Psicothema 2008. Vol. 20, nº 4, pp. 830-838
www.psicothema.com

ISSN 0214 - 9915 CODEN PSOTEG

METODOLOGÍA

Construcción de modelos jerárquicos en contextos aplicados

Guillermo Vallejo Seco, Jaime Arnau Grass* y Roser Bono Cabré*
Universidad de Oviedo y * Universidad de Barcelona

Los modelos lineales jerárquicos se han convertido en una herramienta muy popular para analizar datos que presentan una estructura jerarquizada. Esta metodología reconoce la estructura anidada de los datos y permiten obtener estimaciones insesgadas de las variaciones acaecidas en los distintos niveles de la jerarquía. El objetivo de este artículo es ilustrar la construcción de modelos jerárquicos en contextos transversales y longitudinales implicando tres y cuatro niveles, respectivamente. Para mostrar el proceso de modelado estadístico se evalúa la eficacia de un programa de intervención diseñado para incrementar el rendimiento matemático de estudiantes de Primaria. El ejemplo empleado se analiza con los paquetes estadísticos SAS y SPSS, cuya sintaxis se detalla convenientemente en el manuscrito.

Construction of hierarchical models in applied contexts. Hierarchical linear models have become a very popular tool for analyzing data with a hierarchical structure. This methodology recognizes the nested structure of the data and allows obtaining unbiased estimates of the variations found in the different levels of the hierarchy. The goal of this article is to illustrate the construction of hierarchical models both in cross-sectional and longitudinal contexts involving three and four levels, respectively. The efficiency of an intervention program designed to improve the mathematical performance of primary school students is evaluated to illustrate the statistical modelling process. The example used is analyzed by the SAS and SPSS packages, whose syntax is duly detailed in the manuscript.

Los programas de tratamiento destinados a mejorar la calidad de vida de los ciudadanos se aplican comúnmente en estudios donde las unidades de muestreo están agrupadas de forma natural tales como colegios, empresas, hospitales u otras instituciones (Oliver, Rosel y Jara, 2000). En definitiva, a sistemas organizacionales caracterizados por poseer una estructura de anidamiento, en los cuales las unidades de observación de un determinado nivel se hallan agrupadas en otras de un nivel jerárquico superior. Por ejemplo, los residentes de una determinada ciudad están agrupados en barrios, los empleados de una empresa en departamentos, los pacientes de un hospital en plantas o los estudiantes de un centro escolar en clases. En la última situación, los estudiantes constituyen el nivel inferior de la jerarquía, las clases el nivel intermedio y los centros escolares el nivel superior. Estos agrupamientos, además de representar una estructura específica de los datos, constituyen unidades de análisis que tienen interés por sí mismas, y ello debido a la importancia que tiene considerar las variables medidas en los distintos niveles de la jerarquía para entender las ejecuciones de los participantes en el estudio. De ahí que, como ha destacado Goldstein (2002), el término multinivel se utilice tanto para designar los niveles en que se pueden agrupar los datos, como para caracterizar a las técnicas utilizadas para modelar las relaciones que se dan dentro y a través de los diferentes niveles.

A pesar de que los datos obtenidos a partir de un sistema organizacional, principalmente educativo, social o sanitario, presentan una estructura jerarquizada, la mayor parte de las veces los datos son registrados en el nivel más bajo de la jerarquía y analizados mediante técnicas estadísticas basadas en el modelo lineal clásico. Por ejemplo, en el terreno educativo, resulta frecuente evaluar el rendimiento alcanzado con diferentes métodos de enseñanza utilizando el análisis de la regresión múltiple. Si se pasa por alto la posible falta de independencia de este tipo de datos, la regresión clásica constituye una buena opción analítica para poder responder al objetivo planteado. Con todo, conviene tener presente que si un programa resulta efectivo, la técnica de la regresión no indica ni dónde ni cuándo tiene su efecto óptimo. Mientras que si un programa no surte los efectos deseados, la técnica tampoco permite explorar qué otras variables (individuales, grupales o contextuales) están controlando la conducta, ni conocer si la relación entre el

Fecha recepción: 15-11-07 • Fecha aceptación: 20-2-08
Correspondencia: Guillermo Vallejo Seco
Facultad de Psicología
Universidad de Oviedo
33003 Oviedo (Spain)
E-mail: gvallejo@uniovi.es

programa y la respuesta está siendo afectada por la acción moderadora (interacción) de alguna característica de los diferentes niveles de la jerarquía.

Un enfoque que permite abordar las cuestiones reseñadas es el modelo lineal mixto, del cual los modelos jerárquicos o multinivel son casos especiales (Raudenbush y Bryk, 2002). Además de los trabajos citados a lo largo del manuscrito, una referencia recomendable sobre análisis multinivel es Hox (2002). En español, también puede resultar de utilidad el texto de Ato y Vallejo (2007). El modelo lineal mixto, o el modelo lineal mixto generalizado si los datos carecen de una métrica cuantitativa, reconoce la estructura anidada de los datos y permiten estimar las variaciones acaecidas en los distintos niveles de la jerarquía. Para entender mejor lo dicho, supongamos que interese saber ¿cómo afectan al rendimiento académico el nivel socioeconómico de los estudiantes y el tipo de escuela? De acuerdo con Snijders y Bosker (2004), esta hipotética relación se puede expresar como:

$$Y_{ij}= \beta_0+\beta_1 X_{ij}+\beta_2 Z_j+e_{ij} \quad (1)$$

donde Y_{ij} es el valor de la respuesta dada por el i estudiante dentro de la j escuela, β_0 es el intercepto, β_1 es el coeficiente asociado a la variable nivel económico, β_2 es el coeficiente asociado a la variable tipo de escuela y e_{ij} es el término de error residual.

El modelo de regresión múltiple de la ecuación (1) no recoge adecuadamente la estructura jerárquica de los datos, pues si bien es cierto que tiene en cuenta las variables explicativas, no es menos cierto que obvia la posibilidad de que los coeficientes asociados con ellas varíen de una escuela a otra; es decir, el hecho de que los parámetros β_0 y β_1 no sean independientes del grupo al que pertenecen los escolares. Esta relación de anidación se especifica añadiendo un subíndice extra a estos coeficientes, es decir:

$$Y_{ij}= \beta_{0j}+\beta_{1j} X_{ij}+\beta_2 Z_j+e_{ij} \quad (2)$$

Nótese la diferencia con el modelo (1). En la ecuación 2, cada escuela tiene un valor dado de β_{0j} y β_{1j}. Ya no puede decirse que no importa cuál sea el valor de Z. De este modo, es posible que una determinada escuela j muestre un nivel alto o bajo de β_{0j}. Esto significa que los escolares que pertenecen a una escuela parten de un nivel de desempeño que puede ser alto o bajo. De igual modo, el grado en que el nivel económico explica el rendimiento depende también del tipo de escuela. Así, por ejemplo, si un grupo escolar tiene un valor alto de β_{1j} cabe esperar que el efecto de X sobre Y resulte superior al de otro que lo tenga inferior. Por último, β_2 queda inalterado por ser Z una variable grupal.

Las consecuencias de no prestar atención a la estructura jerárquica de los datos son múltiples, pero tan sólo destacaremos las dos más importantes, a saber: el problema de la unidad de análisis y el denominado sesgo de agregación. El primer problema surge cuando unidades de análisis colectivas configuradas de antemano, más que unidades de análisis individuales, constituyen el referente observacional al que va dirigido el programa de intervención. Por regla general, en estos casos el investigador selecciona al azar un número «X» de unidades que asigna al programa, mientras que el número «Z» restante sirve de control. Ahora bien, para aplicar técnicas basadas en el modelo lineal clásico se requiere satisfacer ciertos supuestos, especialmente el que alude a la falta de relación entre las observaciones. Sin embargo, la independencia estocástica resultará cuestionable cuando lo que se asigna al programa son unidades experimentales colectivas, en lugar de unidades experimentales individuales. Por diversas razones, que ahora no vienen al caso, es altamente probable que los datos extraídos a partir de agrupaciones naturales mantengan un cierto grado de parecido entre sí, lo cual puede afectar a la precisión de las estimaciones y la amplitud de las inferencias (para detalles véase Vallejo, Fernández y Secades, 2006).

El segundo problema también se deriva del empleo de unidades de investigación colectivas. En concreto, el denominado sesgo de agregación se presenta cuando se obtienen resultados diferentes para una misma variable dependiendo de cuál sea el nivel de análisis adoptado y, por ende, la unidad de análisis utilizada. Por ejemplo, puede ocurrir que una variable medida en el nivel más alto de la jerarquía tenga un efecto superior, e inclusive más duradero, que la misma variable medida en el nivel más bajo. Por lo tanto, de utilizarse el modelo de regresión clásico o uninivel surgirían problemas interpretativos; pues los coeficientes estimados serían diferentes dependiendo de que el análisis se efectuase en el nivel superior utilizando medidas de respuesta grupal, o en el nivel inferior utilizando medidas de respuesta individual. Dicho lo anterior, la pregunta que surge de inmediato es la siguiente: ¿resulta apropiado ignorar la variabilidad individual agregando las medidas dentro de los grupos para garantizar la independencia de las observaciones? Al igual que Krull y MacKinnon (2001), pensamos que esta práctica constituye más un problema que una solución, ya que se reduce la potencia de las pruebas estadísticas y se olvida que la interdependencia entre los individuos tiene interés por derecho propio.

Los modelos jerárquicos ayudan a resolver ambos problemas. Pues además de estimar los componentes de varianza correspondientes a las múltiples fuentes de variación aleatoria existentes, también evitan la confusión derivada del sesgo de agregación formulando modelos e hipótesis para procesos y relaciones dentro y entre los niveles. Como será puesto de manifiesto en las secciones siguientes, la técnica capacita al investigador para formalizar la estructura anidada de sus datos de una manera compacta. Ello requiere especificar un modelo adecuado para cada uno de los niveles en que se hallen agrupados los datos y concatenarlos convenientemente. Al combinar la información obtenida a través de los diferentes niveles, la técnica permite examinar simultáneamente los efectos de cada una de las variables presentes en la investigación, así como las interacciones entre las variables de un mismo nivel y de niveles diferentes.

Construcción del modelo jerárquico

Básicamente, la construcción del modelo jerárquico conlleva dos acciones diferentes: desarrollar un modelo estadístico que permita responder a los objetivos que motivaron el estudio e implementar dicho modelo dentro de un programa de ordenador para analizar los datos. Por lo que se refiere a la primera de las dos acciones reseñadas aquí, inicialmente se procede desarrollando un modelo para las unidades de menor tamaño, luego otro para las unidades de tamaño inmediatamente superior, y así sucesivamente. Por regla general, las unidades de tamaño más pequeño se consideran muestras al azar de una población de unidades de tamaño mayor. Por consiguiente, sus efectos son tratados como aleatorios. Cuando las condiciones de interés de una determinada variable sean incluidas en el estudio de manera arbitraria por el investigador, el factor es considerado fijo.

Para ilustrar la construcción del modelo se evaluará la eficacia de un programa destinado a potenciar el rendimiento matemático de estudiantes de Primaria. Los datos de este ejemplo simulado tomado de la investigación educativa, y disponibles desde el primer autor, se consiguieron al evaluar el desempeño de 1.200 estudiantes obtenidos tras seleccionar al azar 60 centros de enseñanza y 220 clases de una gran ciudad, con la condición de que se disponga de un mínimo de dos clases por centro y que al menos una de ellas reciba el programa de tratamiento. Por consiguiente, el programa se aplica en todos los centros, pero no en todas las clases. Además, con el fin de extender la formulación del modelo al ámbito longitudinal, todos los participantes en el diseño parcialmente anidado de grupos al azar estratificado son medidos antes y después de administrar el programa. En concreto, comprobaremos mediante sendas extensiones del análisis de covarianza de puntuaciones pretest y postest (Hedeker y Gibbons, 2006) y del análisis de varianza (ANOVA) de medidas repetidas, si el rendimiento observado en las clases que recibieron el programa resulta similar al de otras clases que no lo recibieron.

A lo largo de la estrategia de modelado, la variable dependiente es el logro predicho (medido en una escala de uno a diez) por un conjunto de variables explicativas registradas en el nivel estudiante (nivel 1), en el nivel clase (nivel 2) y en el nivel centro (nivel 3). Las variables de nivel 1 son el rendimiento previo a la presentación del programa o línea base (LB), el nivel socioeconómico de los estudiantes (NSE) en su ámbito familiar centrados con respecto a su media, el género de los mismos (GNR, hombres= 1 y mujeres= 0) y el nivel educativo de los padres (NEP, titulados= 1 y restantes= 0). Con independencia del nivel de medida utilizado en su registro, las características de los estudiantes son incluidas en el análisis con propósito de control. Además del programa propiamente dicho (PRG, las clases que reciben programa toman el valor de uno y las que no lo reciben un valor de cero), las variables de nivel 2 son la experiencia docente del profesor (EDP, desde cero hasta treinta años de servicio) y el tamaño de la clase (TCL, desde cinco participantes hasta veinte). En el nivel 3 tan sólo se incluye la titularidad del centro docente (TCD, centros de titularidad pública= 1 y centros de titularidad privada-concertada= 0). Se excusa decir que otras muchas características de los niveles uno, dos y tres podrían haber sido usadas para construir el modelo.

Desarrollo del modelo estadístico

El modelado jerárquico comienza formulando un modelo en el nivel 1. Supongamos que Y_{ijk} denote la calificación obtenida después de administrar el programa por el i estudiante dentro de la j clase anidada en el k centro escolar, entonces para los datos de este estudiante podemos escribir el modelo que sigue:

$$Y_{ijk} = \pi_{0jk} + \pi_{1jk}\text{LB} + \pi_{2jk}\text{NSE} + \pi_{3jk}\text{GNR} + \pi_{4jk}\text{NEP} + e_{ijk} \tag{3}$$

donde π_{0jk} denota el rendimiento promedio de la j clase anidada en el k centro, π_{1jk} representa el cambio promedio en el postest asociado con una unidad de cambio en el pretest, π_{2jk} señala si las diferencias entre el nivel económico de los estudiantes de la j clase anidada en el k centro afectan al rendimiento, π_{3jk} indica si el rendimiento promedio de los hombres y las mujeres de la j clase anidada en el k centro difieren entre sí, π_{4jk} especifica si el nivel educativo de los padres afecta diferencialmente al rendimiento promedio de los estudiantes de la j clase anidada en el k centro y e_{ijk} denota la diferencia entre la puntuación obtenida por el i estudiante y la media de la j clase dentro del k centro (varianza que permanece sin explicar tras controlar las variables recursos económicos, género y nivel educativo de los padres). Por simplicidad asumimos que los errores de nivel 1 son variables aleatorias que siguen una distribución normal con media cero y varianza constante a través de las aulas anidadas dentro de los centros, esto es, $e_{ijk} \sim N(0, \sigma^2)$.

A continuación, se procede a incorporar la naturaleza jerárquica de los datos en el modelo de nivel 2. Para ello asumiremos que el coeficiente π_{0jk} varía entre las clases de un mismo centro en función de una media, de las covariables de efectos fijos registradas en el nivel clase y del efecto aleatorio asociado con cada una de las clases del k centro escolar. También es asumido que los efectos de las covariables estatus, género y formación de los padres permanecen constantes a nivel grupal inferior; es decir, las covariables registradas a nivel individual no varían aleatoriamente entre las clases de un determinado centro. De acuerdo con lo dicho, el modelo referido al nivel clase se puede escribir como:

$$\begin{gathered}\pi_{0jk} = \beta_{00k} + \beta_{01k}\,\text{PRG} + \beta_{02k}\,\text{EDP} + \beta_{03k}\,\text{TCL} + u_{0jk},\\ \pi_{1jk} = \beta_{10k} + \beta_{11k}\,\text{PRG},\ \pi_{2jk} = \beta_{20k},\ \pi_{3jk} = \beta_{30k},\ \pi_{4jk} = \beta_{40k},\end{gathered} \tag{4}$$

donde β_{00k} es el rendimiento matemático promedio del k centro, β_{01k} señala si dentro del k centro el desempeño promedio de las clases que reciben tratamiento es superior al de las clases que no lo reciben, β_{02k} indica si dentro del k centro el desempeño promedio varía en función de los años que llevan los profesores impartiendo la materia objeto de estudio, β_{03k} muestra si dentro del k centro el desempeño promedio varía en función del tamaño de la clase, β_{10k} especifica si dentro del k centro el desempeño promedio de las clases que no reciben el programa varía del pretest al postest, β_{11k} informa si la relación funcional entre el pretest y el postest se ve afectada por la administración del programa y u_{0jk} denota la diferencia en el rendimiento matemático promedio de la j clase y el k centro (informa de las diferencias existentes entre las medias de las clases pertenecientes a un mismo centro en el rendimiento, cuando se controlan los efectos de las covariables registradas en el nivel 2). Se asume que el parámetro π_{0jk} de cada una de las clases del mismo centro se distribuye normalmente con media cero y varianza constante, esto es, $u_{0jk} \sim N(0, \omega_{00})$.

Finalmente, se procede a incorporar en el modelo el efecto debido a los centros y a su posible interacción con el programa. El modelo resultante adopta la forma que sigue:

$$\begin{gathered}\beta_{00k} = \gamma_{000} + \gamma_{001}\,\text{TCD} + u_{00k},\\ \beta_{01k} = \gamma_{010} + u_{01k},\ \beta_{020k} = \gamma_{020},\ \beta_{03k} = \gamma_{030},\\ \beta_{10k} = \gamma_{100},\ \beta_{11k} = \gamma_{110},\ \beta_{20k} = \gamma_{200},\ \beta_{30k} = \gamma_{300},\ \beta_{40k} = \gamma_{400},\end{gathered} \tag{5}$$

donde γ_{000} es la media de rendimiento general de todos los estudiantes, γ_{001} señala en qué medida la titularidad de los centros afecta al rendimiento promediado entre los mismos, γ_{010} indica si el rendimiento promediado a lo largo de todos los centros difiere entre las clases que recibieron el programa y las que no lo recibieron, γ_{020} denota en qué medida el rendimiento promediado de los estudiantes difiere en función de los años de experiencia de los profesores, γ_{030} indica si el rendimiento promedio difiere en función del tamaño de clase, γ_{110} informa cómo afecta el programa a la relación funcional entre el rendimiento previo y posterior promediado a través de las clases, γ_{100} especifica si rendimiento matemático promedio medido antes de intervenir predice significativamente el registrado después, γ_{200} muestra el grado de relación existente entre el nivel socioeconómico de los estudiantes y el rendimiento promedio de los mismos, γ_{300} informa si el rendimiento promedio de los hombres difiere del rendimiento promedio de las mujeres, γ_{400} advierte si el rendimiento promedio de los estudiantes es independiente del nivel educativo de los padres, u_{01k} indica si las diferencias de rendimiento promedio entre las clases que reciben el programa y las que no lo reciben permanecen estables a través de los distintos centros y u_{00k} denota la diferencia entre el rendimiento promedio del k centro y la media general.

Sustituyendo en la ecuación (3) las expresiones de las ecuaciones (4) y (5), se obtiene el modelo de efectos mixtos:

$$\begin{aligned} Y_{ijk} = {} & \gamma_{000} + \gamma_{001}\,\mathrm{TCD} + \gamma_{010}\,\mathrm{PGR} + \gamma_{020}\,\mathrm{EDP} + \gamma_{030}\,\mathrm{TCL} \\ & + \gamma_{110}\,\mathrm{LB} \times \mathrm{PGR} + \gamma_{100}\,\mathrm{LB} + \gamma_{200}\,\mathrm{NSE} + \gamma_{300}\,\mathrm{GNR} \\ & + \gamma_{400}\,\mathrm{NEP} + u_{01k} \times \mathrm{PGR} + u_{00k} + u_{0\,jk} + e_{ijk} \end{aligned} \tag{6}$$

donde $u_{01k} \times$ PGR indica en qué medida el desempeño promedio de las clases tratadas y las no tratadas depende del centro escolar en que se hallan anidadas. El modelo de la ecuación (6) resulta de sumar dos partes: una fija y otra aleatoria. Además de la media general del proceso (γ_{000}), la parte fija incluye los coeficientes individuales (γ_{100}, γ_{200}, γ_{300}, γ_{400}), los grupales de nivel inferior (γ_{010}, γ_{020}, γ_{030}) y de nivel superior (γ_{001}), que son comunes a todos los estudiantes, con independencia de la clase y centro al que pertenezcan. A su vez, la parte estocástica incluye los coeficientes que varían aleatoriamente entre grupos y entre sujetos (u_{00k}, $u_{0\,jk}$, e_{ijk}).

A fin de simplificar el modelo, la ecuación (6) puede ser escrita para el conjunto de los I estudiantes, J clases y K centros usando la notación matricial del modelo lineal mixto como:

$$\boldsymbol{y} = \boldsymbol{X}\,\boldsymbol{\beta} + \boldsymbol{Z}\,\boldsymbol{u} + \boldsymbol{e} \tag{7}$$

donde $\boldsymbol{y}$ es un vector de respuestas, $\boldsymbol{X}$ y $\boldsymbol{Z}$ son matrices de diseño, $\boldsymbol{\beta}$ es un vector que contiene todos los parámetros de efectos fijos de la ecuación (7) y $\boldsymbol{u}$ y $\boldsymbol{e}$ son vectores que contienen los efectos aleatorios de la interacción entre el centro y el programa (u_{01k}) $\times$ PGR, del centro (u_{00k}), del aula ($u_{0\,jk}$) y del estudiante (e_{ijk}), respectivamente. Se asume que el vector de coeficientes aleatorios $\boldsymbol{u}$ se distribuye normal e independiente de $\boldsymbol{e}$ con media cero y matriz de covarianza $\boldsymbol{G}$, donde $\boldsymbol{G}$ es una matriz diagonal de bloques de dimensión K. La matriz $\boldsymbol{G}_k$ es común para cada centro y las entradas de su diagonal principal recogen la varianza del centro, la varianza del centro para cada condición experimental y la varianza de cada clase.

Similarmente, también se asume que el vector de coeficientes aleatorios $\boldsymbol{e}$ se distribuye normal e independiente con esperanza cero y matriz de covarianza $\boldsymbol{R}$, donde $\boldsymbol{R}$ es una matriz diagonal de bloques de dimensión I. Al igual que ocurría con la matriz $\boldsymbol{G}_k$ correspondiente a los centros, $\boldsymbol{R}_i$ también es idéntica para cada estudiante. La matriz de covarianza global para el modelo de la ecuación (6) es $\boldsymbol{V} = V(\boldsymbol{y}) = \boldsymbol{Z}\,\boldsymbol{G}\,\boldsymbol{Z}' + \boldsymbol{R}$.

Implementación del modelo estadístico dentro del programa

En una determinada aplicación existen diversas vías de implementar un modelo. Sin embargo, como señala Littell (2007), lo que procede no es tomar el camino más directo, sino el que resulte más eficiente. Así pues, comenzaremos el análisis formulando un modelo mixto completamente incondicional o modelo ANOVA con dos factores aleatorios. De acuerdo con Singer (1998), este paso preliminar constituye un indicador de la variabilidad asociada con los diferentes niveles implicados en el análisis y, por consiguiente, sirve como referente para comparar la bondad de ajuste de sucesivos modelos a los datos. En nuestro caso, se trata de verificar si los componentes de varianza asociados con el rendimiento de los estudiantes dentro de las clases, con el rendimiento promedio entre las clases anidadas dentro de los centros escolares y con el rendimiento promedio entre los centros difieren significativamente de cero.

Para responder a las cuestiones planteadas, el modelo de la ecuación (6) se reduce a $Y_{ijk} = \gamma_{000} + u_{00k} + u_{0\,jk} + e_{ijk}$. Las expresiones básicas requeridas por el módulo *Proc Mixed* del programa SAS para ajustar dicho modelo son:

```
proc mixed method = reml noclprint noitprint covtest;
    class centros clases;
    model postest = / solution ddfm=kr;
    random intercept / subject = centros type = un solution;
    random intercept / subject = clases (centros) type = un solution; run;
```

La sintaxis utilizada para su ejecución con el programa SPSS se detalla en el Apéndice A.

Examinando detenidamente las cinco líneas del programa SAS se observan dos partes: la parte estructural (las dos primeras líneas) y la parte de modelado (las tres últimas líneas). En la expresión *proc* inicial de la parte estructural, la opción *method* permite especificar el método que se desea utilizar para estimar los parámetros. Las opciones *noclprint* y *noitprint* impiden que aparezca impresa la información correspondiente a la expresión *class* y a la historia de las iteraciones, respectivamente. La opción *covtest* permite contrastar las hipótesis referidas a los componentes de varianza. La expresión *class* indica que *centros* y *clases* son variables de clasificación cuyos valores no contienen información cuantitativa. En lo referente a la parte de modelado, la expresión *model* establece la estructura de los componentes de efectos fijos, en este caso no se incorpora ninguna variable. La opción *solution* ordena al programa que imprima las estimaciones basadas en la distribución t para la parte fija del modelo, la opción *ddfm = kr* le pide que compute los grados de libertad usando el método desarrollado por Kenward y Roger (1997). Mientras que la expresión *random* especifica la estructura de la matriz $\boldsymbol{G}$ listando los predictores aleatorios que son incluidos en la matriz $\boldsymbol{Z}$ de la ecuación (7), en este caso se incorpora el intercepto y la pendiente. La opción *subject* especifica la estructura jerárquica de los datos necesaria para crear la matriz $\boldsymbol{Z}$ y la opción *solution* ordena al programa que imprima las estimaciones referidas a la parte aleatoria del modelo. Aunque estas especificaciones son las más comunes, la gran flexibilidad del programa permite otras muchas órdenes alternativas (véase Littell, Milliken, Stroup, Wolfinger y Schabenberger, 2006).

En la tabla 1 se muestran los resultados obtenidos para los datos del ejemplo descrito anteriormente. Por lo que respecta a los efectos fijos, se constata que el promedio de rendimiento global difiere de cero en la muestra de centros ($p<.0001$). Sin embargo, el resultado más destacable es la existencia de diferencias estadísticamente significativas en el desempeño de los estudiantes dentro de las clases ($p<.0001$), en el desempeño promediado entre las clases anidadas dentro de los centros ($p<.0001$) y en el desempeño promediado entre los centros (p= .0037). A su vez, es fácilmente comprobable que las observaciones de los estudiantes dentro un mismo centro están ligeramente correlacionadas (.0142/.1844 ≈ .08), mientras que el grado de dependencia entre las observaciones de los estudiantes dentro de una misma clase es moderada (.0348/.1844 ≈ .19).

Dado que las hipótesis referidas a los componentes de varianza difieren de cero, la pregunta es ¿por qué existen diferencias en el rendimiento matemático de los estudiantes? Para responder a dicho interrogante es necesario restringir la atención a la estructura de datos obtenida tras registrar el desempeño de los estudiantes a través de las clases y de los centros, junto con las variables explicativas registradas en el nivel 1. En concreto, se modela la relación entre el desempeño en el postest y las cuatro variables predictoras que siguen: el desempeño en el pretest (LB), el nivel socioeconómico de los estudiantes (NSE), el género de los mismos (GNR) y el nivel educativo de los padres (NEP). Se asume que la relación entre las variables se mantiene estable a través de las diferentes clases y centros. En el caso hipotético de que el desempeño promedio no variase a través de los centros, ni tampoco lo hiciese la relación entre las variables entre las clases dentro de los centros, se debería concluir que ni las clases ni los centros difieren entre sí en cuanto al desempeño se refiere. Por ende, el modelo de la regresión de coeficientes aleatorios no sería necesario para ajustar los datos, bastaría con el modelo de un único componente aleatorio.

Para comprobar lo dicho, el modelo mixto de la ecuación (6) se simplifica a $Y_{ijk}= \gamma_{000} + \gamma_{100}\,\mathrm{LB} + \gamma_{200}\,\mathrm{NSE} + \gamma_{300}\,\mathrm{GNR} + \gamma_{400}\,\mathrm{NEP} + u_{00k} + u_{0\,jk} + e_{ijk}$, mientras que las expresiones básicas requeridas por SAS *Proc Mixed* para ajustar dicho modelo son:

```
proc mixed method = reml noclprint noitprint covtest;
    class centros clases;
    model postest = lb nse gnr nfp / solution ddfm=kr;
    random intercept / subject = centros type = un solution;
    random intercept / subject = clases (centros) type = un solution; run;
```

Tabla 1
Resumen de los resultados obtenidos con el modelo mixto completamente incondicional

Solución para los efectos fijos

Efecto	Estimador	Error estándar	GL	Valor *t*	Pr > |t|
γ000	5.2154	0.0218	59	239.58	<.0001

Solución para los efectos aleatorios[a]

Estimadores parámetros de covarianza

Par Cov	Efecto	Estimador	Error estándar	Valor Z	Pr > Z
$u_{00\,k}$	Centros	0.0142	0.0053	2.68	0.0037
$u_{0\,jk}$	Clases/Centros	0.0206	0.0055	3.70	<.0001
e_{ijk}	Residual	0.1496	0.0068	21.93	<.0001

Estadísticos de ajuste

Descripción	*Valor*
-2 logaritmo de la función ML	1285.1
Criterio AIC	1291.1

Nota: [a] Para ahorrar espacio se omite la información correspondiente a la solución aleatoria

Un listado con la sintaxis utilizada por el programa SPSS se detalla en el Apéndice A.

En la tabla 2 se muestran los resultados más importantes que se obtuvieron tras ajustar un modelo de intercepto aleatorio con múltiples predictores de nivel 1. Se constata que, en promedio, el desempeño medido antes de intervenir predice significativamente el logro alcanzado después de intervenir (p<.0001). Se encuentra, además, que tanto el nivel socioeconómico de los estudiantes (p= .0159) como el nivel educativo de los padres (p= .0173) afectan significativamente al rendimiento matemático de los participantes; sin embargo, no ocurre lo mismo con la variable género. Con respecto a las varianzas de nivel 2 y 3, se mantienen las diferencias estadísticamente significativas en el rendimiento promedio de los estudiantes entre las clases (p= .0099), no así entre los centros (p= .2424).

Una vez que se ha puesto de relieve que el logro académico medio es más elevado en unas clases que en otras y que la relación funcional entre el rendimiento antes y después de administrar el programa varía a través de las diferentes aulas escolares, procede intentar construir un modelo explicativo que dé cuenta de dicha variabilidad. Esto es, se requiere comprender ¿por qué en algunas clases el desempeño académico es mayor que otras? Para dar cuenta de este interrogante llevamos a cabo un nuevo análisis incorporando las covariables registradas en el nivel clase; prestando especial atención al programa elaborado para incrementar el rendimiento. Parece sensato pensar, en principio, que el programa sea la variable explicativa más importante de la presente investigación.

Tabla 2
Resumen de los resultados obtenidos con el modelo condicional de intercepto aleatorio con múltiples predictores de nivel 1

Solución para los efectos fijos

Efecto	Estimador	Error estándar	GL	Valor *t*	Pr > |t|
Inter	2.2096	0.1134	59	19.48	<.0001
LB	0.6282	0.0246	1125	25.57	<.0001
NSE	0.0330	0.0137	1125	2.41	0.0159
GNR	0.0126	0.0188	1125	0.67	0.5017
NEP	0.0557	0.0234	1125	2.38	0.0173

Solución para los efectos aleatorios[a]

Estimadores parámetros de covarianza

Par Cov	Efecto	Estimador	Error estándar	Valor Z	Pr > Z
$u_{00\,k}$	Centros	0.0011	0.0015	0.70	0.2424
$u_{0\,jk}$	Clases/Centros	0.0068	0.0029	2.33	0.0099
e_{ijk}	Residual	0.0985	0.0045	21.94	<.0001

Estadísticos de ajuste

Descripción	*Valor*
-2 logaritmo de la función ML	723.6
Criterio AIC	729.6

Nota: [a] Para ahorrar espacio se omite la información correspondiente a la solución aleatoria

Bajo esta nueva situación el modelo mixto de la ecuación (6) se simplifica a: $Y_{ijk} = \gamma_{000} + \gamma_{010}\,\text{PRG} + \gamma_{020}\,\text{EDP} + \gamma_{030}\,\text{TCL} + \gamma_{110}\,\text{LB} \times \text{PRG} + \gamma_{100}\,\text{LB} + \gamma_{200}\,\text{NSE} + \gamma_{400}\,\text{NEP} + u_{01k} \times \text{PRG} + u_{00k} + u_{0\,jk} + e_{ijk}$, mientras que las expresiones básicas requeridas por SAS *Proc Mixed* para ajustar dicho modelo son:

```
proc mixed method = reml noclprint noitprint covtest;
    class centros clases;
    model postest = lb nse nep prg lb * prg edp tcl / solution
    ddfm=kr;
    random intercept prg / subject = centros type = un solution;
    random intercept / subject = clases (centros) type = un solution; run;
```

Expresiones similares utilizando el programa SPSS se muestran en el Apéndice A.

En la tabla 3 aparecen los resultados más importantes que se obtuvieron tras ajustar un modelo de intercepto y pendiente aleatorios con múltiples predictores de nivel 1 y 2. Una vez ajustado este nuevo modelo condicional se aprecia que el tratamiento diseñado para incrementar el rendimiento académico ha surtido los efectos deseados. En concreto, se observa que, tras ajustar los efectos de las diversas covariables, el programa fue efectivo como efecto principal (p<.0001) y como efecto secundario (p= .0002); pues, además de afectar al rendimiento promedio dentro de las clases, debilita significativamente la relación funcional entre el desempeño previo y posterior a su administración. Debido al efecto moderador del programa, puntuar bajo en el pretest no conlleva necesariamente hacerlo en el postest. La suerte de las dos covariables restantes fue dispar, así mientras el tamaño de la clase parece afectar negativamente al rendimiento, la experiencia de los profesores apenas si se deja sentir. En otro orden de cosas, conviene advertir que, cuando se controlan los efectos del programa y del tamaño de las clases, desaparece la significación estadística del componente de varianza correspondiente a las clases anidadas dentro de los centros. En consecuencia, estas variables parecen ser la causa de que en algunas clases el desempeño sea mayor que otras. Es importante resaltar que la variación observada en el desempeño promedio entre las clases se redujo casi a la mitad (100 × (.0068 – .0035) /.0068). Esto pone de relieve que una gran parte de la variabilidad observada entre las clases está asociada con el programa y, en menor medida, con el tamaño de las mismas.

Tabla 3
Resumen de los resultados obtenidos con el modelo de intercepto y pendiente aleatorios con múltiples predictores de nivel 1 y 2

Solución para los efectos fijos

Efecto	Estimador	Error estándar	GL	Valor *t*	Pr > |t|
Inter	1.8284	0.1508	59	14.88	<.0001
LB	0.7122	0.0322	1122	22.73	<.0001
NSE	0.0319	0.0136	1122	3.59	0.0189
NEP	0.0518	0.0232	1122	2.77	0.0256
PRG	0.8819	0.2165	1122	3.05	<.0001
LB × PRG	-0.1757	0.0462	1122	-2.80	0.0002
EDP	0.0013	0.0011	1122	1.31	0.2279
TCL	-0.0160	0.0080	1122	-1.92	0.0452

Solución para los efectos aleatorios[a]

Estimadores parámetros de covarianza

Par Cov	Efecto	Estimador	Error estándar	Valor Z	Pr > Z
$u_{00\,k}$	Centros	0.0009	0.0018	0.51	0.3046
$u_{01\,k}$	Centros × PRG	0.0045	0.0040	1.11	0.1330
$u_{0\,jk}$	Clases/Centros	0.0035	0.0028	1.26	0.1041
e_{ijk}	Residual	0.0972	0.0044	21.96	<.0001

Estadísticos de ajuste

Descripción	*Valor*
-2 logaritmo de la función ML	715.7
Criterio AIC	723.7

Nota: [a] Para ahorrar espacio se omite la información correspondiente a la solución aleatoria

El paso final en el proceso de selección del modelo mixto consiste en realizar un nuevo análisis incorporando la variable de nivel 3, titularidad del centro docente, y todas aquellas otras variables que durante el proceso de modelado han resultado estadísticamente significativas. Aunque no se muestran los resultados, la variable de nivel 3 no afecta al rendimiento. En consecuencia, el modelo que explica de una manera más parca los datos del ejemplo simulado debería incorporar las variables referidas al pretratamiento, a los recursos económicos, a la formación educativa de los padres, al programa y al tamaño de las clases. También debe incorporar la información referida a la interacción entre las variables fijas programa y pretratamiento, así como la información referida a la interacción entre la variable fija programa y la variable aleatoria centros escolares.

Extensión de la construcción del modelo al caso longitudinal

Los modelos lineales mixtos también proporcionan un enfoque general para modelar datos estructurados jerárquicamente contemplando el efecto del paso del tiempo (para una exposición detallada del tema véase Singer y Willet, 2003; Fitzmaurice, Laird y Ware, 2004). El objetivo de estos diseños es el estudio del cambio intra-individual e inter-individual, así como de sus posibles causas. Esencialmente, cuando los datos incorporan la dimensión temporal la efectividad de los programas se puede evaluar mediante diseños en los cuales las ocasiones de observación se realizan en los mismos conglomerados de sujetos en intervalos temporales frecuentes y poco distantes, o bien en intervalos temporales distantes y poco frecuentes. También puede ocurrir que las ocasiones de observación se obtengan utilizando muestras diferentes de sujetos de la misma población a las que se evalúa en distintos momentos o tandas.

Aunque es posible acomodar muchos diseños a este tipo de datos, solamente ilustraremos el análisis del anterior diseño de medidas parcialmente repetidas de muestras divididas con variables dependientes continuas. Detalles de modelos jerárquicos referentes a variables dependientes dicotómicas pueden ser hallados en Livert, Rindskopf, Saxe y Stirratt (2001). El modelado comienza formulando una ecuación que recoja el paso del tiempo dentro del estudiante (nivel 1), después se define un modelo que recoja la variación entre los estudiantes anidados dentro de las clases (nivel 2), luego se continúa con la definición de un modelo que recoja la variación entre las clases anidadas dentro de los centros (nivel 3) y, finalmente, se concluye definiendo un modelo que recoja la variación entre los centros escolares (nivel 4). Dado que los modelos que incluyen la variación entre-sujetos, es decir, los formulados para los niveles dos, tres y cuatro, son paralelos a los presentados en las ecuaciones (3-5), escribimos directamente el modelo multinivel resultante de combinar el sistema de ecuaciones. En el Apéndice B se ofrecen los detalles seguidos en el proceso de construcción.

El modelo combinado para la calificación obtenida en el t tiempo, por el i estudiante dentro de la j clase anidada en el k centro escolar se puede escribir como sigue:

$$\begin{aligned} Y_{tijk} = {} & \delta_{0000} + \delta_{0001}\,\mathrm{TCD} + \delta_{0010}\,\mathrm{PRG} + \delta_{0020}\,\mathrm{EDP} + \delta_{0030}\,\mathrm{TCL} \\ & + \delta_{0100}\,\mathrm{NSE} + \delta_{0200}\,\mathrm{GNR} + \delta_{0300}\,\mathrm{NEP} + \delta_{1000}\,\mathrm{TMP} \\ & + \delta_{1010}\,\mathrm{PRG}\times\mathrm{TMP} + u_{000k} + u_{001k}\times\mathrm{PRG} + u_{00\,jk} + u_{100k}\times\mathrm{TMP} \\ & + u_{101k}\times\mathrm{PRG}\times\mathrm{TMP} + u_{10\,jk}\times\mathrm{TMP} + u_{0\,ijk} + u_{1\,ijk}\times\mathrm{TMP} \\ & + e_{ijkt} \end{aligned} \tag{8}$$

La parte fija de la ecuación (8) contiene todos los términos de efectos fijos definidos en la ecuación (6), además incorpora de manera específica los términos referidos al factor tiempo (δ_{1000} TMP) y a la interacción de este factor con el tratamiento (δ_{1010} PRG × TMP). La parte aleatoria de la ecuación (8) también contiene todos los términos de efectos aleatorios recogidos en la ecuación (6), en concreto, los términos asociados con el factor centros (u_{000k}), con el factor clases anidadas dentro los centros ($u_{00\,jk}$), con el factor sujetos anidados simultáneamente dentro de clases y centros ($u_{0\,ijk}$) y con la interacción entre los centros y el programa ($u_{001k}\times$ PRG). Además, añade de manera específica la interacción de los centros con el tiempo ($u_{100k}\times$ TMP), la interacción entre los centros, el programa y el tiempo ($u_{101k}\times$ PRG × TMP), la interacción de las clases con el tiempo ($u_{10\,jk}\times$ TMP) y la interacción de los sujetos con el tiempo ($u_{1\,ijk}\times$ TMP). Es decir, se han incorporado las interacciones de todos aquellos factores que se hallan cruzados con el tiempo y el programa, a excepción de las covariables registradas en el nivel individual (estatus socioeconómico, género y nivel educativo de los padres), grupal inferior (experiencia docente y tamaño de las clases) y grupal superior (titularidad del centro).

Al igual que en el caso transversal, una vez construido el modelo procede trasladarle dentro de un programa. La implementación del modelo estadístico longitudinal dentro del programa requiere seleccionar una estructura lo más parsimoniosa posible, tanto para la parte fija como para la parte aleatoria. Para identificar la estructura asociada con la parte fija conviene seguir los pasos descritos en la sección anterior, mientras que para caracterizar la estructura asociada con la parte aleatoria, específicamente la forma de las matrices ***G*** y ***R***, resulta informativo realizar análisis separados con cada una de las ocasiones de medida antes de pasar a combinarlas en un modelo total.

Para ahorrar espacio no se muestran las salidas, pero los resultados obtenidos por separado con las puntuaciones registradas en el pretest y en el postest, formulando modelos completamente incondicionales, pusieron de relieve que las varianzas en el nivel de la escuela, clase y estudiante eran estrechamente coincidentes. A continuación, procedimos a evaluar la significación de los efectos fijos de cada nivel del modelo teórico de interés en el conjunto de los datos. Después evaluamos el componente aleatorio, teniendo sumo cuidado en comprobar si las estimaciones del análisis conjunto eran compatibles con las halladas a partir de los análisis separados. Concluimos examinando de nuevo los componentes de efectos fijos a la luz de los componentes aleatorios y viceversa.

Aplicando la rutina descrita el modelo mixto de la ecuación (8) se simplifica a $Y_{tijk} = \delta_{0000} + \delta_{0010}\,\mathrm{PRG} + \delta_{0100}\,\mathrm{NSE} + \delta_{0300}\,\mathrm{NFP} + \delta_{1000}\,\mathrm{TMP} + \delta_{1010}\,\mathrm{PRG}\times\mathrm{TMP} + u_{000k} + u_{00\,jk} + u_{10\,jk}\times\mathrm{TMP} + u_{0ijk} + u_{1ijk}\times\mathrm{TMP} + e_{ijkt}$. Las expresiones requeridas por SAS *Proc Mixed* para ajustar dicho modelo son:

```
proc mixed method = reml noclprint noitprint covtest;
    class sujetos centros clases prg tmp;
    model y = tmp nse nep prg prg*tmp / solution ddfm=kr;
    random intercept / subject = centros type = vc;
    random intercept tmp / subject = clases (centros) type = vc;
    repeated tmp / subject = sujetos (clases centros) type = cs;
    lsmeans prg*tmp /slice = tmp; run;
```

Expresiones similares utilizando el programa SPSS aparecen recogidas en el Apéndice A.

La tabla 4 recoge los resultados que se obtuvieron tras ajustar el modelo jerárquico de cuatro niveles que contempla el efecto del paso del tiempo. Las pruebas de efectos fijos muestran la existencia de diferencias estadísticamente significativas entre las observaciones registradas antes y después de intervenir (p<.0001) y entre las clases que reciben y no reciben el programa (p= .0094). Lo mismo sucede con la interacción entre el programa y el tiempo (p= .0017). Un análisis pormenorizado de esta interacción indica que, si bien antes de intervenir el rendimiento promedio de las clases

Tabla 4

Resumen de los resultados obtenidos con el modelo de intercepto y pendiente aleatorios con múltiples predictores de nivel 1 y 2

Solución para los efectos fijos

Efecto	Estimador	Error estándar	GL	Valor *t*	Pr > |*t*|
Inter	5.0329	0.0416	59	120.87	<.0001
TMP	-0.5907	0.0160	59	-36.83	<.0001
NSE	0.0912	0.0146	2313	6.26	<.0001
NEP	0.1676	0.0270	2313	6.28	<.0001
PRG	-0.0723	0.0269	59	-2.69	0.0094
TMP × PRG	0.0764	0.0231	59	3.30	0.0017

Solución para los efectos aleatorios[a]

Estimadores parámetros de covarianza

Par Cov	Efecto	Estimador	Error estándar	Valor Z	Pr > Z
$u_{000\,k}$	Centros	0.0084	0.0033	2.56	0.0053
$u_{00\,k}$	Clases (Centros)	0.0096	0.0037	2.55	0.0054
$u_{10\,jk}\times$ TMP	Clases × Tiempo	0.0038	0.0015	2.56	0.0053
$u_{0\,ijk}$	Sujetos (Clases/Centros)	0.0587	0.0027	15.41	<.0001
$u_{10\,jk}\times$ TMP	Sujetos × Tiempo	0.0791	0.0051	22.01	<.0001
e_{ijkt}	Residual	0.0791	0.0051	22.01	<.0001

Estadísticos de ajuste

Descripción	***Valor***
-2 logaritmo de la función ML	1763.7
Criterio AIC	1785.7

Nota: [a] Para ahorrar espacio se omite la información correspondiente a la solución aleatoria asignadas a la condición de tratamiento era similar al rendimiento de las clases asignadas a la condición de control, tras la intervención las diferencias entre ambos grupos de clases fueron sustanciales. También se observa que tanto el nivel socioeconómico de los estudiantes (p<.0001) como el nivel educativo de los padres (p<.0001) afectan significativamente al rendimiento matemático de los participantes. A su vez, se constata la significación estadística de los componentes de varianza correspondientes a los centros, a las clases anidadas dentro de los centros y a los sujetos anidados dentro de las clases. Y todo ello pese a controlar los efectos del programa, del nivel socioeconómico de los estudiantes y del nivel educativo de los padres.

Consideraciones finales

Antes de iniciar el análisis de los datos es necesario especificar claramente a qué interrogantes se pretende responder. Cuando se comparan modelos mixtos lo dicho alcanza mayor relevancia, si cabe. Si las cuestiones no están bien planteadas, es prácticamente imposible alcanzar las soluciones adecuadas. Asimismo, conviene tener presente que si bien algunos problemas de investigación tienen que ver con parámetros asociados con la parte aleatoria del modelo, la gran mayoría son respondidos con los incluidos en la parte fija. Por ende, a la hora de construir la matriz de diseño correspondiente a los componentes de efectos fijos, no sólo se deberán incluir aquellas variables que atañen directamente a las cuestiones planteadas en la investigación, sino también a aquellas otras que lo hacen indirectamente por confundir las inferencias y/o moderar los resultados.

Si los efectos incluidos en la parte fija y aleatoria del modelo no están bien especificados, el contraste de las hipótesis correspondientes a dichos efectos resultará sesgado. También es necesario resaltar que ambos efectos implican cuestiones diferentes, pero complementarias. Por tal motivo, la exactitud de las inferencias referidas a los efectos fijos estará condicionada por la correcta especificación de la parte aleatoria del modelo, mientras que lo inverso también será cierto. Por consiguiente, durante el proceso de modelado se recomienda seguir alguna estrategia para comprobar si los efectos fijos incluidos en el modelo son relevantes, asumiendo que los parámetros de efectos aleatorios están perfectamente especificados, y viceversa. Debido a la complejidad de los modelos mixtos no es sencillo encontrar una estrategia de modelado ideal, no obstante, concluiremos resumiendo sucintamente los pasos seguidos en el manuscrito.

Inicialmente, la rutina requiere evaluar la significación de los efectos fijos de cada nivel del modelo teórico de interés. En esta fase conviene seleccionar una estructura simple para la parte aleatoria del modelo y una estructura con un relativo número de parámetros para la parte fija. Para obtener un componente de efectos fijos lo más parco que sea posible se eliminan del modelo las variables grupales e individuales que no resulten significativas. Se excusa decir que con la finalidad de poder estudiar la cuestión de interés correctamente, algunas variables carentes de significación deberán permanecer en el sistema. Por ejemplo, no conviene eliminar el programa de entrenamiento del modelo a pesar de que esta variable no surta los efectos deseados (Kreft, 1997).

Seguidamente, se procede a evaluar el componente aleatorio, incluyendo en el modelo los factores de efectos fijos que han mostrado su relación con la conducta de interés en la fase inicial. En el ámbito longitudinal, Wallace y Green (2002) recomiendan comenzar con una estructura aleatoria relativamente compleja e ir descartando modelos hasta alcanzar la estructura más simple posible. Para comparar modelos anidados que no impliquen cambios en los efectos fijos es costumbre utilizar el contraste de razón de verosimilitudes. Si los modelos no están anidados, se puede elegir entre ellos utilizando alguno de los múltiples criterios de información existentes (véase Littell et al., 2006).

Finalmente, dado que la exactitud de las pruebas referidas a los componentes de efectos fijos dependen de la correcta especificación de los componentes de efectos aleatorios, es conveniente volver a examinar los primeros a la luz de los segundos, y viceversa. Obviamente, conviene procurar que la rutina no se convierta en bucle sin fin.

Agradecimientos

Este trabajo ha sido financiado mediante sendos Proyectos de Investigación concedidos por el Plan Nacional de Investigación Científica, Desarrollo e Innovación Tecnológica del Ministerio de Educación y Ciencia (Ref.: SEJ2005-01883 y SEJ2005-01923).

Apéndice A

Expresiones básicas requeridas por el programa SPSS para obtener los resultados mostrados en las tablas 1, 2, 3 y 4, respectivamente.

```
mixed postest by clases centros
/fixed = | sstype(3)
/method = reml
/print = solution
/random = clases (centros) | covtype(un)
/random = centros centros*prg | covtype(un).

mixed postest with lb nse gnr nep by clases centros
/fixed = lb nse gnr nep | sstype(3)
/method = reml
/print = solution
/random = clases (centros) | covtype(un)
/random = centros centros*prg| covtype(un).

mixed postest with lb nse ngr nep prg edp tcl by clases centros
/fixed = lb nse gnr nep prg lb*prg edp tcl | sstype(3)
/method = reml
/print = solution
/random = clases (centros) | covtype(un)
/random = centros centros*prg | covtype(un).

mixed respuesta with nse nep by tiempo prg clases centros
/fixed = prg tiempo prg*tiempo | sstype(3)
/method = reml
/print = solution
/random intercept tiempo prg tiempo*prg | subject (centros) covtype(vc)
/random intercept tiempo | subject (clases*centros) covtype(vc)
/repeated tiempo | subject (id*clases*centros) covtype(cs)
```

Apéndice B

Nivel 1 (Modelo dentro del sujeto)

$Y_{tijk} = \pi_{0ijk} + \pi_{1ijk}\,\text{TIEMPO} + e_{tijk}$

Nivel 2 (Modelo entre sujetos/clases y centros)

$\pi_{0ijk} = \beta_{00\,jk} + \beta_{01\,jk}\,\text{NSE} + \beta_{02\,jk}\,\text{GNR} + \beta_{03\,jk}\,\text{NEP} + u_{0\,ijk}$,
$\pi_{0ijk} = \beta_{10\,jk} + u_{1\,ijk}$

Nivel 3 (Modelo entre clases/centros)

$\beta_{00\,jk} = \gamma_{000\,k} + \gamma_{001\,k}\,\text{PGR} + \gamma_{002\,k}\,\text{EDP} + \gamma_{003\,k}\,\text{TCL} + u_{00\,jk}$,
$\beta_{10\,jk} = \gamma_{100\,k} + \gamma_{101\,k}\,\text{PGR} + u_{10\,jk}$, $\beta_{01\,jk} = \beta_{010\,k}$,
$\beta_{02\,jk} = \gamma_{020\,k}$, $\beta_{03\,jk} = \beta_{030\,k}$

Nivel 4 (Modelo entre centros escolares)

$\gamma_{000\,k} = \delta_{0000} + \delta_{0001}\,\text{TCD} + u_{000\,k}$, $\gamma_{001\,k} = \delta_{0010} + u_{001\,k}$,
$\gamma_{100\,k} = \delta_{1000} + u_{100\,k}$, $\gamma_{101\,k} = \delta_{1010} + u_{101\,k}$,
$\gamma_{010\,k} = \delta_{0100}$, $\gamma_{020\,k} = \delta_{0200}$, $\gamma_{030\,k} = \delta_{0300}$

Referencias

Ato, M., y Vallejo, G. (2007). *Diseños experimentales en Psicología*. Madrid: Pirámide.

Fitzmaurice, G.M., Laird, N.M., y Ware, J.H. (2004). *Applied longitudinal analysis*. Hoboken, NJ: Wiley.

Goldstein, H. (2002). *Multilevel statistical models* (3rd ed.). London: Arnold.

Hedeker, D., y Gibbons, R.D. (2006). *Longitudinal data analysis.* Hoboken, NJ: Wiley.

Hox, J.J. (2002). *Multilevel analysis: Techniques and applications*. Mahwah, NJ: LEA.

Kenward, M.G., y Roger, J.H. (1997). Small sample inference for fixed effects from restricted maximum likelihood. *Biometrics, 53*, 983-997.

Kreft, I.G.G. (1997). The interactive effect of alcohol prevention program in high school classes: An illustration of item homogeneity scaling and multilevel analysis techniques. En K.J. Bryant, M. Windle y S.G. West (Eds.): *The Science of Prevention: Methodological Advances from Alcohol and Substance Abuse Research* (pp. 251-277). Washington, DC: American Psychological Association.

Kull, L.J., y MacKinnon, D.P. (2001). Multinivel modeling of individual and group level mediated effects. *Multivariate Behavioral Research, 36*, 249-277.

Littell, R.C. (2007). *Repeated measures analysis with clustered subjects*. SAS Global Forum 2007, April, 16-19, Orlando, Florida.

Littell, R.C., Milliken, G.A., Stroup, W.W., Wolfinger, R.D., y Schabenberger, O. (2006). *SAS® for mixed models* (2nd ed.). SAS Press, Cary, NC.

Livert, D., Rindskopf, D., Saxe, L., y Stirratt, M. (2001). Using multilevel modeling in the evaluation of community-based treatment program. *Multivariate Behavioral Research, 36*(2), 155-183.

Oliver, J.C., Rosel, J., y Jara, P. (2000). Modelos de regresión multinivel: aplicación en psicología escolar. *Psicothema, 12,* 487-494.

Raudenbush, S.W., y Bryk, A.S. (2002). *Hierarchical linear models. Applications and data.* Thousand Oaks, CA: Sage.

Singer, D.J. (1998). Using SAS PROG MIXED to fit multilevel models, hierarchical models and individual growth models. *Journal of Educational and Behavioral Statistics, 24*, 323-353.

Singer, D.J., y J.B. Willet (2003). *Applied longitudinal data analysis.* New York: Oxford University Press.

Snijders, T.A.B., y Bosker, R.J. (2004). *Multilevel analysis. An introduction to basic and advanced multilevel modelling*. Thousand Oaks, CA: Sage.

Vallejo, G., Fernández, J.R., y Secades, R. (2006). Amenazas a la validez analítica de las técnicas usadas habitualmente en la evaluación de programas. *Análisis y Modificación de Conducta, 32*, 86-101.

Wallace, D., y Green, B.S. (2002). Analysis of repeated measures designs with linear mixed models. En D.S. Moskowitz y S.L. Hershberger (Eds.): *Modeling intraindividual variability with repeated measures data: Methods and applications* (pp. 135-170). Mahwah, NJ: LEA.