla de contingencia para construir los datos para el análisis de clases latentes, resulta que el total de casos perdidos asciende a 102, esto es un 10,68% del total. Este dato es relativamente asumible. Sin embargo, puesto que incluso porcentajes pequeños de pérdida de datos pueden tener consecuencias serias para los hallazgos de un estudio, más abajo se tratará de minimizar esta pérdida.

P1⁴⁹ y P7⁵⁰ se han dicotomizado y P5 se ha reducido a tres categorías de respuesta⁵¹ para evitar el problema de *escasez* (sparseness) que se produce cuando, en determinados análisis estadísticos, algunas celdas no contienen ninguna observación, esto es N=0 —o bien cuando contienen números muy bajos⁵²—. Este es un problema bien conocido ya en el análisis de tablas de contingencia de dos variables. Por diversos motivos que incluyen desde tamaños muestrales pequeños e incluso moderados a *ceros estructurales* que se refieren a celdas que representan situaciones imposibles (como *hombres embarazados*) o al menos no observables (individuos que forman parte de una población teórica pero *no aparecen en un censo ni tienen teléfono*), algunas celdas pueden no incluir ningún caso⁵³. Esta dificultad

⁴⁹ Se han refundido las categorías de respuesta «Multa», «Trabajo comunitario», «Que se le condene, pero que se suspenda la condena» y «Otra pena» por un lado; mientras que «Prisión» se ha mantenido independiente. Los valores asignados han sido, respectivamente, 1 y 2.

⁵⁰ Se han refundido las categorías de respuesta «Muy en desacuerdo», «Más bien en desacuerdo» y «Ni de acuerdo ni en desacuerdo» por un lado; y «Más bien de acuerdo» y «Muy de acuerdo» por otro. Los valores asignados han sido, respectivamente, 1 y 2.

⁵¹ En primer lugar, se han refundido las categorías de respuesta «Multa», «Trabajo comunitario» y «Que se le condene, pero que se suspenda la condena»; mientras que «Prisión» y «Otra pena» se han mantenido independientes. El motivo de mantener separada ahora, en contraposición con P1, «Otra pena» no es fundamentalmente su relativamente alta frecuencia, sino que en una muestra de entrevistas escuchada por este investigador, quienes elegían esta categoría de respuesta tendían a añadir de modo espontáneo comentarios que evocaban castigos muy severos, no contemplados por nuestras leyes penales, por ejemplo del tipo «Que le hagan a él lo mismo» (*sic*). Debido a nuestras limitaciones, en ningún caso se contempló la posibilidad de incluir preguntas abiertas, verbigracia solicitando especificar a qué otra pena se estaban refiriendo. Los valores asignados han sido, respectivamente, 1, 2 y 3.

⁵² AGRESTI, 2002: 542; BIEMER, 2011: 188-190, 326-327 y 330; COLLINS y LANZA, 2010: 85-86; KAPLAN, 2009: 188.

⁵³ Acerca de la distinción, AGRESTI, 2002: 392.

no aparece, como es sabido, en análisis también habituales como los factoriales o los mínimos cuadrados ordinarios. Ello es debido a que en estos supuestos se asume que las correlaciones de segundo orden y superior son cero. En nuestro caso de análisis de clases latentes también existen correlaciones que no son bivariadas sino de órdenes superiores y por eso requieren mayores tamaños muestrales⁵⁴.

En efecto, este problema se presenta en el análisis de clases latentes. En realidad, en los mismos estas situaciones son habituales debido a que, al incluirse por regla más de dos variables, el número de celdas se dispara⁵⁵. En este sentido, el método de las clases latentes es *intensivo al nivel de los casos* —o sea que suele exigir tamaños muestrales elevados—. Bajo estas circunstancias y en particular para las celdas con cero observaciones, los tests de Chi^2 se vuelven problemáticos porque la verdadera distribución del estadístico no es aproximada por la distribución teórica del Chi^2 de modo correcto⁵⁶. Lo mismo cabe decir para L^2 . La literatura propone la razón N/k (siendo k el número de celdas) para valorar cuándo una tabla puede considerarse *escasa*⁵⁷. Para nuestro caso y utilizando todas las categorías de respuesta posibles de las preguntas de la Tabla 1, esta razón es muy baja. En efecto, puesto que $N=853$ y $k=5*2*5*5=250$, $N/k=3,412$. Ello deja clara la imposibilidad de utilizar nuestros items sin reducir sus categorías de respuesta. Con nuestra estrategia de reducción todavía persisten cinco celdas (20,8% del total de celdas) con menos de cinco observaciones, si bien ninguna con cero. Aunque la literatura pone reparos cuando se supera el 20% de celdas con menos de cinco observaciones⁵⁸ —lo cual se da, como vemos, aunque muy ligeramente, en nuestro estudio—, no deberíamos esperar serias complicaciones aquí. Para comenzar, la razón N/k es ahora elevada (=35,54); en segundo lugar, todas las celdas problemáticas, salvo una, cuentan con al menos tres casos; y, en último lugar, todas las celdas contienen alguna observación. Esto último, que suena redundante, no debería pasarse por alto en procedimientos que siguen un planteamiento probabilístico. Aunque con imprecisión, una observación, una, permite una aproximación a la probabilidad, en nuestro caso, de que un sujeto dé una contestación determinada. Dicho con otras palabras, se puede valorar el tamaño de una muestra en la que sería esperable encontrar un determinado patrón de respuesta. No sé,

⁵⁴ Vid. sobre ello BASILEVSKY, 1994: 611-612.

⁵⁵ AGRESTI, 2002: 392.

⁵⁶ LANGEHEIME et al., 1996: 492-514, proponen métodos de remuestreo como el *bootstrap* para evaluar los parámetros en casos como éstos de tablas con *escasez*.

⁵⁷ BIEMER, 2011: 188 y 329.

⁵⁸ Así, AGRESTI, 2002: 396.

digamos, uno de cada 1.000, uno de cada 10.000... es esperable que conteste de un determinado modo. O sea que ello permite *aproximar una probabilidad*. Pero con cero observaciones sencillamente no es posible ninguna estimación, ni siquiera imprecisa. En una muestra dada no aparece dicho patrón de respuesta, pero... ¿cómo de grande debería ser la muestra como para poder esperar que alguien lo hiciera? No se puede saber.

2.3. Estrategia analítica

Como se ha adelantado, un instrumento que cuenta con una cierta tradición en ciencias sociales⁵⁹ y que es cada vez más utilizado es el *análisis de clases latentes*⁶⁰. A menudo la literatura apunta que en una de sus versiones, probablemente la más habitual a día de hoy, representa un equivalente de los análisis factoriales en el caso de variables manifiestas y latentes categóricas⁶¹. Dicho con otros términos más técnicos, uno de los usos más habituales es la identificación de variables latentes que pueden ser medidas a partir de una serie de variables observadas. No importa repetir que en estos análisis tanto las variables latentes como las manifiestas están medidas a nivel categórico —u ordinal—. Esto representa un *modelo de medición*⁶². Ésta será la orientación que inicialmente seguiremos aquí, en consonancia con nuestras hipótesis.

En nuestra disciplina se pueden encontrar algunos interesantes ejemplos de análisis de clases latentes como los aquí propuestos o semejantes. Así, BRITT utilizó una estrategia múltiple que incluía análisis de clases latentes para defender la versatilidad de los delincuentes y, por lo tanto, las dificultades para establecer tipologías. Encontró que el mejor modelo era el de dos clases: delincuentes y no delin-

⁵⁹ Acerca de su evolución histórica, vid. COLLINS y LANZA, 2010: 7-8; GOODMAN, 2002: 5 y 27-31; SKRONDAL y RABE-HESKETH, 2007: 714-715.

⁶⁰ COLLINS y LANZA, 2010: 3-9 y 23-47; McCUTCHEON, 1987: 5-27. Es importante, por lo tanto, insistir en la existencia de una familia de análisis de variables latentes, a la cual pertenecen, sin duda, los análisis de clases latentes, KLINE, 2005: 16-17.

⁶¹ Así, BASILEVSKY, 1994: 611-612; BIEMER, 2011: xiii y 117; COLLINS y LANZA, 2010: 6; DAYTON, 1998: 1; McCUTCHEON, 1987: 7; el mismo, 2002: 56; McCUTCHEON y HAGENNAARS, 1997: 266; VERMUNT, 2003: 24. También puede verse como un análogo a los análisis de conglomerados, UEBERSAX, 1993: 1. Merece la pena señalar que al menos una investigación ha encontrado diferencias según se utilizan preguntas de este tipo o continuo, KURY et al., 1996: 310 y 318.

⁶² BOLLEN, 1989: 16-20; GOODMAN, 2002: 21; SKRONDAL y RABE-HESKETH, 2007: 714. Acerca de los problemas y falta de flexibilidad de las tablas de contingencia de tres y más variables, vid. POWERS y XIE, 2000: 87 y 129-135.

cuentas⁶³. D'UNGER y sus colegas aplicaron el mismo esquema a las carreras criminales, aunque con notables particularidades analíticas, y encontraron entre cuatro y cinco clases latentes de criminales en las distintas muestras que tuvieron a su disposición⁶⁴. KRUTTSCHNITT y MACMILLAN emplearon análisis de clases latentes para clasificar a una muestra de mujeres que habían sido víctimas de violencia en relaciones personales de distinta naturaleza y a continuación utilizaron las clases como la variable dependiente en modelos de regresión logit multinomiales⁶⁵. Aunque, por supuesto, no son comportamientos delictivos, SACCO y sus compañeros identificaron tres clases de bebedores en una elevada muestra de sujetos de sesenta o más años de edad del *National Epidemiologic Survey on Alcohol and Related Conditions*⁶⁶; mientras que TIMBERLAKE encontró cinco clases de fumadores utilizando una también amplia muestra procedente del conocido estudio *Add Health*⁶⁷. En ambos casos se utilizaron análisis de clases latentes. VAUGHN y sus asociados observaron mediante esta herramienta un modelo de cinco clases de niños según su nivel de autocontrol, con una quinta categoría que nítidamente incluía a los más bajos en esta variable⁶⁸. Confío en que estos ejemplos sirvan como muestra. Nótese, en todo caso, que algunas de las aplicaciones precedentes utilizan datos longitudinales, con lo cual las diferencias metodológicas y analíticas con el presente estudio son significativas.

Una importante advertencia que habrá hecho el crítico lector es que nuestro análisis —que en lo fundamental trata de estimar una variable no observada— utiliza sólo cuatro ítems o preguntas —cuando en el ámbito de la medición—, por ejemplo bajo la influencia de la Psicología clásica, suelen utilizarse muchos más. Esto es importante porque llama la atención sobre cómo en el fondo de nuestro enfoque general de variables latentes figuran muchas importantes diferencias con las formas más tradicionales de medición⁶⁹. Una revisión de esta cuestión, ni que decir tiene, excede la misión de este artículo; pero una de estas consecuencias es que escalas de medición cortas pueden arrojar mediciones plausibles e incluso superiores a otras más largas⁷⁰. A mayor abundamiento, UEBERSAX alerta contra el abuso de análisis de clases latentes con muchas variables manifiestas, y

⁶³ BRITT, 1994: 180 y 186-189.

⁶⁴ D'UNGER et al., 1998: 1593-1625.

⁶⁵ KRUTTSCHNITT y MACMILLAN, 2006: 147-161.

⁶⁶ SACCO et al., 2009: 829-837.

⁶⁷ TIMBERLAKE, 2008: 709-715.

⁶⁸ VAUGHN et al., 2009: 18-21, 26 y 28 sobre todo.

⁶⁹ EMBRETSON y REISE, 2000: 3 y 13-39; REEVE y MÄSSE, 2004: 248, 256-258 y 271.

⁷⁰ EMBRETSON y REISE, 2000: 18-21; REEVE y MÄSSE, 2004: 256-257.

escribe que «Con más de dos o tres [...], los resultados [...] pueden ser difíciles de interpretar»⁷¹. Nuestra experiencia aquí, como veremos, le da la razón.

Aunque, de modo sorprendente, la investigación y la literatura suelen prestarles relativamente poca atención, los modelos de clases latentes llevan a cabo una serie de importantes asunciones —algunas de las cuales pueden considerarse, además, relativamente fuertes—. BIEMER, a quien seguiremos en este punto con fidelidad, ofrece una taxonomía de las mismas: el modelo exige que las observaciones hayan sido recogidas mediante *muestreo aleatorio simple*; los indicadores deben ser *localmente independientes*, esto es independientes una vez controlada la variable latente de que se trate; las probabilidades de respuesta deben ser *homogéneas*, lo cual quiere decir que las probabilidades sean iguales para cualesquiera dos unidades seleccionadas; y los indicadores *unívocos*, o bien que lo sean todos ellos de la misma variable latente⁷². Aunque ahora mismo me limito a su enumeración, habrá que volver con más detalle sobre la comprobación de las mismas más abajo.

Quizá no está de más recalcar que, de entre estas asunciones, una que puede considerarse nuclear es la de independencia local o condicional⁷³. Como señala con claridad McCUTCHEON, esta situación se da cuando las relaciones entre un conjunto de variables es cero dentro de las categorías de alguna otra variable X; en estos casos puede decirse que aquellas variables son localmente independientes respecto de la variable latente subyacente X que puede considerarse explicativa de las primeras. Y esto es decisivo porque este criterio «proporciona un método para determinar si las relaciones entre un conjunto de medidas observadas se deben a alguna variable explicatoria no medida»; cuando se cumple esta asunción, entonces, «decimos que la variable adicional “explica” las relaciones observadas»⁷⁴. Por ello, así las cosas, más abajo dedicaremos un esfuerzo especial al test de esta hipótesis.

Finalmente, debe mencionarse que los análisis de clases latentes han sido también acreedores de algunas importantes consideraciones críticas. El propio BIEMER recoge y discute las que siguen: no hay pruebas de que la variable latente estimada realmente sea la que

⁷¹ UEBERSAX, 1993: 2.

⁷² BIEMER, 2011: 76, 83, 116, 125-128, 196-203 y 325-331.

⁷³ BASILEVSKY, 1994: 609; COLLINS y LANZA, 2010: 44-47; HEINEN, 1996: 6-9; McCUTCHEON, 1987: 14-16; McCUTCHEON y MILLS, 1998: 84-85; UEBERSAX, 1993: 1-2.

⁷⁴ McCUTCHEON, 1987: 16.

subyace a los indicadores y es de interés; realiza asunciones bastante fuertes; en casos de escasez sus resultados son pobres; sus hallazgos son difíciles de replicar; o, finalmente, sus resultados son fácilmente mal interpretados⁷⁵.

Los cálculos han sido realizados utilizando el programa LEM⁷⁶. También se ha utilizado CONDEP⁷⁷ para testar la hipótesis de independencia local en nuestro modelo de dos clases latentes. Es muy importante subrayar que ambos son gratuitos y accesibles fácilmente por internet. Para cuestiones puntuales que se anunciarán en su momento se ha recurrido a Mplus⁷⁸.

3. RESULTADOS

3.1. Análisis ordinarios

Nuestra primera hipótesis sustantiva, como se recordará, es que *existe una variable latente categórica de firmeza frente al delito con indicadores también categóricos de actitudes ciudadanas sobre el castigo* (H_1). Procedemos a su test mediante un análisis de clases latentes y comenzamos seleccionando el modelo, en particular su número de clases. La orientación es aquí básicamente exploratoria. La siguiente Tabla 2 ofrece los estadísticos de bondad de ajuste para tres modelos

Tabla 2. Estadísticos de bondad de ajuste para tres modelos de clases latentes

Modelo	Chi ² [p]	L ² [p]	BIC (L ²)	AIC (L ²)	BIC _{aj}	GL
<i>Modelo de independencia</i> (M_0)	42,3586 [0,001]	42,2665 [0,0005]	-77,2112	8,2665	4371,302	18
<i>2 clases latentes</i> (M_1)	18,2038 [0,1096]	17,815 [0,1214]	-63,1701	-6,185	4369,415	12
<i>3 clases latentes</i> (M_2)	3,9583 [0,6823]	3,8686 [0,6945]	-36,624	-8,1314	4374,801	6

N=853.

⁷⁵ BIEMER, 2011: 324-328.

⁷⁶ VERMUNT, 1997a: *passim*; el mismo, 1997b: 4-6 y *passim*.

⁷⁷ UEBERSAX, 2009: *passim*.

⁷⁸ MUTHÉN y MUTHÉN, 2010: *passim*.

de clases latentes⁷⁹, los cuales permitirán valorar cada uno de los mismos y compararlos entre sí.

Como puede observarse, el modelo de mutua independencia o de homogeneidad completa M_0 —en el que nuestras variables observadas se asumen mutuamente independientes entre sí y, de este modo, se especifica la existencia de una única clase latente— no se ajusta bien a los datos y debe rechazarse (no sólo p para $\text{Chi}^2=0,001$; e índice de disimilitud $d=0,0731$; sino que también existen cuatro residuos típicos superiores a $|2|$)⁸⁰. Su bajo BIC es, por lo tanto, ignorable. Se testan a continuación dos modelos de clases latentes con dos (M_1) y tres (M_2) clases respectivamente. Ambos, pues, incluyen una única variable latente con varias categorías o clases latentes. Ninguno de los dos incluye restricción alguna y pueden considerarse exploratorios⁸¹. Ambos modelos se ajustan de manera más que aceptable a los datos según estadísticos bien conocidos y ampliamente utilizados en análisis de esta naturaleza⁸². De entre los dos, M_1 es superior según BIC, pero no según AIC —valores (negativos) más bajos son indicadores de ajustes mejores—. Aunque pueden encontrarse divergencias en la literatura, BIC parece ser preferido para análisis como el nuestro⁸³. También BIC ajustado por el tamaño de la muestra⁸⁴ —que, en comparación con sus hermanos, no está basado en L^2 sino en la log-probabilidad, de ahí las diferencias en sus magnitudes— apoya el modelo con dos clases latentes, aunque su uso es mucho menos común. Estos estadísticos, como es sabido, permiten comparaciones incluso entre modelos no anidados —siempre que los datos utilizados sean idénticos⁸⁵—. En este caso, y no es inhabitual que no coincidan a la hora de

⁷⁹ Como deber ser evidente y de hecho se repetirá *infra* en el texto, nuestros modelos contienen una única variable latente, ANDERSON y VERMUNT, 2000: 85-89.

⁸⁰ Sobre los modelos de independencia, vid. en particular AGRESTI, 2002: 542 y 544.

⁸¹ McCUTCHEON, 2002: 58; VERMUNT, 1997b: 32-33. VAN DER HEIJDEN et al., 2002: 127, sugieren con razón que una buena estrategia es, precisamente, decidir primero qué número de clases latentes se precisa para describir correctamente los datos. A falta de hipótesis con base teórica, esta estrategia será la seguida en este trabajo.

⁸² Acerca de la valoración de los modelos y, en particular, de su ajuste, vid. AGRESTI, 2002: 216-217 y 257; BIEMER, 2011: 158-163, 169 y 190-191; COLLINS y LANZA, 2010: 82-88 y 109; DAYTON, 1998: 14-24; HEINEN, 1996: 39-43; LONG, 1997: 109-113; McCUTCHEON, 2002: 66-69 y 85 nota 7; POWERS y XIE, 2000: 65-71, 92, 99-101, 104-107 y 145-146; VERMUNT y MAGIDSON, 2002: 98. Una mirada más cautelosa sobre algunos de los estadísticos de bondad de ajuste aquí utilizados en RUDAS, 2002: 345-348.

⁸³ Vid. LI y NYHOLT, 2001: S272-S277.

⁸⁴ En particular: $n^* = (n + 2) / 24$.

⁸⁵ Aunque estos modelos pueden considerarse anidados, en estos casos no es posible utilizar la partición de L^2 . McCUTCHEON y MILLS, 1998: 89, añaden que BIC y AIC negativos son indicativos de un modelo superior al saturado.

seleccionar el modelo preferible, son inconcluyentes como diagnósticos⁸⁶. Tampoco el índice de disimilitud es definitivo puesto que en ambos casos $d < 0,05$, un criterio propuesto por la doctrina (d para $M_1 = 0,049$; d para $M_2 = 0,0198$) —a la par que, como vemos para los dos modelos, $d > 0,01$ —. Análisis con datos perdidos tampoco resuelven nuestro dilema. Por ejemplo, cuando se lleva a cabo un análisis de clases latentes que utiliza EM para imputar datos perdidos⁸⁷ —una estrategia alternativa a la que se seguirá más abajo— y, por lo tanto, se incluyen todas las observaciones ($N = 953$), de modo que puede esperarse una estimación algo más precisa, encontramos que AIC, BIC y BIC ajustado por el tamaño muestral siguen sin coincidir respecto a qué modelo es mejor.

Existen otros modos formales que pueden ayudar a decidir cuál de nuestros modelos es preferible, en concreto el test de la razón de verosimilitud de VUONG, LO, MENDELL y RUBIN y su versión ajustada⁸⁸. En ambos casos, el modelo favorecido es M_2 , esto es el de tres clases latentes. Lo que hacen estos procedimientos estadísticos es testar que un modelo con una clase menos es superior al de referencia, que cuenta con una clase más. A continuación, la Tabla 3 ofrece los resultados de estos tests.

Tabla 3. Tests de razón de verosimilitud para modelo con k-1 clases

Test	Valor [p]
<i>Razón de verosimilitud de Vuong-Lo-Mendell-Rubin</i>	-2164,294 [0,0367]
<i>Razón de verosimilitud ajustada de Lo-Mendell-Rubin</i>	14,323 [0,0393]

N=853.

⁸⁶ Vid. aquí en particular BIEMER, 2011: 162; COLLINS y LANZA, 2010: 88.

⁸⁷ Mplus permite esta opción.

⁸⁸ Esta posibilidad es ofrecida por Mplus, MUTHÉN y MUTHÉN, 2010: 657-658.

Cuando hay más de dos clases, Mplus sugiere de modo rutinario un aumento del número de comienzos aleatorios. Por defecto, éstos están fijados en 10; siguiendo la recomendación, aquí los hemos elevado a 100. Mplus también ofrece el *test de la razón de verosimilitud con bootstrap*, MUTHÉN y MUTHÉN, 2010: 658-659. No se informa aquí del mismo debido a que aparecen problemas porque la mejor verosimilitud no pudo ser replicada en casi la mitad de las sacas de bootstrap, con la consecuencia de que el valor p deja de ser confiable. Este problema no ha podido ser corregido por este

Como puede observarse, los dos tests rechazan la hipótesis nula de que un modelo con una clase menos al de tres es suficiente para describir nuestros datos ($p < 0,05$)⁸⁹. De este modo, como se dijo, favorecen M_2 . Es importante tener esto en cuenta. Sin embargo, la selección de un modelo no es un test de hipótesis. En palabras de LI y NYHOLT en un contexto estadístico semejante al nuestro, «No extrae conclusiones sobre si un modelo nulo es incorrecto [...] Para enfermedades humanas complejas [igual que para el estudio de actitudes también complejas], uno no debería pensar que los métodos estadísticos por sí solos proporcionarán la respuesta final»⁹⁰. Así, pues, seguimos obligados a una decisión.

A mi modo de ver, M_1 tiene a su favor ser un modelo más parsimonioso tanto desde el punto de vista *empírico* de que estima menos parámetros y por ello cuenta con más grados de libertad como del *lógico* puesto que clasifica las observaciones recurriendo a un número menor de clases. También puede mencionarse que la interpretación de M_1 es sencilla —identifica dos clases latentes (ordinales), una de *punitivos* o *firmes frente al delito* y otra de *no punitivos* o *benévolos*—, mientras que M_2 parece separar la clase de los punitivos en dos cuyas diferencias son inconsistentes y ambiguas excepto que en $X=3$ existe una probabilidad relativamente elevada ($=0,6804$) de elegir «Otra pena» en MGF (C); mientras que en $X=2$ la alta probabilidad se predica de la categoría de respuesta «Prisión» ($=0,9577$) ($X=1$ serían los «Benévolos») ⁹¹. Puede decirse que M_1 capitaliza las respuestas a MGF (C), pero existe ambigüedad cuando se toma el conjunto y se mira también a A, B y D —o bien podría decirse que no existen diferencias dignas de mención para dichos items entre $X=2$ y $X=3$ —. Bajo estas circunstancias, pues, la interpretación de M_2 parece más problemática que la de M_1 . Un último apunte es que existen algunos pequeños indicios de *sobreajuste* (overfitting)⁹² en el caso de M_2 , aunque las pruebas son, en el mejor de los casos, inconcluyentes. Y otro más es que ambos modelos podrían mejorar su ajuste introduciendo constreñimientos *ad hoc*, esto es *a posteriori* una vez observados los resultados. Aunque no es ni mucho menos una estrategia ilícita, aquí no estamos interesados en la construcción de modelos⁹³ ni tampoco en la especificación de modelos sin una base teórica.

investigador pese a que se ha incrementado notablemente el número de comienzos aleatorios.

⁸⁹ Por supuesto, si fijamos $\alpha=0,01$, el diagnóstico se invierte.

⁹⁰ LI y NYHOLT, 2001: S276.

⁹¹ Asumo que en este modelo existe *intercambio de clases latentes*.

⁹² BIEMER, 2011: 160.

⁹³ Seguramente debo aclarar que entiendo por este término el proceso de *introducción de mejoras sucesivas* en los modelos buscando mejores ajustes o menores

Modelos con más clases latentes no pueden estimarse sin restricciones debido a que no se encuentran identificados, y en el momento actual de la discusión sobre nuestro constructo *firmeza frente al delito* no existen criterios teóricos claros para imponer dichas restricciones⁹⁴. Así, el modelo con cuatro clases latentes no se encuentra bien identificado y tiene cero grados de libertad. En todo caso, tampoco existe justificación para este intento al haber hallado modelos más parsimoniosos con un ajuste aceptable⁹⁵.

Como acaba de decirse y puede comprobarse, el modelo con dos clases latentes M_1 —el que hemos seleccionado como el mejor— se ajusta aceptablemente bien a los datos (p para $\chi^2=0,1096$; p para $L^2=0,1214$; $d=0,049$; residuos típicos $\leq |1,671|$). El modelo, como ya sabemos, es parsimonioso⁹⁶ y tiene una muy sencilla interpretación⁹⁷ y sentido desde un punto de vista teórico. Nada sugiere que no haya habido convergencia del algoritmo⁹⁸. Igual que sus hermanos, no existen indicios de que M_1 no se encuentre identificado (por ejemplo, autovalores $\geq 1,5338$)⁹⁹. No parece haber problemas tampoco con *soluciones* u *óptimos locales*¹⁰⁰. De hecho, se han utilizado distintos valores de inicio ($J=10$), siempre con los mismos resultados, lo cual sugiere ausencia de problemas evidentes para diversas cuestiones¹⁰¹. No se han encontrado ahora indicios de *sobreajuste*¹⁰² ni de *intercambio de clases latentes* (flippage)¹⁰³.

errores típicos. Formas típicas de este procedimiento son la introducción de interacciones, transformaciones de las variables, etc. pero habitualmente de un modo *ad hoc*, no guiado por la teoría sustantiva —aunque también a posteriori se pueden ofrecer justificaciones más o menos plausibles—.

⁹⁴ ANDERSON y VERMUNT, 2000: 94-95; BÖCKENHOLT, 2002: 165.

⁹⁵ COLLINS y LANZA, 2010: 99.

⁹⁶ COLLINS y LANZA, 2010: 82.

⁹⁷ La interpretación es, en efecto, un importante elemento evaluador, McCUTCHEON, 1987: 13-14; McCUTCHEON y MILLS, 1998: 88 y 90. Vid., por ejemplo, KRUTTSCHNITT y MACMILLAN, 2006: 152-153 para unas clases a mi modesto entender difíciles de interpretar.

⁹⁸ BIEMER, 2011: 89.

⁹⁹ BIEMER, 2011: 183-185 y 330; COLLINS y LANZA, 2010: 89-92; HAGENAARS, 1990: 112; McCUTCHEON, 1987: 25-26; el mismo, 2002: 66; RABE-HESKETH y SKRONDAL, 2001: 1256-1259.

¹⁰⁰ BIEMER, 2011: 89, 91-92, 192-194 y 330; McCUTCHEON, 2002: 65; VERMUNT y MAGIDSON, 2002: 97. Con pocas clases, como es nuestro caso, los óptimos locales son mejores que cuando aquéllas aumentan, BÖCKENHOLT, 2002: 165.

¹⁰¹ AGRESTI, 2002: 541; COLLINS y LANZA, 2010: 92.

¹⁰² BIEMER, 2011: 160.

¹⁰³ BIEMER, 2011: 194-195.

Nuestros indicadores no pueden considerarse medidos todos ellos a nivel ordinal, lo cual podría sugerir estrategias analíticas alternativas, AGRESTI, 2010: 282-288; BIEMER, 2011: 231-235; POWERS y XIE, 2000: 119.

Así las cosas, nuestros resultados sugieren, en efecto, que existe en nuestros datos una variable latente categórica de firmeza frente al delito en el sentido de H_1 . Pero adviértase desde ya que el test de nuestra segunda hipótesis de que existe una variable latente categórica *unidimensional* de firmeza frente al delito con indicadores también categóricos de actitudes ciudadanas *directas e indirectas* sobre el castigo (H_2) exige una mayor profundización. *Indirectas*, como sabemos, se refiere aquí a actitudes sobre objetos de naturaleza procesal-penal. Volveremos sobre ella.

Los miembros de nuestras dos clases latentes pueden denominarse, respectivamente, «Benévolos» ($X=1$) y «Firmes» ($X=2$) frente al delito. Esto no debe interpretarse necesariamente como que existen en la población dos grupos discretos, cualitativamente distintos, sino que es perfectamente posible que sean la consecuencia de un corte en una variable que se distribuye de modo continuo —más habrá que decir sobre esta cuestión—. Para el caso de A y B, los sujetos que son clasificados entre los «Benévolos» tienen una probabilidad relativamente elevada de ofrecer respuestas favorables a las sanciones más leves que se les plantean (que coinciden con las categorías de respuesta numeradas como «1»); mientras que los «Firmes» tienen una probabilidad significativamente más elevada de ofrecer una respuesta alternativa («2») (ver Tabla 5). Aún así, los «Firmes» es más probable que se muestren favorables a sanciones que excluyan la privación de libertad para A y, desde luego, se opongan a la pena de muerte en el caso de B. Para C, lo más probable en ambas clases es mostrarse favorable a la pena de «Prisión»; y ambas tienen una probabilidad prácticamente idéntica —y además relativamente baja— de seleccionar «Otra pena». La diferencia sustancial reside en que los «Firmes» es algo más probable que se muestren favorables a la privación de libertad cuando se les pregunta por la mutilación genital femenina. Las probabilidades de nuestro análisis son semejantes para el caso de D: ambas clases tienen una mayor tendencia a posicionarse a favor del enjuiciamiento de los acusados de mutilación genital femenina, con una probabilidad ligeramente superior, como parece lógico, para los «Firmes». Como puede observarse desde ahora mismo, las tasas de error en que se incurre en nuestro análisis son muy elevadas, y eso complica enormemente la interpretación. La clase de los «Firmes» es mayor en tamaño en nuestro estudio ya que existe una probabilidad más alta de pertenecer a la misma ($=0,6768$) en comparación con los «Benévolos» ($=1-0,6768=0,3232$). Estos datos pueden transformarse en porcentajes con facilidad, y para nuestro estudio la frecuencia esperada de cada clase es, respectivamente, 577,31 y 275,69 —no se olvide la naturaleza probabilística del enfoque—.

La Tabla 4 con las frecuencias observadas y estimadas para cada patrón de respuesta y la ya mencionada Tabla 5 con los resultados de clase latente ofrecen información valiosa para la comprensión del modelo e incluso para la valoración del error.

Así, las tasas de error para los ítems A, B, C y D son 0,0505, 0,1761, 0,5749 y 0,697 (falsos negativos), respectivamente, para que-

Tabla 4. Frecuencias

Patrón: ABCD	Frecuencia observada	Frecuencia estimada	Residuos típicos
1111	15	17,924	-0,691
1112	46	42,235	0,579
1121	75	77,325	-0,264
1122	217	209,871	0,492
1131	23	20,326	0,593
1132	46	53,935	-1,08
1211	7	4,713	1,053
1212	8	11,777	-1,101
1221	34	40,634	-1,041
1222	114	120,022	-0,55
1231	15	9,776	1,671
1232	37	28,463	1,6
2111	3	1,6561	0,044
2112	3	4,438	-0,683
2121	23	23,349	-0,072
2122	78	71,144	0,813
2131	4	5,415	-0,608
2132	11	16,382	-1,33
2211	1	,829	0,188
2212	3	2,429	0,367
2221	18	17,972	0,007
2222	57	55,683	0,176
2231	6	4,082	0,949
2232	9	12,621	-1,019
<i>Total</i>	853	853,001	

Tabla 5. Resultados de clase latente para M_1

		Clase 1 (No punitivo)	Clase 2 (Punitivo)
		0,3232	0,6768
A [ROBAR]	1	0,9495	0,65
	2	0,0505	0,35
B [PENA_CAPITAL]	1	0,8239	0,5489
	2	0,1761	0,4511
C [MGF]	1	0,2513	0,029
	2	0,5749	0,7925
	3	0,1738	0,1785
D [JUZGAR]	1	0,304	0,2428
	2	0,697	0,7572

$E=0,2558$

$\lambda=0,2085$

$OCC_j < 5$

nes pertenecen a la clase latente «Benévolos»; mientras que la tasa de error es 0,65, 0,5489, 0,029 y 0,2428, respectivamente, para los «Firmes» (falsos positivos). Partiendo de la base de que lo natural es pronosticar a los «Firmes», las primeras probabilidades, en efecto, representan *falsos negativos* y las segundas *falsos positivos*. La impresión general tanto para $X=1$ como para $X=2$ es muy pobre. En términos generales, además, las tasas de error *no son asumibles* si tomáramos los items de modo aislado.

Por lo que se refiere a las clases en concreto, $X=1$ es relativamente homogénea, desde luego en comparación con $X=2$, que no brilla precisamente por esta característica. Esto complica una interpretación clara y dificulta la ya de por sí difícil (potencial) labor de clasificación. En parte por lo anterior, la separación entre ambas clases latentes es modesta¹⁰⁴. La impresión global poco optimista —para la evaluación— de la clasificación con los patrones e items individuales en mente es confirmada por los estadísticos globales de *errores de clasificación* ($E=0,2558$) y de *reducción en la proporción de errores* ($\lambda=0,2085$)¹⁰⁵, de los que también informa la Tabla 5. Un ulterior diagnóstico, que resulta redundante, es el propuesto nada más y nada menos que por Nagin bajo la denominación de *ventajas de cla-*

¹⁰⁴ COLLINS y LANZA, 2010: 56-67 y 75-76, 56-57 en particular.

¹⁰⁵ Sobre su cálculo e interpretación, vid. McCUTCHEON, 1987: 36-37.

sificación correcta, que son también aplicables a casos como el nuestro¹⁰⁶. El mismo añade más sensaciones pesimistas puesto que no alcanza el 5 para ambos grupos, el punto de corte que sugiere este autor.

Puesto que aquí estamos especialmente interesados en ellos, insistamos en nuestros ítems particulares. Aunque nuestro modelo de dos clases M_1 ha sido elegido como el mejor de los ensayados y sugiere que una variable latente única subyace a nuestros indicadores, no está nada claro que JUZGAR (D) forme parte de la misma (en el marco de nuestro modelo). Una mirada a las probabilidades condicionales de la Tabla 5 deja claro que *la respuesta que se dé por un entrevistado a dicha pregunta prácticamente no depende de la clase latente a la que pertenece*. Así, las probabilidades condicionales para elegir la categoría de respuesta más benévola (=1) son 0,3040 para los «Benévolos» y 0,2428 para los «Firmes». Ello puede incluso sugerir independencia¹⁰⁷, aunque no se han encontrado ulteriores pruebas claras de ello. Los pseudocoefficientes de determinación para $P(D|X)$ en M_1 son tan insignificantes que la interpretación va también en esta línea, incluso aunque esté claro que estos estadísticos deben contemplarse con cautela¹⁰⁸. Por ejemplo, $\text{verosimilitud}^{(-2/N)} = 0,0042/0,0061$. Como se adelantó, ello pone muy en duda H_2 . Ahora bien, es menester no perder de vista la especial complejidad de esta pregunta D —remito aquí al crítico lector *supra* al texto en la Tabla 1—, con lo que parece plausible que al menos algunos entrevistados hayan malinterpretado la misma. Aquí se requeriría investigación cualitativa, la cual quedaba fuera de nuestras posibilidades. Sin la misma, lo más cauto es asumir que nos encontramos ante una cuestión abierta.

Debido a su novedad, un ítem importante en el presente estudio es el relativo a la pena que se impondría en un caso de circuncisión o

¹⁰⁶ NAGIN, 2005: 88-89.

¹⁰⁷ Ya me he decantado para este trabajo en contra de la construcción de modelos y a favor de la orientación teórica inicial. Sin embargo, el lector crítico puede preguntarse legítimamente por el modelaje de A, B y C, dejando fuera D. El modelo de independencia claramente está muy lejos de ajustarse bien a los datos y debe rechazarse sin ningún género de duda. El modelo de dos clases latentes se ajusta a los datos de modo marginal y su aceptación puede ser asumible. En particular, para el modelo de dos clases latentes M_{1bis} : $\text{Chi}^2=8,1113$ ($p=0,0173$); $L^2=8,1553$ ($p=0,0169$); $gl=2$; $d=0,028$; $\text{BIC}(L^2)=-5,4137$; $\text{AIC}(L^2)=4,1553$. Autovalores y residuos típicos no ofrecen ningún indicio de problema alguno. Modelos exploratorios con más clases no se encuentran identificados. Ahora bien, casi todo lo que digamos en el texto para el modelo con A, B, C y D (M_1) vale para el limitado a A, B y C (M_{1bis}) ya que, sin ir más lejos y como era de esperar, las probabilidades condicionales apenas varían de uno a otro (para las variables incluidas, se entiende).

¹⁰⁸ LONG, 1997: 102; POWERS y XIE, 2000: 71.

mutilación genital femenina. Las probabilidades condicionales sugieren que, como se avanzó, la categoría de respuesta «Otra pena» es problemática. En efecto, la probabilidad de que un sujeto de la clase «Firmes frente al delito» y de la clase «Benévolos» seleccione esta respuesta es prácticamente la misma (0,1738 vs. 0,1785). Ello puede ser consistente con que para una parte de los encuestados esta categoría de respuesta implique una sanción leve, para otra grave —en comparación, por ejemplo, con la privación de libertad— y para otra quizá más simplemente una no respuesta. Por nuestras limitaciones en la duración de las entrevistas no fue posible preguntar qué se entendía por «Otra pena» cuando un encuestado la seleccionaba. Sin embargo, este ítem MGF forma parte de modo más claro de nuestra variable latente, aunque los pseudocoefficientes de determinación son modestos —si bien ya no despreciables—: así, $\text{verosimilitud}^{(-2/N)} = 0,1081/0,1374$. Si obviamos la categoría de respuesta «Otra pena», los «Firmes frente al delito» es muy probable que opten por una pena privativa de libertad en MGF (C). Los «Benévolos», por su parte, también es probable que muestren la misma preferencia, lo cual no es sorprendente si se considera la conducta por la que se interroga, pero de modo significativamente menor que en el caso de sus compañeros «Firmes».

Para el caso de un análisis de clases latentes $M_{1\text{bis}}$ limitado a A, B y C y excluyendo, por lo tanto, D, el panorama que ofrecen las probabilidades condicionales es muy semejante. Quizá merezca la pena advertir que ahora la probabilidad de elegir «Otra pena» es un poco más elevada en el caso de los «Firmes» que en el de los «Benévolos» (0,1932 vs. 0,1363), respaldando quizá la intuición originaria. Por lo demás, las consideraciones del párrafo precedente son igualmente aplicables.

ROBAR (A) y PENA_CAPITAL (B) son dos de los ítems más habitualmente utilizados en la investigación de la firmeza frente al delito o punitividad. De nuevo, los pseudocoefficientes de determinación, siempre para M_1 , son modestos, pero es asumible que ambas preguntas forman parte de una variable categórica latente superior. Así, $\text{verosimilitud}^{(-2/N)}$ para $P(A|X) = 0,1184/0,1748$; y para $P(B|X) = 0,0739/0,1012$. A y B hacen un buen trabajo de discriminación para $X=1$. En ambos casos, la probabilidad de que los «Benévolos» ofrezcan una respuesta no punitiva es muy alta. En efecto, sus probabilidades condicionales ascienden para la categoría de respuesta «1», respectivamente, a 0,9495 y 0,8239. Ahí, pues, la discriminación es notable. Sin embargo, no puede decirse lo mismo para $X=2$, o sea para la clase latente de los «Firmes». Incluso éstos es más probable que prefieran una pena no privativa de libertad para el joven ladrón

(=0,65) y que se opongan a la pena de muerte (=0,5489). No está de más recordar aquí el mayor tamaño de esta segunda clase latente en comparación con la primera, lo cual puede contribuir a explicar esta mayor similitud en las probabilidades condicionales, o dicho de otro modo su menor capacidad discriminatoria dentro de la clase. Eso sí, la probabilidad de una respuesta punitiva en A y B es notoria y significativamente más elevada en $X=2$ que en $X=1$. Todo esto es igualmente válido para M_{1bis} , que excluye D.

A nivel individual, entonces, la valoración de los items particulares no invita al optimismo. En concreto no sólo debe decirse que la carga de error de medición de cada uno de ellos, aunque con matices, es notable; sino que, consecuentemente, también la relación de cada uno de ellos con la variable latente firmeza/benevolencia frente al delito, o punitividad si se prefiere, es débil. Aunque estas afirmaciones han de tomarse con cautela debido, entre otras cosas, a que no tenemos la certeza de que nuestra variable latente no sea acreedora a su vez de consideraciones de algún tipo, el escenario apunta a las claras a que *la utilización aislada de items como los aquí empleados es altamente problemática*. Podemos ir con algo más de detalle siguiendo las indicaciones de la literatura. En primer lugar, está claro que puede excluirse independencia al menos entre A, B y C y nuestra variable latente puesto que las probabilidades de respuesta a lo largo de las clases no son idénticas ni son iguales a las proporciones marginales. En segundo lugar, las probabilidades condicionales no se aproximan de modo regular a 0 y 1, lo cual es indicador de un bajo grado de certeza. Vistas en conjunto, estas consideraciones apuntan a que al menos la relación entre A y B —así como, sin duda, D— por un lado y la variable latente por otro es débil —aunque no importa repetir que, sin duda, no son independientes¹⁰⁹—.

Como se anunció y probablemente resulta evidente después de los párrafos precedentes, el enfoque de clases latentes abre enormes posibilidades, potencialmente revolucionarias, para la evaluación del *error de medición*¹¹⁰. Hasta ahora, sin embargo, este camino ha sido menos explorado por los analistas que el de la reducción de la información, la construcción de tipologías o el test de hipótesis. BIEMER se extiende en los detalles de esta utilización particular de los modelos de clases latentes¹¹¹. En concreto, escribe que se trata de un «método

¹⁰⁹ COLLINS y LANZA, 2010: 50-53 sobre todo, aunque también 53-55. C presenta particularidades propias que dificultan una valoración como la del texto.

¹¹⁰ BIEMER, 2004: 226-228 y 242-246 sobre todo; el mismo, 2011: 1-24 y 115-125; MCCUTCHEON y MILLS, 1998: 82-83.

¹¹¹ BIEMER, 2011: 1-24 sobre todo.

estadístico para la predicción de la verdadera clasificación de individuos de acuerdo con sus clasificaciones observadas. Adicionalmente, el LCA [análisis de clases latentes] proporcionará estimaciones de las probabilidades de los encuestados de ser mal clasificados por la pregunta», «Estas estimaciones pueden ser utilizadas para identificar ítems de cuestionario que sean problemáticos, y en última instancia para la mejora de la encuesta. Un objetivo secundario es la estimación de la prevalencia de características de una población corregidas por el error de clasificación»¹¹². Sin embargo y como se ha adelantado, el *enfoque de análisis de error* sólo puede ser seguido aquí de modo *aproximado*. El motivo es que nuestro análisis no asume que la variable latente con dos clases que hemos identificado en M_1 sea la característica o variable latente *verdadera* que se buscaba. Sólo en este caso, por supuesto, es posible establecer con seguridad las probabilidades de error de cada ítem particular, y ello exigiría conocer con precisión y *a priori* la variable latente X , y para que esta situación se dé, X debe poder ser medida directamente y sin error¹¹³. Por supuesto, esto no excluye que sea posible, para nuestra investigación en particular, una aproximación a las probabilidades de error en que incurrimos sobre todo porque no existe manera de medir directamente la *firmeza frente al delito*, en el bien entendido que se trata de una actitud latente, no observable directamente¹¹⁴. En todo caso, funciones mencionadas como la selección de ítems y la exclusión de los que resultan problemáticos son de mucho más sencilla implementación y pueden mejorar una investigación de modo significativo.

También a estas alturas debe haber quedado claro que, si como regla general, pese a que es posible utilizar este enfoque para la clasificación individual de sujetos, esto es su adscripción a una u otra clase, con la idea en mente, verbigracia, de utilizar esta información para ulteriores análisis, la clasificación que se hace es probabilística¹¹⁵ y no es aconsejable seguir este procedimiento por las tasas de error en que se incurre, en casos como el nuestro esta vía es simplemente impensable.

¹¹² BIEMER, 2011: 22-23.

¹¹³ BIEMER, 2011: 118, 126 y 324-325.

¹¹⁴ Este enfoque permite, a mayor abundamiento, una valoración de la fiabilidad en el caso de variables latentes dicotómicas —un terreno relativamente atrasado en comparación con las continuas—. Aunque puede señalarse que la misma es pobre, aquí no estamos interesados en este punto, sino en la evaluación de las mediciones más habituales de *firmeza frente al delito*, a partir de ítems únicos. Vid. sobre todo ello ALWIN, 1992: 61-67, 105 y 270; CLOGG y MANNING, 1996: 173-174.

¹¹⁵ GOODMAN, 2002: 22; VERMUNT y MAGIDSON, 2002: 91; los mismos, 2003: 531-537.

3.3. ¿Existen «ideólogos» en nuestra muestra?

Como se mencionó, algunos autores han sugerido que, en determinadas situaciones, algunos encuestados muestran un nivel de consistencia en sus respuestas muy elevado¹¹⁶. En particular, estos individuos tienen una *probabilidad perfecta* ($p=1$) de ofrecer una determinada respuesta. Puede pensarse en estar siempre «Muy de acuerdo» con una serie de afirmaciones en una escala o, para nuestro caso, con seleccionar siempre las respuestas más punitivas. Dicho de otro modo más intuitivo, me refiero a sujetos que siguen un patrón de respuesta $y_j=(2, 2, 2, 2)$ con $p=1$. Al parecer, esta idea procede nada más y nada menos que de Otis Duncan, quien denominó «ideólogos (ideologues)» a estas personas¹¹⁷. Se recordará que ésta era precisamente nuestra tercera hipótesis sustantiva de que *existe en la variable latente categórica* firmeza frente al delito *una clase latente de «ideólogos»* (H_3).

La investigación criminológica ha realizado alguna referencia, si no al concepto, sí a figuras cercanas. Por ejemplo, KUHN sugiere la existencia de un pequeño grupo de encuestados «particularmente punitivo» que estarían afectando los estadísticos descriptivos de medidas de firmeza frente al delito¹¹⁸. Quizá más próximo a nuestra hipótesis, Varona Gómez divide su muestra entre quienes habían contestado a todos los escenarios de su cuestionario de modo punitivo y el resto. Aunque este autor no sigue un modelo probabilístico ni de variables latentes y asume que su clase (observada) está compuesta por los más firmes, aquí se despierta la duda de si este grupo podría tener alguna particularidad de naturaleza cualitativa¹¹⁹.

Para testar esta hipótesis es preciso construir un modelo que incluya una tal clase latente de «ideólogos» —en realidad «ideólogos-firmes»—. La opción más natural, a la luz de nuestros hallazgos previos, es hipotetizar dos clases semejantes a las de M_1 más una tercera para los ideólogos (M_3)¹²⁰. A tal fin, es menester introducir restricciones en el modelo fijando una respuesta punitiva para cada uno de nuestros cuatro items para el caso de los «ideólogos», esto es cons-

¹¹⁶ McCUTCHEON y MILLS, 1998: 87.

¹¹⁷ Vid. McCUTCHEON y MILLS, 1998: 87.

¹¹⁸ KUHN, 2002: 123.

¹¹⁹ VARONA GÓMEZ, 2008: 24.

¹²⁰ La opción de un modelo de sólo dos clases latentes (M_3 bis), una para los «ideólogos» y otra para todos los demás parece intuitivamente poco plausible. De todos modos hemos testado esta opción con resultados muy poco halagüeños para la misma.

truir una clase latente con probabilidades condicionales de 1,0 para los cuatro ítems¹²¹. Para C se considera que la categoría de respuesta más punitiva es «Prisión» (=2). A continuación, la Tabla 6 ofrece los estadísticos de bondad de ajuste y los grados de libertad para el modelo de «ideólogos» (M_3) y para nuestro modelo de dos clases latentes sin restricciones, que hemos aceptado como el mejor.

Tabla 6. Estadísticos de bondad de ajuste para dos modelos de clases latentes (M_1 y M_3)

Modelo	Chi ² [p]	L ² [p]	BIC (L ²)	AIC (L ²)	GL
2 clases latentes (M_1)	18,2038 [0,1096]	17,815 [0,1214]	-63,1701	-6,185	12
3. ^a clase de «ideólogos» (M_3)	18,2038 [0,0516]	17,815 [0,0582]	-49,6726	-2,1850	10 [†]

N=853.

[†]Grados de libertad para M_3 correctos, devueltos mediante matriz de diseño en LEM¹²².

El modelo de tres clases latentes que incluye una tercera de «ideólogos» o «ideólogos-punitivos» (M_3) se ajusta razonablemente bien a los datos (p para Chi²=0,0516; p para L²=0,0582). Verbigracia, el *índice de disimilitud* (d=0,049), que se interpreta como la menor proporción de observaciones que habría que cambiar de unas celdas a otras para un ajuste perfecto, y que Biemer¹²³ considera indicativo de un buen ajuste cuando es inferior a 0,05 —o bien 0,01 en otras situaciones— invita al optimismo. Sin embargo, el modelo es claramente inferior a M_1 atendiendo a BIC y AIC. Además es obviamente menos parsimonioso. Finalmente pero igual de importante es que la probabilidad estimada de pertenencia a dicha tercera clase es, en M_3 , bajísima, esto es que su tamaño requeriría una población elevada como para contar con algún miembro. Por todo lo anterior debe rechazarse esta opción¹²⁴.

¹²¹ Sobre la imposición de restricciones en general, vid. COLLINS y LANZA, 2010: 79-80 y 102-106; McCUTCHEON, 1987: 37-44.

¹²² Vid. BIEMER, 2011: 168.

¹²³ BIEMER, 2011: 160.

¹²⁴ El ajuste es prácticamente idéntico y en todo caso inferior a M_1 cuando se ha interpretado que la respuesta más punitiva a C era «Otra pena» (=3) e incluso cuando se ha concedido como muy punitivo contestar, con la misma probabilidad, «Otra pena» o «Prisión». Ni que decir tiene que, en contraposición, también puede pensarse en una

Así las cosas, al menos para nuestro estudio, no existen pruebas de la existencia de una clase latente significativa de «ideólogos»¹²⁵. Con ello no se encuentra apoyo para nuestra hipótesis sustantiva H_3 .

3.4. Comprobación de las asunciones del modelo de clases latentes (M_1)

BIEMER, a quien como se dijo seguiremos en este punto, ofrece una excelente discusión de las asunciones del modelo de clases latentes —algo que suele pasar más bien desapercibido en el caso de un elevado número de analistas—. Para empezar, el modelo debe realizarse a partir de una muestra extraída de una amplia población mediante muestreo aleatorio simple sin reemplazo. Esta condición no se cumple en nuestro estudio. Aunque existen formas relativamente sofisticadas de tener en cuenta los aspectos complejos del muestreo¹²⁶, en la práctica de nuestra disciplina, al menos hasta ahora, rara vez se recurre a estos expedientes aunque son asunciones de un conjunto de técnicas habituales.

La asunción nuclear de nuestro enfoque es, como se dijo, la independencia condicional dada la pertenencia a una clase latente particular: los indicadores deben ser *localmente independientes*, o sea independientes una vez que se controla la variable latente X de que se trate¹²⁷. Cuando no se cumple la asunción de independencia local, tanto los estadísticos de bondad de ajuste como los errores típicos pueden resultar demasiado elevados, con lo que sus efectos a la hora de evaluar el modelo pueden ser críticos, aunque difíciles de predecir. Del mismo modo, las estimaciones de los parámetros pueden ser incorrectas. Lo más importante, sin embargo, es que el modelo es puesto en duda. Existen, eso sí, formas de modelado de la de-

clase latente de sujetos *benévolos* que responden de modo perfectamente consistente. Testado este modelo de «ideólogos-benévolos», el veredicto es el mismo que para los «ideólogos-firmes».

¹²⁵ Quizá se pueda acusar a este modelo de demasiado determinista. Es también posible, mediante la imposición de restricciones, construir un modelo en el que la tercera clase de «ideólogos-firmes» simplemente tenga una probabilidad relativamente alta e idéntica de contestar de modo punitivo a cada uno de nuestros cuatro ítems. Este modelo, que también puede tener varias versiones, es siempre inferior a M_1 .

¹²⁶ BIEMER, 2011: 209-230 y 330-331.

¹²⁷ BASILEVSKY, 1994: 609; COLLINS y LANZA, 2010: 44-47; HEINEN, 1996: 6-9; KAPLAN, 2009: 185-186; McCUTCHEON, 1987: 14-16; el mismo, 2002: 58; McCUTCHEON y MILLS, 1998: 84-85; SKRONDAL y RABE-HESKETH, 2004: 74; UEBERSAX, 1993: 1-2; VERMUNT, 2003: 23.

pendencia condicional¹²⁸. Existen varias formas, siguiendo siempre a la literatura, para detectar la dependencia condicional o local —o sea para testar nuestra asunción—. Un diagnóstico muy interesante es el propuesto por UEBERSAX, que utiliza una versión modificada del *test de la razón de las log-ventajas* (LORC) originariamente desarrollado por GARRET y ZEGER. Aunque para los detalles he de remitirme a los trabajos originales, en realidad LORC trata de estimar bien conocidos valores z para cada par de items o variables, para lo cual utiliza la diferencia entre las razones de las log-ventajas observadas y esperadas, divididas por el error típico de las segundas. Naturalmente, la comprobación de si los valores z exceden los valores críticos habituales —en nuestro caso $|1,96|$ (o, si se prefiere, $|1,645|$)— nos ofrece evidencia de que tales items son condicionalmente dependientes. Esto es, uno de los procedimientos más habituales en análisis estadísticos, por lo tanto bien conocido. El programa CONDEP permite realizar esta operación con una relativa sencillez y rapidez. La Tabla 7 nos ofrece los resultados de los tests de dependencia condicional entre nuestros pares de items, por supuesto para M_1 .

Como puede observarse, $z \leq |0,54|$, de modo que no existen en estos análisis pruebas de una violación de la asunción de referencia, cuyo carácter es, no importa repetirlo, nuclear en análisis de clases latentes.

BIEMER ofrece una manera de testar la hipótesis de *homogeneidad* —una asunción importante para la identificabilidad del modelo y la estimación de los parámetros— mediante la introducción de *variables*

Tabla 7. Diagnósticos para dependencia condicional (M_1)

Tests	G^2	Razón de las log-ventajas esperada	E.T.	Razón de las log-ventajas observada	Valor z
<i>i j</i>					
A B	13,23	-4,62	1,42	-4,68	-0,05
A C	-3,59	0,98	0,436	1,07	0,21
A D	13,27	0,1	0,247	0,12	0,07
B C	-2,67	0,53	0,291	0,52	-0,02
B D	13,05	0,07	0,185	0,01	-0,31
C D	-3,34	0,17	0,26	0,31	0,54

¹²⁸ Vid., por ejemplo, UEBERSAX, 1999: 283-295.

de grupo o de agrupación. Una opción habitual, a la par que intuitivamente natural y que seguiremos aquí, es elegir el sexo (S) como variable de grupo. Nuestro autor sugiere aquí especificar el modelo {XS}... {XD}, dejando fuera XSA... XSD. Esta exclusión implica homogeneidad de grupo para cada una de nuestras probabilidades de error. Este modelo M_{hom} se ajusta de modo asumible a los datos ($\text{Chi}^2=44,7762$ [$p=0,1023$]; $L^2=45,9947$ [$p=0,0822$]; $gl=34$; $d=0,0771$; $\text{BIC}[L^2]=-183,4631$; $\text{AIC}[L^2]=-22,0053$)¹²⁹. De este modo, nuestros hallazgos apuntan a que XA... XD no dependen de S, o lo que es lo mismo que existe homogeneidad, como exige nuestro modelo de clases latentes¹³⁰.

La *univocalidad* es otra importante asunción de los modelos de clases latentes. La misma implica que todos los indicadores lo son de la misma variable latente. De nuevo BIEMER ofrece una estrategia para testar esta hipótesis añadiendo una segunda variable latente (Y). A y B se modelan como indicadores de X y, por otro lado, C y D de Y¹³¹. El modelo M_{mult} así estimado se ajusta bastante bien a los datos ($\text{Chi}^2=17,1708$ [$p=0,0707$]; $L^2=16,7225$ [$p=0,0807$]; $gl=10$; $d=0,0463$; $\text{BIC}[L^2]=-50,7651$; $\text{AIC}[L^2]=-3,2775$). Sin embargo, es claramente inferior a M_1 desde diversas ópticas. Así las cosas, tampoco hemos encontrado pruebas de una potencial infracción de la asunción de univocalidad¹³².

3.5. Análisis de rasgo latente

Hasta aquí, pues, nuestros análisis de clases latentes. Nuestras dos clases en M_1 identifican, como ya sabemos bien, dos grupos de modo claro: los «Firmes» ($X=2$) y los «Benévolos frente al delito» ($X=1$) —o, si se prefiere, punitivos y no punitivos—. Adviértase que se trata de clases que no son, según nuestro punto de partida, estrictamente nominales, sino que *guardan un orden entre ellas*. Esto es, que las dos

¹²⁹ También es superior a su alternativo, que incluye los términos interactivos referidos, puesto que $\text{Chi}^2=32,7446$ ($p=0,1096$); $L^2=36,0484$ ($p=0,0543$); $gl=24$; $d=0,0633$; $\text{BIC}(L^2)=-125,9218$; $\text{AIC}(L^2)=-11,9516$.

Acerca de supuestos en los que los estadísticos de ajuste no coinciden, como aquí, vid. McCUTCHEON y MILLS, 1998: 90. Adviértase, por otro lado, que los datos no coinciden con el caso de M_1 , ya que han debido adaptarse para añadir S, con lo que no es posible llevar a cabo comparaciones.

¹³⁰ Sobre todo ello, vid. BIEMER, 2011: 148 sobre todo, así como 72-77, 126 y 144-155.

¹³¹ BIEMER, 2011: 126, 203 sobre todo y 329.

¹³² Esta asunción guarda una íntima relación con nuestra hipótesis de unidimensionalidad, que aquí asumimos se cumple cuando nuestro modelo alcanza un ajuste aceptable; vid. así las reflexiones de EMBRETSON y REISE, 2000: 228.

clases latentes representan «diferentes niveles de una escala ordenada»¹³³. CROON señala incluso cómo en ocasiones las clases pueden ordenarse *a lo largo de un continuo* en el sentido de que éstas sólo representan la división de una única población. El mismo autor aclara que los análisis ordinarios de clases latentes no son suficientes para confirmar este orden¹³⁴. Así las cosas, aquí podemos, incluso, dar un paso más e hipotetizar que nuestra variable latente X de interés es no ya ordinal, sino continua. Ello no sólo ofrecería una mejor descripción, sino que en hipotéticos análisis posteriores se podría utilizar más información. De este modo procedemos a testar nuestra cuarta hipótesis: *la variable latente firmeza frente al delito tiene una naturaleza continua* (H_4).

En efecto, es bien sabido que las variables latentes pueden perfectamente ser continuas, naturalmente también en el caso de ítems nominales u ordinales. Estos modelos que combinan variables manifiestas categóricas u ordinales con variables latentes que se encuentran en su origen de tipo continuo son denominados de *rasgo latente* —o *de teoría de respuesta al ítem*¹³⁵—. Los mismos también pueden entenderse como una forma de análisis factorial cuyos indicadores son de nuevo categóricos u ordinales y el o los factores continuos. Se trata de una herramienta muy flexible y puede utilizarse, como nos interesa aquí, para valorar hipótesis sobre la naturaleza y distribución de un constructo, o bien simplemente para la reducción de datos o para otras formas exploratorias¹³⁶.

Procedemos, pues, a testar nuestro modelo de rasgo latente con A, B, C y D (M_4), utilizando los datos a nuestra disposición. La siguiente Tabla 8 ofrece los estadísticos de bondad de ajuste más los grados de libertad para el mismo, así como para nuestro modelo de dos clases latentes (M_1), sobre cuyo buen ajuste nos hemos extendido más arriba.

Como puede observarse, el modelo de rasgo latente (M_4) se ajusta a los datos de un modo marginal¹³⁷. Cuenta, eso sí, con el BIC más bajo de todos los modelos ensayados en este trabajo, a la par que des-

¹³³ DAYTON, 1998: 2; también COLLINS y LANZA, 2010: 34.

¹³⁴ CROON, 2002: 135-136 y 160-161, así como 141-152 para un modelo de clases ordinales.

¹³⁵ BIEMER, 2011: 119; EMBRETSON y REISE, 2000: 40-61; HEINEN, 1996: 26-28, 91-120, 93-97 sobre todo y 150-190; REEVE y MÄSSE, 2004: 249-155; SKRONDAL y RABE-HESKETH, 2007: 715-717 y 727-728; UEBERSAX, 1997: 188-193; VERMUNT, 1997b: 38-40.

¹³⁶ HEINEN, 1996: 26-28, 55-60, 69-70, 91-120 y 151-152.

¹³⁷ Sobre la evaluación de estos modelos, con un enfoque bastante sofisticado, vid. EMBRETSON y REISE, 2000: 233-238.

Tabla 8. Estadísticos de bondad de ajuste para un modelo de clases latentes y un modelo de rasgo latente

Modelo	Chi ² [p]	L ² [p]	BIC (L ²)	AIC (L ²)	GL
2 clases latentes (M ₁)	18,2038 [0,1096]	17,815 [0,1214]	-63,1701	-6,185	12
Rasgo latente (M ₄)	24,4162 [0,0408]	17,815 [0,0240]	-68,2273	-1,7446	14

N=853.

taca también por su parsimonia, con hasta 14 grados de libertad¹³⁸. Igualmente $d=0,0568$ sobrepasa ligeramente los límites recomendables. No presentan problemas dignos de mención los residuos típicos ($\leq |1,781|$), salvo para el caso de $y_i=(2, 1, 1, 2)$ en el que se superan por poco los límites de la significación ($=-2,235$) —ello sugiere que el modelo no ofrece para dicho patrón una buena estimación, pero, por lo demás, no debería ser muy problemático—. No se aprecian problemas de identificación (autovalores $\geq 0,0641$), de óptimos locales o de otra naturaleza. Hemos testado M₄ bajo la asunción de distribución normal¹³⁹.

De una comparación entre M₁ y M₄ parece que el primero es superior. Así, los análisis parecen arrojar en principio evidencia que invita a rechazar nuestra cuarta hipótesis si la interpretamos como que identifica el mejor modelo¹⁴⁰. Ahora bien, esta afirmación puede matizarse si atendemos a análisis con datos perdidos —que poten-

¹³⁸ Como era de esperar dada su naturaleza, los estadísticos globales para la clasificación *errores de clasificación* ($E=0,604$) y *reducción en la proporción de errores* ($\lambda=0,1379$) son también ahora pobres. Por supuesto, es esperable que se incurra en muchos errores cuando, digamos, existen muchas más clases latentes —esto es cuando se asume la distribución a lo largo de un continuo—.

¹³⁹ En particular, hemos especificado que el método para aproximar la distribución normal sea el de *densidad reescalada* y que el rango de la distribución normal sea -3 a $+3$. Esto se puede realizar mediante la opción «nor(·)» de LEM.

Los modelos de rasgo latente también realizan una serie de asunciones, vid. EMBRETSON y REISE, 2000: 231-233; REEVE y MÄSSE, 2004: 255-256. De nuevo, la asunción de independencia local es nuclear. REEVE y MÄSSE, 2004: 256, sin embargo, mantienen que ningún programa estadístico es capaz de testar esta hipótesis, a la vez que recomiendan una estrategia que se aproxima a la construcción de modelos —algo que, ya se ha dicho, no forma parte de nuestros intereses—.

¹⁴⁰ Aunque se muestra cauteloso, UEBERSAX afirma que modelos de clases latentes y de rasgo latente pueden compararse entre sí utilizando índices como BIC, 1997: 191, si bien recomienda otros criterios.

cialmente pueden ser más precisos— y si interpretamos H_4 de modo más ortodoxo en el sentido de que tanto una *estimación categórica* como una *continua* pueden no sólo ser de utilidad sino ajustarse a la realidad. Así, debemos posponer nuestras conclusiones para esta cuarta hipótesis hasta el siguiente epígrafe sobre análisis con datos perdidos.

3.6. *Análisis con datos perdidos*

Como se dijo, los presentes análisis han sido realizados sin tomar en cuenta la existencia de datos perdidos. La Tabla 1 mostraba el número de ellos para cada uno de los ítems, y en conjunto la pérdida ascendía para nuestros análisis a 102 (10,68%). Aunque tal porcentaje de pérdida de datos es habitualmente asumible, no se puede descartar que, aún así, se produzcan serios problemas. Ello aconseja explorar hasta qué punto esta situación afecta a nuestros modelos principales, esto es M_1 y M_4 . La literatura ha dedicado una importante atención a esta circunstancia ya en el caso de datos categóricos¹⁴¹. Una primera opción que viene a la mente de modo intuitivo es la de imputar los datos perdidos en la matriz originaria a partir de algún procedimiento válido para variables dico o politómicas y, a partir de ahí, calcular las celdas para el análisis de clases latentes¹⁴². Vermunt, por el contrario, sugiere una interesante estrategia alternativa consistente en especificar tipos diferentes de tablas de frecuencias observadas correspondientes a subgrupos distintos de individuos, subgrupos compuestos por aquellos para los que se dispone del mismo tipo de información¹⁴³. La lógica es muy simple¹⁴⁴: es posible que algunos sujetos no entren a formar parte de la muestra si su información está perdida para algún ítem, pero no para los demás; de este modo, es posible recuperar la información disponible mediante tablas que incluyan los ítems a los que sí hayan contestado. De este modo, en efecto, se pueden especificar subgrupos de individuos según sus patrones de respuesta. En nuestros análisis para datos perdidos nos hemos limitado a añadir la Tabla (y el subgrupo) AB debido a que al

¹⁴¹ BIEMER, 2011: 311-315; COLLINS y LANZA, 2010: 25, 39, 80-81 y 85; VERMUNT et al., 2008: 369-391; WINSHIP et al., 2002: 408-430.

¹⁴² Vid. una estrategia general en esta línea en ALLISON, 2002: 19-27; COLLINS y LANZA, 2010: 80-81; ENDERS, 2006: 315-339.

¹⁴³ VERMUNT, 1997b: 49-50.

¹⁴⁴ *Muy simple* si se considera la forma en que se introducen habitualmente los datos en LEM mediante tablas de contingencia. Existen, eso sí, otras maneras de indicar al programa qué datos debe utilizar.

agregar otras se producía un problema de escasez que se hacía cada vez más serio, a la par que la información que se iba añadiendo era cada vez menor. Con nuestra estrategia, la menor frecuencia de una celda es seis. Ello es debido en buena medida a que D es, como vimos, el ítem con más datos perdidos. Siguiendo este procedimiento propuesto por Vermunt se utilizan finalmente 911 casos de un total de 955 (953) observaciones. Ello representa un 95'39% del total, o lo que es lo mismo, los datos perdidos se han rebajado a menos del 5% del total, esto es a menos de la mitad de los primeros análisis. La Tabla 9 ofrece los estadísticos de bondad de ajuste y los grados de libertad para nuestros modelos de referencia M_1 y M_4 .

Tabla 9. Estadísticos de bondad de ajuste para un modelo de clases latentes y un modelo de rasgo latente, análisis para datos perdidos (Tablas ABCD y AB)

Modelo	Chi ² [p]	L ² [p]	BIC (L ²)	AIC (L ²)	GL
2 clases latentes (M_1)	21,4278 [0,1237]	20,9335 [0,139]	-81,2846	-9,0665	15
Rasgo latente (M_4)	27,568 [0,0502]	29,3316 [0,0316]	-86,5156	-4,6684	17

N=911.

Es claro que nuestro interés ahora se centra en comprobar si la pérdida de datos está afectando a nuestros análisis principales limitados a datos con información completa. Esta es una vocación fundamentalmente confirmatoria para M_1 y más sustantiva, como se adelantó, para M_4 —y ya estamos familiarizados con los dos, especialmente con el primero de ellos—. El modelo de dos clases latentes sin restricciones (M_1) sigue ajustándose bien a los datos (aunque ahora $d=0,0508$). Ello refuerza nuestra confianza en el análisis principal así como en nuestras pruebas a favor de la hipótesis H_1 ¹⁴⁵.

El modelo de rasgo latente (M_4) apunta un ajuste también ahora marginal. Así, para Chi² $0,1 > p > 0,05$ y para L² $0,05 > p > 0,01$, si bien no puede olvidarse que estos estadísticos tienden a ser, a nuestros intereses, *conservadores*, al menos con muestras grandes y que, por ese

¹⁴⁵ Puesto que los datos utilizados no coinciden, no es posible emplear BIC y AIC para comparar estos análisis con los realizados con datos con información completa. Ello, en todo caso, carece de mayor interés en el conjunto de nuestra estrategia.

motivo, incluso buenos modelos podrían ser rechazados conforme a ellos. En todo caso, debe insistirse en que L^2 (sólo) es significativo al nivel $\alpha=0,05$, pero no 0,01. De nuevo el índice de disimilitud es aceptable sólo marginalmente (d para $M_4=0,0564$). Tampoco, pues, este esfuerzo ofrece resultados completamente nítidos, definitivos sobre H_4 , aunque ahora es un poco más optimista.

Una ulterior estrategia es tratar de optimizar aún más la información disponible. Así, puede ser recomendable llevar a cabo análisis de rasgo latente o de modelos de teoría de respuesta al ítem mediante imputación de datos perdidos con EM¹⁴⁶. De este modo es posible repetir nuestra estimación incluyendo todas las observaciones ($N=953$)¹⁴⁷. La Tabla 10 ofrece los estadísticos de bondad de ajuste y los grados de libertad para M_4 con imputación de datos perdidos¹⁴⁸.

Estos resultados, aunque tienen sus complicaciones¹⁴⁹, favorecen más claramente la posibilidad de que la variable latente firmeza fren-

Tabla 10. Estadísticos de bondad de ajuste para un modelo de modelo de rasgo latente, análisis para datos perdidos

Modelo	Chi ² [p]	L ² [p]	BIC	AIC	GL
<i>Rasgo latente</i> (M_4)	0,508 [0,7756]	0,542 [0,7625]	4818,687	4779,81	2

N=953.

Estimador=Máxima verosimilitud (ML).

Enlace=Logit.

p para test de Chi² para MCAR>0,2.

¹⁴⁶ Mplus ofrece esta opción, MUTHÉN y MUTHÉN, 2010: 39-40, 60-61 y 186-187. Nótese en todo caso, y esta es una advertencia importante, que el procedimiento de Mplus en análisis de rasgo latente es notablemente distinto del de LEM en estos casos (1). Por este motivo no es posible atribuir los cambios simplemente al mayor número de observaciones utilizadas. (1) Vid., sobre enfoques de máxima verosimilitud frente a los puramente probabilísticos, McCUTCHEON, 1987: 21-27.

¹⁴⁷ Con el estimador ML no es posible mantener las ponderaciones en Mplus (de ahí que $N=953$). Otros estimadores sí lo permiten, como MLR, que tiene una naturaleza robusta; cuando se utiliza *MLR con ponderaciones* (e imputación de datos perdidos, claro está), los resultados no se alteran de modo significativo y todas las conclusiones del texto siguen siendo igual de válidas. Sobre estos estimadores, vid. MUTHÉN y MUTHÉN, 2010: 533.

¹⁴⁸ Para nuestra variable nominal, esto es no ordinal en el lenguaje de Mplus, utilizamos C#2.

¹⁴⁹ Sobre algunas de las mismas, derivadas de sus asunciones, vid. SERRANO MAÍLLO, 2008: 164-166.

te al delito se distribuya en nuestro estudio de modo continuo. La evidencia, entonces, favorece ahora H_4 , aunque sería, en efecto, muy deseable contar aquí con ulteriores y más sofisticados estadísticos de bondad de ajuste¹⁵⁰. Vista en su conjunto, la evidencia que hemos obtenido sobre H_4 parece más bien mixta. No es preciso, sin embargo, elegir entre una variable latente categórica o continua, sino que ambas son opciones lícitas y que pueden resultar útiles en situaciones muy heterogéneas¹⁵¹. En palabras de COLLINS y LANZA, «Nuestro punto de vista es que muchos fenómenos pueden tener características tanto continuas como categóricas»¹⁵². Por último, debe subrayarse que una mirada a los resultados del modelo sugieren a las claras, una vez más, que es dudoso que D forme parte de esta variable latente (continua).

4. DISCUSIÓN

A pesar de que la firmeza frente al delito se ha convertido en uno de los objetos de estudio más importantes de la Criminología contemporánea, todavía queda mucho por hacer en ámbitos básicos tan relevantes como su medición, precisamente lo que aquí más nos ha preocupado. En el presente estudio hemos estado interesados en la identificación de una variable latente de firmeza o punitividad a partir de opiniones sobre el castigo, a nivel individual. Así, hemos recurrido a modelos de clases latentes para testar, con evidencia favorable a la misma, la hipótesis H_1 de que *existe una variable latente categórica de firmeza frente al delito con indicadores también categóricos de actitudes ciudadanas sobre el castigo*. En particular, el modelo de dos clases latentes (M_1) se ajusta bien a los datos y no se han encontrado problemas de ningún tipo ni tampoco violaciones de las asunciones más importantes; aunque existen dudas razonables sobre si este modelo es superior al de tres clases, algo que aquí hemos decidido sobre la base de la parsimonia. Análisis con datos perdidos han confirmado los hallazgos ordinarios. Nuestra segunda hipótesis de que *esta variable latente categórica es unidimensional e incluye indicadores de actitudes ciudadanas directas e indirectas sobre el castigo* (H_2), sin embargo, debe rechazarse puesto que nuestra variable observada indirecta de tipo procesal-penal no parece formar parte de modo claro de aquélla si se atiende a las probabilidades condicionales y a los pseudoeficientes de determinación. Puesto que la pregunta utilizada

¹⁵⁰ EMBRETSON y REISE, 2000: 229.

¹⁵¹ SERRANO MAÍLLO, 2011: 32

¹⁵² COLLINS y LANZA, 2010: 10.

presentaba serios problemas, es menester, sin embargo, poner en cuarentena esta cuestión de la dimensionalidad. Dicho de otro modo, si la firmeza frente al delito es empíricamente un constructo unidimensional o si, por el contrario y como más bien sugiere nuestra investigación, tiene una naturaleza mixta, heterogénea —y la opinión sobre aspectos procesal-penales constituye una de sus dimensiones— requiere más estudios¹⁵³. La evidencia claramente rechaza la hipótesis tercera, de acuerdo con la cual *existiría en la variable latente categórica firmeza frente al delito una clase de «ideólogos»* (H_3). Finalmente, para nuestra última hipótesis —*la variable latente firmeza frente al delito tendría una naturaleza continua* (H_4)— las pruebas encontradas, aunque favorables, probablemente no pueden considerarse definitivas.

Como se ha repetido con insistencia, el enfoque de clases latentes está llamado a desempeñar un rol decisivo en la medición de nuestro objeto de estudio, donde la investigación ha tendido a ser hasta ahora limitada. Muchos estudios utilizan una única medida de punitividad o firmeza frente al delito; y algunos otros que incluyen varias se limitan a análisis particulares para cada una de ellas, pero sin tratar de combinar los distintos ítems¹⁵⁴. Estas modestas aproximaciones son claramente inferiores a otros esfuerzos. HARTNAGEL y TEMPLETON utilizan siete preguntas para medir la punitividad, pero a continuación se limitan a sumar las respuestas (y dividir el resultado entre siete), con lo cual ignoran completamente la cuestión de la dimensionalidad que aquí nos preocupa —a la vez que también siembran serias dudas sobre sus análisis¹⁵⁵—. Algo semejante hace BRILLON cuando crea una escala de punitividad mediante la suma de las respuestas a cuatro ítems, si bien igualmente informa del coeficiente de fiabilidad alfa para los mismos¹⁵⁶; así como CALLANAN en su investigación sobre las leyes de *tres strikes y estás eliminado*¹⁵⁷. Avanzando un pequeño paso, KUTATELADZE dedica mucha atención a este tema utilizando cuarenta y cuatro ítems sobre punitividad estatal. Sin embargo, primero anuncia unidimensionalidad usando un análisis de fiabilidad mediante el alfa de CRONBACH y más adelante de multidimensionalidad a partir de una matriz de correlaciones con los ítems agrupados en cinco conjuntos¹⁵⁸. Estos enfoques, así las cosas, son insuficientes,

¹⁵³ KURY, OBERGFELL-FUCHS y WÜRGER, 2002: 3.

¹⁵⁴ Así, por ejemplo, CESARONI y DOOB, 2003: 438-439; o SECRET y JOHNSON, 1989: 366-370 por ejemplo.

¹⁵⁵ HARTNAGEL y TEMPLETON, 2008: 359.

¹⁵⁶ BRILLON, 1988: 95-96.

¹⁵⁷ CALLANAN, 2005: 100-101.

¹⁵⁸ KUTATELADZE, 2011: 156 nota 7 y 158 sobre todo.

como vamos a ver enseguida, para una cuestión ciertamente compleja y es dudoso que puedan, por lo tanto, asumirse¹⁵⁹. NEIL y LYNCH combinan varias preguntas relativamente heterogéneas y afirman que el apoyo a la pena de muerte y la percepción de la benevolencia de los tribunales son cosas diferentes, apuntando a la hipótesis de la pluridimensionalidad¹⁶⁰. Sin embargo, no hay aquí ningún esfuerzo de test más formal y, además, la validez de la segunda pregunta para la medición de la firmeza frente al delito no es evidente. Entre las aproximaciones más sofisticadas que aparecen con frecuencia en la literatura figuran los análisis de componentes principales o factoriales de naturaleza exploratoria. KURY y OBERGFELL-FUCHS utilizaron preguntas sobre veintiséis formas delictivas distintas para medir la punitividad a nivel individual, las cuales incluían desde la conducción sin carnet y el consumo de hachís hasta el infanticidio y varias modalidades de violación y otros delitos sexuales. Sometieron dichos items, previa retirada de dos de ellos, a un análisis de componentes principales y encontraron que una solución de cinco factores era la preferible¹⁶¹. Concluyen KURY y OBERGFELL-FUCHS que «la punitividad no es un constructo unificado, no existe la punitividad como tal, sino solamente aspectos diferentes de la misma»¹⁶²; así como que «Claramente, la punitividad no es un concepto unitario. Asume varias dimensiones [...] La punitividad no existe»¹⁶³. Más adelante someten dichos cinco factores a análisis de regresión lineal con las mismas variables independientes, con resultados muy diferentes entre sí¹⁶⁴. Sobre esto último, sin embargo, debe advertirse que al menos en parte los hallazgos de la regresión deben encontrarse influenciados por el tipo de rotación que utilizan, varimax¹⁶⁵, la cual impone ortogonalidad en los componentes o factores¹⁶⁶. Aquí deben también mencionarse las consideraciones críticas a que han sido sometidos los análisis exploratorios de componentes principales y factoriales¹⁶⁷; la limitación al primer orden de análisis; así como, quizá, que tan poco plausible intuitivamente como la solución unidimensional es la imagen que proyectan KURY y sus colegas de un constructo con un elevado número de dimensiones absolutamente heterogéneas entre sí. También utiliza un análisis de componentes principales ROBBERS: a

¹⁵⁹ BOLLEN, 1989: 226-232, por ejemplo y entre otros pasajes.

¹⁶⁰ NELLIS y LYNCH, 2008: 45-46.

¹⁶¹ KURY y OBERGFELL-FUCHS, 2008: 288-289.

¹⁶² KURY y OBERGFELL-FUCHS, 2008: 299.

¹⁶³ KURY et al., 2008: 130.

¹⁶⁴ KURY y OBERGFELL-FUCHS, 2008: 296-297.

¹⁶⁵ KURY y OBERGFELL-FUCHS, 2008: 288.

¹⁶⁶ DUNTEMAN, 1989: 49.

¹⁶⁷ BASILEVSKY, 1994: ix-xii.

partir de una escala de diez items sobre la pena de muerte se decanta por una solución de un único factor, lo cual apunta a la unidimensionalidad¹⁶⁸ —aunque se echa de menos más información para valorar esta decisión—. Nótese en todo caso que se trata de un estudio con un objeto más limitado que el nuestro. Y algo semejante cabe decir del trabajo de TYLER y WEBER, quienes informan de una solución de cuatro factores principales a partir de doce interrogaciones sobre la pena de muerte, a la par que igualmente indican que estos cuatro factores correlacionan entre sí de modo significativo. Estos factores se toman como motivos para el apoyo a la pena de muerte¹⁶⁹.

Así las cosas, se requiere a mi juicio superar enfoques metodológicamente limitados o al menos insuficientemente sofisticados como los precedentes —que no permiten, verbigracia, excluir efectos de la muestra; ni testar que los items compartan una causa común— y sustituirlos por otros más avanzados tales como los de variables latentes, de los que forma parte el análisis de clases latentes¹⁷⁰. En especial HATTIE ha mostrado de modo convincente que enfoques tales como los basados en la consistencia de los patrones de respuesta, estimaciones de fiabilidad o análisis de componentes principales o factoriales (exploratorios), entre otros, son en efecto insuficientes para nuestra misión por diversas razones, entre las que se encuentran su incapacidad para distinguir entre soluciones uni- y pluridimensionales, carecer a menudo de distribuciones muestrales conocidas, ignorarse cómo es su comportamiento bajo determinadas circunstancias y faltar guías claras para su empleo¹⁷¹.

Nuestro estudio, entonces, se ha centrado en el modelo de medición, y a tal fin ha recurrido con especial decisión a los *análisis de clases latentes* —aunque, como debería ser ya obvio, hemos seguido la recomendación general de utilizar varias técnicas de modelaje¹⁷²—. Los mismos son especialmente prometedores en nuestra disciplina y en las ciencias humanas y sociales en general debido a que asumen que sus indicadores se encuentran medidos únicamente a nivel nominal u ordinal. No puede exagerarse la importancia de esta circunstancia habida cuenta de que los criminólogos no siempre disponen de variables

¹⁶⁸ ROBBERS, 2006: 209.

¹⁶⁹ TYLER y WEBER, 1982: 32-33. Existen más ejemplos de investigaciones que incluyen varias preguntas pero todas ellas sobre la pena de muerte, y que se preocupan por la cuestión de la dimensionalidad, vid. por ejemplo KEIL y VITO, 1991: 457 sobre todo.

¹⁷⁰ Existen otras alternativas, aunque a mi juicio menos prometedoras, vid. Hox, 2008: 397-401.

¹⁷¹ HATTIE, 1984: 49-78; el mismo, 1985: 139-164.

¹⁷² UEBERSAX, 1997: 188.

medidas a nivel de razón o de intervalo¹⁷³, a pesar de lo cual recurren habitualmente a herramientas de análisis estadístico que asumen, entre otras cosas, este nivel de medición¹⁷⁴. Del mismo modo, en ocasiones es perfectamente imaginable que se hipoteticen una variable latente categórica. La técnica es flexible y poderosa, en el sentido de que abre la puerta al test de muchos modelos distintos, con y sin base teórica; y, a la vez, permite relajar asunciones de métodos más tradicionales¹⁷⁵. Su flexibilidad, así, permite su utilización en diversos frentes, como los que siguen: reducción de datos, análisis de heterogeneidad no observada, refinamiento de escalas de medición, investigación de datos perdidos, test de determinadas hipótesis, exploración o confirmación de la existencia de variables latentes unidimensionales, etc.¹⁷⁶. Como mínimo, en cuestiones extremas tiene un valor heurístico innegable¹⁷⁷. Así las cosas no puede sorprender que haya llegado a proponerse su utilización de modo rutinario¹⁷⁸.

Otra función importante de los enfoques de variables latentes que debe añadirse a las anteriores y que puede resultar especialmente útil para el estudio de la firmeza frente al delito —en realidad para infinitísimos objetos de estudio— tiene que ver con el *error de medición*¹⁷⁹. Al combinar varios indicadores individuales —que a menudo, como digo, han sido medidos a nivel categórico u ordinal—, permiten una evaluación (incluso un control) del error en que se incurre con cada uno de ellos. El razonamiento es fácil de seguir. Cuando se utiliza un único indicador manifiesto, esto es de modo aislado, se asume que se trata de un *indicador perfecto*, no afectado por error de medición¹⁸⁰. Es bien sabido tanto que esta asunción tiende a ser insostenible porque toda medición de este tipo conlleva un cierto grado de error¹⁸¹, como que suele ser la estrategia habitual en la investigación criminológica en la materia. Por el contrario, el enfoque de variables latentes, al permitir varios indicadores de modo simultáneo y extraer lo que de común tienen —en cuanto que, como

¹⁷³ CLARKE y McCUTCHEON, 2009: 713.

¹⁷⁴ Debe añadirse que otras herramientas, como es el caso de los modelos de ecuaciones estructurales, están avanzando mucho en la incorporación de variables que no son continuas ni se distribuyen de modo normal, FINNEY y DiSTEFANO, 2006: 269-301.

¹⁷⁵ BIEMER, 2004: 227, 243 y 245; McCUTCHEON, 1987: 7-8 y 79-80; McCUTCHEON y HAGENAARS, 1997: 266 y 276.

¹⁷⁶ McCUTCHEON y MILLS, 1998: 94 nota 2.

¹⁷⁷ EMBRETSON y REISE, 2000: 228.

¹⁷⁸ UEBERSAX y GROVE, 1990: 559.

¹⁷⁹ BIEMER, 2004: 226-228 y 242-246 sobre todo; COLLINS y LANZA, 2010: 74-75.

¹⁸⁰ McCUTCHEON y MILLS, 1998: 82.

¹⁸¹ Vid. COLLINS y LANZA, 2010: 27, 45 y 47; HOX, 2008: 387 y 401.

pensamos en esta investigación, están causados por otra variable no observada—, permite una valoración del error de cada uno de ellos en particular¹⁸². Desde esta perspectiva, los indicadores individuales ya no se asumen mediciones perfectas, sino «mediciones falibles de una variable latente»¹⁸³.

Nuestro análisis sobre el error —con la notable limitación, como se dijo, de que no se asume que la variable latente de interés teórico haya sido verdaderamente medida sin error— deja claras las limitaciones de las mediciones de la firmeza frente al delito basadas en una única cuestión o pregunta¹⁸⁴, en particular si se realizan de modo acrítico y acontextual o aislado de ulteriores investigaciones. Así, KURY y sus asociados confirman que «Los resultados dejan claro [...] que, por lo tanto, también es problemática su operacionalización con el ítem sobre la pena de muerte o con cualquier otro procedimiento de medición reducido»¹⁸⁵.

Se dirá, de manera intuitivamente irreprochable, que estos resultados abiertamente pesimistas para las preguntas únicas, aisladas son consecuencia de haber incluido en nuestro modelo una variable que en realidad no se encuentra relacionada con la variable latente identificada y de mi obstinación en no introducir cambios —*mejoras* si se prefiere— en los modelos a partir de una base puramente empírica. A continuación, la Tabla 5bis —así llamada por su parentesco con la Tabla 5, *supra*— ofrece los resultados de clase latente para un modelo de dos clases limitado ahora a A, B y C, esto es excluyendo el ítem D de naturaleza procesal-penal¹⁸⁶.

Una somera mirada a la misma deja clara no sólo la tremenda carga de error de medición de los ítems individuales, sobre todo para $X=2$, sino las sobresalientes semejanzas con las tasas de error derivadas de las probabilidades condicionales de M_1 , esto es incluyendo D.

Pero a mi modo de ver, como se dijo, estas consideraciones no quieren decir que necesariamente deba renunciarse a estos indicadores individuales. Si somos capaces de replicar determinados hallazgos con operacionalizaciones y mediciones distintas de un mismo

¹⁸² Ni que decir tiene que la familia de los modelos de variables latentes ofrece más opciones, incluyendo la identificación de variables latentes con control del error en los indicadores.

¹⁸³ BIEMER, 2004: 227.

¹⁸⁴ CULLEN et al., 2009: 77; KEIL y VITO, 1991: 455; KURY, OBERGFELL-FUCHS y WÜRGER, 2002: 4 y 170; KURY et al., 2004: 97; MAYHEW y VAN KESTEREN, 2002: 67.

¹⁸⁵ KURY, OBERGFELL-FUCHS y WÜRGER, 2002: 170.

¹⁸⁶ Sobre su ajuste, vid. *supra* nota 107.

Tabla 5bis. Resultados de clase latente para un modelo de dos clases latentes con indicadores A, B y C (D excluido)

		Clase 1 (No punitivo)	Clase 2 (Punitivo)
		0,2735	0,7265
A [ROBAR]	1	0,9724	0,6616
	2	0,0276	0,3384
B [PENA_CAPITAL]	1	0,8812	0,5480
	2	0,1188	0,4520
C [MGF]	1	0,2611	0,0387
	2	0,6026	0,7681
	3	0,1363	0,1932

$E=0,2225$

$\lambda=0,1865$

constructo —incluidas las basadas en un único ítem—, siempre que gocen de una mínima validez, entonces nuestra confianza en dichos hallazgos se verá reforzada. Es lo que BRAITHWAITE denomina *la concurrencia de las debilidades*¹⁸⁷. Verbigracia, a menudo los investigadores desean utilizar datos secundarios con muestras elevadas y una información relativamente rica, pero con mediciones pobres para constructos nucleares, como es el caso de la firmeza frente al delito¹⁸⁸. Así, a modo de ejemplo, algunas encuestas del CIS incluyen una pregunta sobre la pena de muerte¹⁸⁹. Hemos comprobado que ésta es una estimación muy pobre de la firmeza frente al delito, pero, por un lado, análisis de clases latentes pueden ofrecer una aproximación al error de medición en que incurrir; y, por otro, un conjunto de estudios empíricos independientes con medidas modestas (aunque mínimamente asumibles) pero distintas que arrojan unos mismos resultados, pueden ofrecer resultados verosímiles cuando coinciden¹⁹⁰.

En la medida de lo posible, sin embargo, deben evitarse las preguntas únicas. Nuestro estudio sugiere que las mismas pueden en-

¹⁸⁷ BRAITHWAITE, 1979: 22.

¹⁸⁸ RIEDEL, 2000: 7-8, incidiendo en que ésta suele ser una de las principales dificultades de los datos secundarios.

¹⁸⁹ Vid. SERRANO MAÍLLO y KURY, 2008: 328-329. Nada de ello debe hacernos olvidar las hondas y merecidas consideraciones críticas que ha merecido este enfoque basado en la pregunta sobre la pena de muerte, así KURY et al., 2008: 130; NELLIS y LYNCH, 2008: 34, 37, 41 y 46.

¹⁹⁰ CULLEN et al., 2000: 6, son incluso más optimistas y señalan que la pregunta sobre la pena de muerte puede valer para una estimación «global».

contrarse fuertemente contaminadas por error de medición y relacionarse sólo débilmente con la variable latente de interés. Aunque nuestra investigación se centra en el nivel individual de análisis, es esperable que conclusiones como éstas sean ampliables a otros niveles y que, en particular, la habitual mención en solitario de las tasas de encarcelamiento sea una opción en el mejor de los casos exigua.

Finalmente, los modelos de clases latentes no deben interpretarse demasiado¹⁹¹, pero en la presente investigación parece existir un orden o incluso un continuo a lo largo de los cuales se puede bien *ordenar* bien *colocar a distancias conocidas* a los grupos de individuos. ¿Es nuestra variable latente esencialmente ordinal o continua? Existe una cierta polémica acerca de si las variables categóricas son así de modo inherente o si, por el contrario, una mayoría de variables son conceptualmente continuas, si bien a menudo sólo son observables o mensurables a nivel nominal¹⁹². En palabras de COLLINS y LANZA, «Nuestro punto de vista es que muchos fenómenos pueden tener características tanto continuas como categóricas», «Más que un debate acerca de si el consumo de alcohol es verdaderamente continuo o categórico, más bien consideraríamos si una operacionalización continua o categórica es más relevante para las preguntas de la investigación de que se trate». Desde este punto de vista, que parece razonable, tanto el modelo de dos clases latentes como el de rasgo latente pueden ser plenamente útiles. Concluyen los autores con que «simplemente hacemos la modesta asunción de que, dado un conjunto particular de preguntas de investigación y un conjunto de datos disponible, un modelo categórico puede proporcionar información útil»¹⁹³.

La investigación empírica española sobre el nivel individual de análisis de nuestro objeto de estudio ha sido hasta la fecha muy escasa. En la mayor parte de los casos, además, se ha basado en preguntas únicas. En algunas investigaciones, incluso, había más de un indicador de punitividad, pero cada uno se utilizaba de modo aislado, esto es que no se combinaban entre sí de ninguna manera¹⁹⁴. Existen al menos tres investigaciones en España que han utilizado más de una pregunta, ítem o variable observada para la medición de la punitividad o de la firmeza/benevolencia frente al delito a nivel individual.

¹⁹¹ AGRESTI, 2002: 544; UEBERSAX, 1997: 191.

¹⁹² POWERS y XIE, 2000: 7-11, 8 en particular.

¹⁹³ COLLINS y LANZA, 2010: 10. Incluso la elección de un modelo frente a otro puede depender de la finalidad, así EMBRETSON y REISE, 2000: 75.

¹⁹⁴ Vid., así, en un riguroso trabajo, FERNÁNDEZ MOLINA y TARANCÓN GÓMEZ, 2010: 5-6 y 13 sobre todo.

VARONA GÓMEZ, en un artículo pionero, llevó a cabo una encuesta enviando por email el cuestionario a los alumnos de la Universidad de Gerona. El mismo incluía cuatro preguntas o escenarios en la terminología del estudio sobre la sanción que se impondría en otros tantos supuestos. Las conductas delictivas sobre las que se interrogaba eran robo en domicilio por parte de un reincidente, conducción bajo los efectos del alcohol que ocasionaba un accidente leve, violencia de género habitual y tráfico de drogas a baja escala. El autor divide su muestra en dos grupos de punitivos y otros, estando la primera categoría conformada por quienes habían elegido la pena de prisión en las cuatro preguntas. Este grupo de punitivos («Muestra punitivos» en la terminología del trabajo) se encontraba integrado por treinta y siete individuos (4,55%) de un total de 813¹⁹⁵. Este proceder es merecedor de algunas consideraciones —aunque resulta superior a otros esfuerzos y marca el camino de los varios indicadores que debe seguirse—. No existe una razón clara para el punto de corte que se elige. Así, aquí se podría estar identificando a la mayoría los *muy firmes*. Aunque puede aventurarse que la tasa de falsos positivos (para los muy punitivos) es muy baja y aceptando dicha clasificación, es improbable que pueda decirse lo mismo de los falsos negativos. El grupo residual, en todo caso, probablemente se encuentra compuesto por una mezcla heterogénea. Al dividir la muestra en dos grupos, esto es al construir una variable ordinal dicotómica, el grado de información que se utiliza es bajo.

Helmut KURY y yo mismo utilizamos la encuesta del Centro de Investigaciones Sociológicas *Sondeo sobre la juventud española, 2005 (Primera oleada)* (estudio número 2596) en otra investigación pionera. Su universo estaba compuesto por adolescentes y jóvenes adultos de entre quince y veintinueve años de edad, de España. La misma interrogaba sobre la pena de muerte para personas con delitos muy graves y sobre la penalización del consumo de drogas. A partir de estas dos preguntas construimos un índice o, más propiamente, una escala de punitividad mediante la pura adición¹⁹⁶. Tampoco es una opción plenamente satisfactoria. La construcción de índices es siempre una labor problemática ya que, en alguna medida, se está alterando el orden natural de las cosas. Puesto que allí asumimos que quienes favorecen la pena de muerte son *los más firmes*, otorgamos un peso especial a esta variable en nuestra escala. El resultado fue una escala

¹⁹⁵ VARONA GÓMEZ, 2008: 24 sobre todo.

¹⁹⁶ SERRANO MAÍLLO y KURY, 2008: 329-333. Sobre los fundamentos metodológicos generales de la construcción de escalas —generalmente a partir de más de dos variables originales—, vid. SPECTOR, 1992: 10-46 sobre todo.

con un rango que iba de 0 (*poco punitivo*) a 5 (*muy punitivo*). Esto es, que el mismo tenía seis categorías. La asunción es que aquellos que tienen puntuaciones más altas en la escala tienen asimismo una mayor punitividad. Este procedimiento tiene la ventaja de medir la firmeza frente al delito de modo ordinal con más de cinco categorías de respuesta, lo cual permitió a continuación análisis de regresión lineal y ordinal. En conjunto y como ya debería estar claro, esta estrategia es claramente inferior a la de variables latentes —aunque en descargo puede decirse que un modelo de medición con tan poca información no está identificado— y que la carga de error de la escala no debe ser despreciable.

En un reciente artículo, yo mismo utilicé análisis de clases latentes para identificar una variable latente de firmeza frente al delito o punitividad sobre la base de cuestiones relativas a derechos fundamentales procesal-penales. Los datos procedían también de una encuesta del CIS, y las variables observadas, que tenían cuatro categorías de respuesta, se encontraban medidas a nivel ordinal¹⁹⁷. Para variables de este tipo y como ya sabemos, los análisis de clases latentes son considerados en la actualidad por la literatura como una de las herramientas estadísticas de referencia. Su principal desventaja, que ya nos es familiar, puede ser que, por su naturaleza probabilística, no permiten una clasificación de los sujetos sin un cierto grado de incertidumbre, con lo cual no es recomendable utilizarlos como variables dependientes en análisis posteriores. Puesto que la preocupación en aquel trabajo recaía en el modelo de medición, esta cuestión quedaba fuera de nuestros intereses, pero resulta evidente que ocupa un lugar primordial en la discusión contemporánea.

Limitadas como son por uno u otro motivo, estas tres aproximaciones pioneras en España apuntan a las claras a la utilización de varias preguntas de modo coordinado en la medición de la firmeza frente al delito. Esta estrategia, unida a los procedimientos adecuados para combinarlas, en particular en el paradigma de las variables latentes, es la propuesta del presente artículo.

5. CONCLUSIONES

En línea con una parte importante de la literatura, nuestra primera conclusión desaconseja, entonces, el recurso a preguntas únicas en la medición de la punitividad o firmeza frente al delito. A menudo,

¹⁹⁷ SERRANO MAÍLLO, 2011: 5-10.

sin embargo, esta habitual advertencia se basa más en la lógica general de la medición, por ejemplo en el marco de la teoría clásica de los tests, que en pruebas empíricas. En efecto, si $V=O+e$, entonces está claro que más ítems reducen el error. Otra cosa, por supuesto, se produce cuando añadimos de modo más realista un término de error sistemático o sesgo (s) a la fórmula como en $V=O+e+s$ ¹⁹⁸. El enfoque de clases latentes, por el contrario, permite, dentro de unos límites, una distinción entre tipos de error y una aproximación a la estimación del error aleatorio y sistemático de medición; y para nuestro caso los hallazgos sugieren que la asunción de que indicadores manifiestos únicos pueden considerarse *indicadores perfectos* no es sostenible. Antes al contrario, las tasas de error para las variables observadas en la medición de la punitividad son muy elevadas en nuestro estudio.

No importa repetir que ello no exige despreiciar *en todo caso* mediciones únicas, como suele ser el caso de encuestas generales que por un lado no dejan alternativa al investigador que desea utilizar datos secundarios o que, simplemente, no tiene otro remedio; pero que, por otro, pueden ofrecer importantes ventajas como muestras elevadas, mayor representatividad, trabajos de campo de calidad, etc. No sólo eso, los datos secundarios —que cuando tienen una vocación general sólo incluirán, como mucho, mediciones aisladas de firmeza frente al delito— son una buena manera de permitir que investigadores con escasos recursos pero independientes puedan también testar sus hipótesis y evitar que una financiación influenciada por factores extracientíficos determine los hallazgos de nuestra Criminología. Por todo ello, uno debe ser muy cauto a la hora de descartar fuentes de información empírica pues se corre el riesgo, si se me permite la expresión, de tirar el agua con el niño dentro; aunque también debe exigirse en estos casos una mínima validez.

Aunque es desde luego preferible, nuestro estudio también desaconseja la mera suma de las puntuaciones individuales cuando se utilizan varias preguntas o ítems como en algunas investigaciones españolas que se han mencionado más arriba. Como hemos visto, no sólo las tasas de error son elevadas y en parte este error no es aleatorio, sino que no es posible asumir de modo ingenuo que los errores simplemente se cancelan recíprocamente. Hox escribe en este sentido que «Esta no es una estimación muy buena, debido a que la relación entre la variable latente [...] y la suma de valores observada no es lineal. Además, si el número de ítems es pequeño [como suele ser el

¹⁹⁸ SPECTOR, 1992: 4-6 y 10-12.

caso], la medida no es fiable y la incerteza asociada con las estimaciones individuales de [...] [la variable latente] es demasiado grande como para ser ignorada»¹⁹⁹.

Con ciertas limitaciones, nuestro estudio apunta a que, tal y como resulta intuitivamente plausible y como mantienen una mayoría de autores, la firmeza frente al delito o punitividad parece ser, al menos a nivel individual, un constructo pluridimensional. De todos modos, nuestras pruebas distan de ser concluyentes, entre otros motivos por la especial complejidad de JUZGAR (D), que como se recordará interrogaba sobre un objeto de naturaleza procesal-penal, potencialmente conectada con el castigo de modo indirecto. Así, no es descabellado que al menos algunos entrevistados la hayan malinterpretado. En conclusión, entonces, si la firmeza es pluridimensional y, en su caso, cuál es su grado de complejidad, precisa de ulteriores investigaciones.

Nuestro enfoque cuantitativo también podría complementarse con investigación cualitativa. La cuestión precedente podría beneficiarse de la misma, pero también otras más puntuales que han ido saliendo a lo largo de nuestro estudio. Por ejemplo, nuestra investigación apunta a que la categoría de respuesta «Otra pena» en MGF (C) es problemática y que no puede interpretarse, al menos para un grupo de entrevistados, como una respuesta significativamente menos punitiva que la de prisión. En efecto, la probabilidad de que un sujeto de la clase «Firmes frente al delito» y de la clase «Benévolos» seleccione esta respuesta es prácticamente la misma (0,1738 vs. 0,1785). A mi modo de ver, la investigación cualitativa también podría arrojar luz sobre cuestiones de este tipo que a menudo se pasan por alto.

En definitiva, la presente investigación acude en apoyo de que la medición de la firmeza frente al delito o punitividad a nivel individual probablemente requiere esfuerzos más sofisticados que los habituales y que, consistentemente, muchas de las medidas contemporáneas están plagadas de error de medición. En este sentido, DeVellis mantiene con razón que «Incluso si una medida pobre es la única disponible, los costes de usarla pueden ser mayores que cualesquiera beneficios que se obtengan»²⁰⁰.

¹⁹⁹ HOX, 2008: 398.

²⁰⁰ DEVELLIS, 2003: 12.

BIBLIOGRAFÍA

- AGRESTI, A. 2002. *Categorical data analysis*, 2.^a ed. Hoboken, NJ: John Wiley and sons.
- 2010. *Analysis of ordinal categorical data*, 2.^a ed. Hoboken, NJ: John Wiley and sons.
- ALLISON, P. D. 2002. *Missing data*. Thousand Oaks, Ca. [etc.]: Sage.
- ALWIN, D. F. 1992. «Information transmission in the survey interview: number of response categories and the reliability of attitude measurement». *Sociological Methodology*, 22, 83-118.
- 2007. *Margins of error. A study of reliability in survey measurement*. Hoboken, NJ: John Wiley and sons.
- ANDERSON, C. y J. K. VERMUNT. 2000. «Log-multiplicative association models as latent variable models for nominal and/or ordinal data». *Sociological Methodology*, 30, 81-122.
- BASILEVSKY, A. 1994. *Statistical factor analysis and related methods. Theory and applications*. New York [etc.]: John Wiley and sons.
- BIEMER, P. P. 2001. Apuntes de cátedra. Inéditos.
- 2004. «Modeling measurement error to identify flawed questions». Pp. 225-246 en *Methods for testing and evaluating survey questionnaires* (S. PRESSER et al. eds.). Hoboken, NJ: John Wiley and sons.
- 2011. *Latent class analysis of survey error*. Hoboken, NJ: John Wiley and sons.
- BIEMER, P. P. y C. L. CHRIST. 2008. «Weighting survey data». Pp. 317-341 en *International Handbook of survey methodology* (E. D. DE LEEUW et al. eds.). New York y London: Lawrence Erlbaum Associates.
- BILLIET, J. y G. LOOSVELDT. 1988. «Improvement of the quality of responses to factual survey questions by interviewer training». *Public Opinion Quarterly*, 52, 190-211.
- BLUMSTEIN, A., M. TONRY y A. VAN NESS. 2005. «Cross-national measures of punitiveness». *Crime and Justice, 33 - Crime and punishment in Western countries, 1980-1999*, 347-376.
- BÖCKENHOLT, U. 2002. «Comparison and choice. Analyzing discrete preference data by latent class scaling models». Pp. 163-182 en *Applied latent class analysis* (J. A. HAGENAARS y A. L. MCCUTCHEON eds.). Cambridge [etc.]: Cambridge University Press.
- BOLLEN, K. A. 1989. *Structural equations with latent variables*. New York [etc.]: John Wiley and sons.

- BRAITHWAITE, J. 1979. *Inequality, crime, and public policy*. London [etc.]: Routledge and Kegan Paul.
- BRILLON, Y. 1988. «Punitiveness, status and ideology in three Canadian provinces». Pp. 84-110 en *Public attitudes to sentencing. Surveys from five countries* (N. WALKER y M. HOUGH eds.). Aldeshot [etc.]: Gower.
- BRITT, C. L. 1994. «Versatility». Pp. 173-192 en *The generality of deviance* (T. HIRSCHI y M. R. GOTTFREDSON eds.). New Brunswick, NJ y London: Transaction Publishers.
- BROWN, E. K. 2006. «The dog that did not bark: punitive social views and the “professional middle classes”». *Punishment and Society*, 2006, 287-312.
- CALLANAN, V. J. 2005. *Feeding the fear of crime. Crime-related media and support of three strikes*. New York: LFB.
- CAMPANELLI, P. 2008. «Testing survey questions». Pp. 176-200 en *International Handbook of survey methodology* (E. D. DE LEEUW et al. eds.). New York y London: Lawrence Erlbaum Associates.
- CESARONI, C. y A. N. DOOB. 2003. «The decline in support for penal welfarism. Evidence of support among the elite for punitive segregation». *British Journal of Criminology*, 43, 434-441.
- CLARKE, H. D. y A. L. MCCUTCHEON. 2009. «The dynamics of party identification reconsidered». *Public Opinion Quarterly*, 73, 704-728.
- CLOGG, C. C. y W. D. MANNING. 1996. «Assessing reliability of categorical measurements using latent class models». Pp. 169-182 en *Categorical variables in developmental research. Methods and analysis* (A. VON EYE y C. C. CLOGG eds.). San Diego, Ca. [etc.]: Academic Press.
- COCHRAN, W. G. 1977. *Sampling techniques*, 3.^a ed. New York [etc.]: John Wiley and sons.
- COLLINS, L. M. y S. T. LANZA. 2010. *Latent class and latent transition analysis with applications in the social, behavioral, and health sciences*. Hoboken, NJ: John Wiley and sons.
- CONRAD, F. G., M. F. SCHOBER y W. DIJKSTRA. 2008. «Cues of communication difficulty in telephone interviews». Pp. 212-230 en *Advances in telephone survey methodology* (J. M. LEPKOWSKI et al. eds.). Hoboken, NJ: John Wiley and sons.
- CROON, M. 2002. «Ordering the classes». Pp. 137-162 en *Applied latent class analysis* (J. A. HAGENAARS y A. L. MCCUTCHEON eds.). Cambridge [etc.]: Cambridge University Press.

- CULLEN, F. T., B. S. FISHER y B. K. APPLGATE. 2000. «Public opinion about punishment and corrections». *Crime and Justice*, 27, 1-79.
- CULLEN, F. T., J. D. UNNEVER, K. R. BLEVINS, J. A. PEALER, S. A. SANTANA, B. S. FISHER y B. K. APPLGATE. 2009. «The myth of public support for capital punishment». Pp. 73-95 en *Public opinion and criminal justice* (J. WOOD y T. GANNON eds.). Cullompton: Willan.
- DAYTON, C. M. 1998. *Latent class scaling analysis*. Thousand Oaks, Ca. [etc.]: Sage.
- DEMING, W. E. [1960] 1990. *Sample design in business research*. New York [etc.]: John Wiley and sons.
- DEVELLIS, R. F. 2003. *Scale development. Theory and applications*, 2.^a ed. Thousand Oaks, Ca. [etc.]: Sage.
- DEVILLE, J.-C. 1991. «A theory of quota surveys». *Survey Methodology*, 17, 163-181.
- DOHERTY, M. 1994. «Probability versus non-probability sampling in sample surveys». *The New Zealand Statistics Review*, marzo, 21-28.
- D'UNGER, A. V., K. C. LAND, P. MCCALL y D. S. NAGIN. 1998. «How many latent classes of delinquent/criminal careers? Results from mixed poisson regression analysis of the London, Philadelphia, and Racine cohorts studies». *American Journal of Sociology*, 103, 1593-1630.
- DUNTEMAN, G. H. 1989. *Principal components analysis*. Newbury Park, Ca. [etc.]: Sage.
- EMBRETSON, S. E. y S. P. REISE. 2000. *Item response theory for psychologists*. Mahwah, NJ y London: Lawrence Erlbaum Associates.
- ENDERS, C. K. 2006. «Analyzing structural equation models with missing data». Pp. 315-344 en *Structural equation modeling. A second course* (G. R. HANCOCK y R. O. MUELLER eds.). Greenwich, CT: IAP.
- FERNÁNDEZ MOLINA, E. y P. TARANCÓN GÓMEZ. 2010. «Populismo punitivo y delincuencia juvenil: mito o realidad». *Revista Electrónica de Ciencia penal y Criminología*, 12-08, 1-25.
- FINNEY, S. J. y C. DISTEFANO. 2006. «Nonnormal and categorical data in structural equation modeling». Pp. 269-314 en *Structural equation modeling. A second course* (G. R. HANCOCK y R. O. MUELLER eds.). Greenwich, CT: IAP.
- FORSYTH, B., J. M. ROTHGEB y G. B. WILLIS. 2004. «Does pretesting make a difference? An experimental test». Pp. 525-546 en *Methods for testing and evaluating survey questionnaires* (S. PRESSER et al. eds.). Hoboken, NJ: John Wiley and sons.

- FOWLER, F. J. 1995. *Improving survey questions. Design and evaluation*. Thousand Oaks, Ca. [etc.]: Sage.
- 2002. *Survey research methods*, 3.^a ed. Thousand Oaks, Ca. [etc.]: Sage.
- FOWLER, F. J. y T. W. MANGIONE. 1990. *Standardized survey interviewing. Minimizing interviewer-related error*. Newbury Park [etc.]: Sage.
- GANNINGER, M. 2011. *By pure chance. The merits of probability sampling*. Conferencia pronunciada en el RECSM, Barcelona, inédita.
- GOODMAN, L. A. 2002. «Latent class analysis: the empirical study of latent types, latent variables, and latent structures». Pp. 3-55 en *Applied latent class analysis* (J. A. HAGENAARS y A. L. MCCUTCHEON eds.). Cambridge [etc.]: Cambridge University Press.
- GONZÁLEZ SÁNCHEZ, I. 2010. «El modo telefónico en las encuestas de victimización» *Revista de Derecho penal y Criminología*, 4, 249-269.
- GROVES, R. M. y M. P. COUPER. 1998. *Nonresponse in household interview surveys*. New York [etc.]: John Wiley and sons.
- GROVES, R. M., F. J. FOWLER, M. P. COUPER, J. M. LEPKOWSKI, E. SINGER y R. TOURANGEAU. 2004. *Survey methodology*. Hoboken, NJ: John Wiley and sons.
- HAGENAARS, J. A. 1990. *Categorical longitudinal data: log-linear panel, trend, and cohort*. Newbury Park, Ca.: Sage.
- HANSEN, S. E. y M. P. COUPER. 2004. «Usability testing to evaluate computer-assisted instruments». Pp. 337-360 en *Methods for testing and evaluating survey questionnaires* (S. PRESSER et al. eds.). Hoboken, NJ: John Wiley and sons.
- HARKNESS, J., B.-E. PENNELL y A. SCHOUA-GLUSBERG. 2004. «Survey questionnaire translation and assessment». Pp. 453-473 en *Methods for testing and evaluating survey questionnaires* (S. PRESSER et al. eds.). Hoboken, NJ: John Wiley and sons.
- HARTNAGEL, T. F. y L. TEMPLETON. 2008. «Perceptions, emotions and experiences of crime: effects on attitudes toward punishment in a Canadian sample». Pp. 349-370 en *Fear of crime – Punitivity. New developments in theory and research* (H. KURY ed.). Bochum: Universitätsverlag Dr. N. Brockmeyer.
- HATTIE, J. 1984. «An empirical study of various indices for determining unidimensionality». *Multivariate Behavioral Research*, 19, 49-78.
- 1985. «Methodology review: assessing unidimensionality of tests and items». *Psychological Measurement*, 9, 139-164.

- VAN DER HEIJDEN, P. G. M., L. A. VAN DER ARK y A. MOOIJAART. 2002. «Some examples of latent budget analysis and its extensions». Pp. 107-136 en *Applied latent class analysis* (J. A. HAGENAARS y A. L. MCCUTCHEON eds.). Cambridge [etc.]: Cambridge University Press.
- HEINEN, T. 1996. *Latent class and discrete latent trait models. Similarities and differences*. Thousand Oaks, Ca. [etc.]: Sage.
- HOX, J. J. 2008. «Accommodating measurement errors». Pp. 387-402 en *International Handbook of survey methodology* (E. D. DE LEEUW et al. eds.). New York y London: Lawrence Erlbaum Associates.
- KALSBECK, W. D. y R. P. AGANS. 2008. «Sampling and weighting in household telephone surveys». Pp. 29-55 en *Advances in telephone survey methodology* (J. M. LEPKOWSKI et al. eds.). Hoboken, NJ: John Wiley and sons.
- KAPLAN, D. 2009. *Structural equation modeling. Foundations and extensions*, 2.^a ed. Los Angeles, Ca.: Sage.
- KEIL, T. J. y G. F. VITO. 1991. «Fear of crime and attitudes toward capital punishment: a structural equations model». *Justice Quarterly*, 8, 447-464.
- KLINE, R. B. 2005. *Principles and practice of structural equation modeling*, 2.^a ed. New York y London: The Guilford Press.
- KRUTTSCHNITT, C. y R. MACMILLAN. 2006. «The violent victimization of women. A life course perspective». Pp. 139-170 en *Gender and crime. Patterns of victimization and offending* (K. HEIMER y C. KRUTTSCHNITT eds.). New York y London: New York University Press.
- KUHN, A. 1993. «Attitudes towards punishment». Pp. 271-288 en *Understanding crime experiences of crime and crime control* (A. ALVAZI DEL FRATE et al. eds.). Roma: UNICRI.
- 2002. «Public and judicial attitudes to punishment in Switzerland». Pp. 115-127 en *Changing attitudes to punishment. Public opinion, crime and justice* (J. V. ROBERTS y M. HOUGH eds.). Cullompton: Willan Publishing.
- KURY, H. y J. OBERGFELL-FUCHS. 2008. «Methodological problems in measuring attitudes to punishment (punitivity)». Pp. 277-302 en *Fear of crime – Punitivity. New developments in theory and research* (H. KURY ed.). Bochum: Universitätsverlag Dr. N. Brockmeyer.
- KURY, H. y E. PUTKARADZE. 2009. «Attitudes to punishment in Georgia. Results from a student survey». Ponencia presentada al Congreso Anual de la *European Society of Criminology*. Inédito.
- KURY, H., U. DÖRMANN, H. RICHTER y M. WÜRGER. 1996. *Opfererfah-*

- rungen und Meinungen zur Inneren Sicherheit in Deutschland. Ein empirischer Vergleich von Viktimisierungen, Anzeigeverhalten und Sicherheitseinschätzung in Ost und West vor der Vereinigung.* Wiesbaden: Bundeskriminalamt.
- KURY, H., J. OBERGFELL-FUCHS y M. WÜRGER. 2000. *Kriminalität und Einstellung. Ein Vergleich zwischen Ost- und Westdeutschland.* Freiburg i.Br.: MPI.
- 2002. *Strafeinstellungen. Ein Vergleich zwischen Ost- und Westdeutschland.* Freiburg i.Br.: MPI.
- KURY, H., J. OBERGFELL-FUCHS y U. SMARTT. 2002. «The evolution of public attitudes to punishment in Western and Eastern Europe». Pp. 93-114 en *Changing attitudes to punishment. Public opinion, crime and justice* (J. V. ROBERTS y M. HOUGH eds.). Cullompton: Willan Publishing.
- KURY, H., H. KANIA y J. OBERGFELL-FUCHS. 2004. «Worüber sprechen wir, wenn wir über Punitivität sprechen? Versuch einer konzeptionellen und empirischen Begriffsbestimmung». *Kriminologisches Journal*, 36, 51-88.
- KURY, H., T. N. FERDINAND y J. OBERGFELL-FUCHS. 2008. «Punitivity in Germany: attitudes to punishment, sentencing, and prison rates». Pp. 107-137 en *International perspectives on punitivity* (H. KURY y T. N. FERDINAND eds.). Bochum: Universitätsverlag Brockmeyer.
- KUTATELADZE, B. 2011. «Measuring State punitiveness in the United States». Pp. 151-179 en *Punitivity. International developments, 1 - Punitiveness – a global phenomenon?* (H. KURY y E. SHEA eds.). Bochum: Universitätsverlag Dr. Brockmeyer.
- KUUSELA, V., M. CALLEGARO y V. VEHOVAR. 2008. «The influence of mobile telephones on telephone surveys». Pp. 87-112 en *Advances in telephone survey methodology* (J. M. LEPKOWSKI et al. eds.). Hoboken, NJ: John Wiley and sons.
- LANGEHEIME, R., J. PANNEKOEK y F. VAN DE POL. 1996. «Bootstrapping goodness-of-fit measures in categorical data analysis». *Sociological Methods and Research*, 24, 492-516.
- LAVRAKAS, P. J. 1993. *Telephone survey methods. Sampling, selection, and supervision*, 2.^a ed. Newbury Park, Ca. [etc.]: Sage.
- LESSLER, J. T., J. EYERMAN y K. WANG. 2008. «Interviewer training». Pp. 442-460 en *International Handbook of survey methodology* (E. D. DE LEUW et al. eds.). New York y London: Lawrence Erlbaum Associates.
- LI, W. y D. R. NYHOLT. 2001. «Marker selection by Akaike information criterion and Bayesian information criterion». *Genetic Epidemiology*, 21, Suplemento 1, S272-S277.

- LONG, J. S. 1997. *Regression models for categorical and limited dependent variables*. Thousand Oaks, Ca. [etc.]: Sage.
- MAYHEW, P. y J. VAN KESTEREN. 2002. «Cross-national attitudes to punishment». Pp. 63-92 en *Changing attitudes to punishment. Public opinion, crime and justice* (J. V. ROBERTS y M. HOUGH eds.). Cullompton: Willan Publishing.
- MCCUTCHEON, A. L. 1987. *Latent class analysis*. Newbury Park, Ca. [etc.]: Sage.
- 2002. «Basic concepts and procedures in single- and multiple-group latent class analysis». Pp. 56-85 en *Applied latent class analysis* (J. A. HAGENAARS y A. L. MCCUTCHEON eds.). Cambridge [etc.]: Cambridge University Press.
- MCCUTCHEON, A. L. y J. A. HAGENAARS. 1997. «Comparative social research with multi-sample latent class models». Pp. 266-277 en *Applications of latent trait and latent class models in the social sciences* (J. ROST y R. LANGEHEINE eds.). Münster [etc.]: Waxman.
- MCCUTCHEON, A. L. y C. MILLS. 1998. «Categorical data analysis: log-linear and latent class models». Pp. 71-94 en *Research strategies in the social sciences. A guide to new approaches* (E. SCARBROUGH y E. TANENBAUM eds.). Oxford: Oxford University Press.
- MUTHÉN, L. K. y B. O. MUTHÉN. 2010. *Mplus user's guide*, 6.^a ed. Los Angeles, Ca.: Muthén and Muthén.
- NAGIN, D. S. 2005. *Group-based modeling of development*. Cambridge, Mass. y London: Harvard University Press.
- NELLIS, A. M. y J. P. LYNCH. 2008. «Crime, fear, and the demand for punishment in the United States». Pp. 33-54 en *International perspectives on punitivity* (H. KURY y T. N. FERDINAND eds.). Bochum: Universitätsverlag Brockmeyer.
- POWERS, D. A. y Y. XIE. 2000. *Statistical methods for categorical data analysis*. San Diego, Ca. [etc.]: Academic Press.
- RABE-HESKETH, S. y A. SKRONDAL. 2001. «Parameterization of multivariate random effects models for categorical data». *Biometrics*, 57, 1256-1264.
- REEVE, B. B. y L. C. MÄSSE. 2004. «Item response theory modeling of questionnaire evaluation». Pp. 247-273 en *Methods for testing and evaluating survey questionnaires* (S. PRESSER et al. eds.). Hoboken, NJ: John Wiley and sons.
- RIEDEL, M. 2000. *Research strategies for secondary data. A perspective for Criminology and Criminal Justice*. Thousand Oaks, Ca.: Sage.

- ROBBERS, M. 2006. «Tough-mindedness and fair play: personality traits as predictors of attitudes toward the death penalty – an exploratory gendered study». *Punishment and Society*, 8, 203-222.
- ROHNE, H.-C. 2008. «Conceptualizing punitiveness from a victims' perspective – Findings in the context of the Al-Aqsa intifada». Pp. 161-186 en *International perspectives on punitivity* (H. KURY y T. N. FERDINAND eds.). Bochum: Universitätsverlag Brockmeyer.
- RUDAS, T. 2002. «A latent class approach to measuring the fit of a statistical model». Pp. 345-365 en *Applied latent class analysis* (J. A. HAGENAARS y A. L. MCCUTCHEON eds.). Cambridge [etc.]: Cambridge University Press.
- SACCO, P., K. K. BUCHOLZ, E. L. SPITZNAGEL. 2009. «Alcohol use among older adults in the National Epidemiologic Survey on alcohol and related conditions: a latent class analysis». *Journal of Studies on Alcohol and Drugs*, 70, 829-838.
- SECRET, P. E. y J. B. JOHNSON. 1989. «Racial differences in attitudes toward crime control». *Journal of Criminal Justice*, 17, 361-375.
- SERRANO MAÍLLO, A. 2006. «Punitivität und Gesetzgebung: Die Situation in Spanien». Pp. 245-252 en *Kriminalität und Kriminalprävention in Ländern des Umbruchs* (H. KURY y E. KARIMOV eds.). Bochum: Universitätsverlag Brockmeyer.
- 2008. «El problema de los *datos perdidos* y su tratamiento en los análisis cuantitativos en ciencias sociales. La imputación múltiple». Pp. 157-169 en *Procesos de infracción de normas y de reacción a la infracción de normas: dos tradiciones criminológicas. Nuevos estudios en Homenaje al Profesor Alfonso Serrano Gómez* (A. SERRANO MAÍLLO y J. L. GUZMÁN DÁLBORA eds.). Madrid: Dykinson.
- 2011. «Actitudes sobre Derechos fundamentales procesal-penales: una dimensión inexplorada de la punitividad. Un análisis de clases latentes». *Revista Electrónica de Ciencias penales y Criminología*, 13-05, 1-38.
- SERRANO MAÍLLO, A. y H. KURY. 2008. «Insecurity feelings and punitivity: relationship in a national sample of adolescents and young adults in Spain». Pp. 321-348 en *Fear of crime – Punitivity. New developments in theory and research* (H. KURY ed.). Bochum: Universitätsverlag Brockmeyer.
- SKRONDAL, A. y S. RABE-HESKETH. 2004. *Generalized latent variable modeling. Multilevel, longitudinal, and structural equation models*. Boca Raton, FL [etc.]: Chapman and Hall/CRC.
- 2007. «Latent variable modelling: a survey». *Scandinavian Journal of Statistics*, 34, 712-745.

- SPECTOR, P. E. 1992. *Summated rating scale construction. An introduction*. Newbury Park, Ca. [etc.]: Sage.
- SPROTT, J. B. 1999a. *Views of the punishment of youth: the dimensions of punitiveness*. Ann Arbor, Mi.: UMI.
- 1999b. «Are members of the public tough on crime?: the dimensions of public “punitiveness”». *Journal of Criminal Justice*, 27, 467-474.
- STALANS, L. J. 2002. «Measuring attitudes to sentencing». Pp. 15-32 en *Changing attitudes to punishment. Public opinion, crime and justice* (J. V. ROBERTS y M. HOUGH eds.). Cullompton: Willan Publishing.
- STEEH, C. 2008. «Telephone surveys». Pp. 221-238 en *International Handbook of survey methodology* (E. D. DE LEEUW et al. eds.). New York y London: Lawrence Erlbaum Associates.
- STEVE, K. W., A. T. BURKS, P. J. LAVRAKAS, K. D. BROWN y J. B. HOOVER. 2008. «Monitoring telephone interviewer performance». Pp. 401-422 en *Advances in telephone survey methodology* (J. M. LEPKOWSKI et al. eds.). Hoboken, NJ: John Wiley and sons.
- TARNAI, J. y D. L. MOORE. 2008. «Measuring and improving telephone interviewer performance and productivity». Pp. 359-384 en *Advances in telephone survey methodology* (J. M. LEPKOWSKI et al. eds.). Hoboken, NJ: John Wiley and sons.
- TIMBERLAKE, T. S. 2008. «A latent class analysis of nicotine-dependence criteria and use of alternate tobacco». *Journal of Studies on Alcohol and Drugs*, 69, 709-717.
- TONRY, M. y D. P. FARRINGTON. 2005. «Punishment and crime across space and time». *Crime and Justice*, 33 – *Crime and punishment in Western countries, 1980-1999*, 1-39.
- TUCKER, C. y J. M. LEPKOWSKI. 2008. «Telephone survey methods: adapting to change». Pp. 3-26 en *Advances in telephone survey methodology* (J. M. LEPKOWSKI et al. eds.). Hoboken, NJ: John Wiley and sons.
- TYLER, T. y R. WEBER. 1982. «Support for the death penalty; instrumental response to crime, or symbolic attitude?». *Law and Society Review*, 17, 21-45.
- UEBERSAX, J. 1993. «Dimension reduction and latent class analysis. A simple method for interpretation of latent class analysis parameters, with possible implications for factor analysis of dichotomous and ordered-category measures». Ponencia presentada al Congreso Anual de la *Classification Society of North America*. Inédito.

- 1997. «Analysis of student problem behaviors with latent trait, latent class, and related probit mixture models». Pp. 188-195 *Applications of latent trait and latent class models in the social Sciences* (J. ROST y R. LANGEHEINE eds.). New York: Waxman.
- 1999. «Probit latent class analysis with dichotomous or ordered category measures: conditional independence/dependence models». *Applied Psychological Measurement*, 23, 283-297.
- 2009. *A practical guide to conditional dependence in latent class analysis*, última versión. Accesible en <<http://www.john-uebersax.com/stat/condep.htm>>.
- UEBERSAX, J. y W. M. GROVE. 1990. «Latent class analysis of diagnostic agreement». *Statistics in Medicine*, 9, 559-572.
- VARONA GÓMEZ, D. 2008. «Ciudadanos y actitudes punitivas: un estudio piloto de población universitaria española». *Revista Española de Investigación Criminológica*, 6-1, 1-38.
- 2009. «¿Somos los españoles punitivos?: actitudes punitivas y reforma penal en España». *InDret*, 1/2009, 1-31.
- VAUGHN, M. G., M. DELISI, K. M. BEAVER y J. P. WRIGHT. 2009. «Identifying latent classes of behavioral risk based on early childhood manifestations of self-control». *Youth Violence and Juvenile Justice*, 7, 16-31.
- VERMUNT, J. K. 1997a. *Log-linear models for event histories*. Thousand Oaks, Ca. [etc.]: Sage.
- 1997b. *LEM: a general program for the analysis of categorical data*. Tilburg: Department of Methodology and Statistics, Tilburg University.
- 1999. «A general class of non-parametric models for ordinal categorical data». *Sociological Methodology*, 29, 187-223.
- 2003. «Applications of latent class analysis in social science research». *Lecture Notes in Artificial Intelligence*, 2711, 22-36.
- VERMUNT, J. K. y J. MAGIDSON. 2002. «Latent class cluster analysis». Pp. 89-106 en *Applied latent class analysis* (J. A. HAGENAARS y A. L. McCUTCHEON eds.). Cambridge [etc.]: Cambridge University Press.
- 2003. «Latent class models for classification». *Computational Statistics and Data Analysis*, 41, 531-537.
- 2007. «Latent class analysis with sampling weights. A maximum likelihood approach». *Sociological Methods and Research*, 36, 87-111.

- VERMUNT, J. K., J. R. VAN GINKEL, L. A. VAN DER ARK y K. SIJTSMA. 2008. «Multiple imputation of incomplete categorical data using latent class analysis». *Sociological Methodology*, 33, 369-297.
- VIKI, G. T. y G. BOHNER. 2009. «Achieving accurate assessment of attitudes toward the criminal justice system: methodological issues». Pp. 96-119 en *Public opinion and criminal justice* (J. WOOD y T. GANNON eds.). Cullompton: Willan.
- WALKER, N., M. HOUGH y H. LEWIS. 1988. «Tolerance of leniency and severity in England and Wales». Pp. 178-202 en *Public attitudes to sentencing. Surveys from five countries* (N. WALKER y M. HOUGH eds.). Aldeshot [etc.]: Gower.
- WINSHIP, C., R. D. MARE y J. R. WARREN. 2002. «Latent class models for contingency tables with missing data». Pp. 408-432 en *Applied latent class analysis* (J. A. HAGENAARS y A. L. MCCUTCHEON eds.). Cambridge [etc.]: Cambridge University Press.
- WRIGHT, B. R. E., A. CASPI, T. E. MOFFITT y R. PATERNOSTER. 2004. «Does the perceived risk of punishment deter criminally prone individuals? Rational choice, self-control, and crime». *Journal of Research in Crime and Delinquency*, 41, 180-213.