

XIII JORNADAS DE LA A.E.D.E.
San Sebastián, 13-14 septiembre 2004

***LA ELECCIÓN DE INDICADORES DE
RESULTADOS EN LA EDUCACIÓN SECUNDARIA.***

María de los Ángeles García Valiñas

Manuel Antonio Muñoz Pérez

Departamento de Economía. Universidad de Oviedo.

ABSTRACT:

La selección de indicadores de resultados y productividad es una cuestión de plena actualidad en el campo de la gestión y evaluación de los servicios públicos en general, y en el caso de la Educación Secundaria en particular. Debemos tener en cuenta que, para lograr una correcta evaluación y calificación de la actividad productiva de cada centro educativo, es necesario prestar la atención que merece al proceso a través del cual se seleccionan los criterios e indicadores finales en base a los que se va a llevar a cabo dicha evaluación. Para dar respuesta a esta cuestión, en este trabajo se plantean y se da respuesta a dos aspectos complementarios. En primer lugar, detectar si se dan o no en la realidad en un número significativo de escuelas las potenciales situaciones que podrían sesgar la evaluación, como por ejemplo la especialización en determinados resultados. En segundo lugar y como continuación del punto anterior, comprobar cómo las distintas alternativas de definición de variables de output relacionadas con la PAU dan o no una respuesta adecuada a los casos problemáticos detectados. Como datos del estudio se emplean los relativos a una muestra de 215 Institutos de Educación Secundaria públicos localizados en las CC. AA. de Asturias, Aragón y Extremadura. La metodología aplicada es el Análisis Envolvente de Datos o DEA, técnica no paramétrica de evaluación de la eficiencia productiva que se ha relevado en los últimos años como la más adecuada para la evaluación de la eficiencia productiva en los servicios públicos.

1. INTRODUCCIÓN.

La selección de indicadores de resultados y productividad es una cuestión de plena actualidad en el campo de la gestión y evaluación de los servicios públicos. Debemos tener en cuenta que, para lograr una correcta evaluación y calificación de la actividad productiva de cada unidad, es necesario prestar la atención que merece al proceso a través del cual se seleccionan los criterios e indicadores finales en base a los que se va a llevar a cabo dicha evaluación. Sin embargo, las dificultades a la hora de especificar una definición concreta de output son muy importantes en la mayoría de los servicios públicos, debido sobre todo a su carácter intangible y/o multidimensional. De hecho, es habitual en las investigaciones empíricas que la decisión sobre la variable final de output se base simplemente en una cierta argumentación teórica general (tomando como base la literatura previa relacionada con el sector estudiado) complementada con la necesaria correlación input/output, sin que se suela analizar si las definiciones empleadas con carácter general en un servicio público siguen siendo válidas íntegramente en el momento de confrontarlas con las circunstancias específicas (legislativas, geográficas, políticas,...) de la prestación del servicio en un entorno concreto.

Como servicio público paradigmático de la necesidad de esta reflexión hemos seleccionado la Educación secundaria en España. Con el fin de delimitar el que será nuestro campo de acción en esta investigación, es necesario distinguir previamente entre las dos vertientes potenciales de la discusión sobre el output educativo, teórica y empírica. Desde un punto de vista teórico o conceptual, el objeto del estudio sería no tanto determinar una medida específica del output sino identificar cuál debe ser el aspecto o concepto a medir en una evaluación del proceso educativo. En este punto existen numerosos trabajos desarrollados por teóricos de la ciencia educativa y pedagógica, en los que se desmenuzan las ventajas y desventajas relativas de las múltiples dimensiones que engloba el resultado escolar, existiendo cierto consenso al menos en cuanto a las distintas acepciones a considerar y su importancia relativa¹. Nuestra atención se centrará en la segunda posibilidad, la vertiente empírica, donde la meta consiste en concretar y definir una medida específica de los resultados de las escuelas, con el fin último de lograr que se dé un tratamiento equitativo y riguroso a la heterogénea problemática y conducta que puede envolver la práctica cotidiana de cada uno de los centros evaluados.

En nuestro país, los escasos trabajos empíricos publicados hasta el momento que evalúan la eficiencia productiva de los centros educativos de enseñanza secundaria públicos toman como referencia los resultados en la Prueba de Acceso a la Universidad, como única posibilidad de examen común a todos los Institutos. Sin poner en duda la utilidad de dicho

¹ A este respecto pueden consultarse las revisiones clásicas sobre la función de producción educativa, como Cohn y Millman (1975), Glasman y Biniaminov (1981) o Hanushek (1986) entre otras, o la reciente recopilación de trabajos editada por Hanushek (2003).

referente, la definición específica de la variable de output que se lleve a cabo debería tener en cuenta las potenciales conductas o situaciones educativas que podrían sesgar la calificación que finalmente se asigne a cada centro. De este modo, aspectos como la desigual proporción de alumnos de cada centro interesados en la PAU o una potencial especialización de algunas escuelas en cantidad (porcentaje de aprobados) o calidad (nota media de los aprobados), pueden hacer que se penalice o sobreestime la calificación de aquellos centros envueltos en alguna de estas dinámicas, dependiendo de la variable de output finalmente empleada. Para dar respuesta a esta cuestión, en este trabajo se plantean y se da respuesta a dos aspectos complementarios:

- En primer lugar, detectar si se dan o no en la realidad en un número significativo de escuelas las situaciones potenciales que podrían sesgar la evaluación (como por ejemplo, la especialización en determinados resultados), de cara a comprobar si tiene pertinencia empírica el análisis de sus consecuencias.
- En segundo lugar y como continuación del punto anterior, comprobar cómo las distintas alternativas de definición de variables de output relacionadas con la PAU dan o no una respuesta adecuada a los casos problemáticos detectados.

Como datos del estudio se emplean los relativos a una muestra de 215 Institutos de Educación Secundaria públicos localizados en las CC. AA. de Asturias, Aragón y Extremadura. La metodología aplicada es el Análisis Envolvente de Datos o DEA, técnica no paramétrica de evaluación de la eficiencia productiva que se ha revelado en los últimos años como la más adecuada para la evaluación de la eficiencia productiva en los servicios públicos. El documento se estructura en los siguientes apartados. En el primero se describe sucintamente cómo se han conectado hasta la fecha las recomendaciones teóricas con la elección de variables específicas en los trabajos empíricos publicados. A continuación se describe esta problemática en el contexto específico de la Educación Secundaria en España, planteándose asimismo las hipótesis a contrastar en el análisis empírico posterior en cuanto a conductas de los centros y posibles indicadores de las mismas. En el siguiente epígrafe se da paso a la descripción del análisis empírico efectuado, en cuanto a los datos y metodología empleada y los resultados obtenidos. Finalmente, el habitual apartado de conclusiones cierra el desarrollo de la investigación.

2. ANTECEDENTES Y PLANTEAMIENTO DEL PROBLEMA EN TÉRMINOS EMPÍRICOS.

A pesar de la falta de acuerdo en torno a las distintas posibilidades teóricas de medida del output educativo y la potencial preponderancia de unas sobre otras, las investigaciones empíricas han simplificado la cuestión optando por una dimensión concreta y específica como representante del output escolar, como son los resultados en exámenes específicos de los alumnos de cada centro escolar, con todas las prevenciones añadidas al reconocimiento de que

dicha elección no deja de ser una aproximación al verdadero producto de los centros escolares. Y aquí es donde surge el problema real en términos prácticos: ¿qué aspectos o variables son más adecuadas para medir esos resultados en exámenes? o dicho de otro modo, ¿cuáles son las opciones que nos aseguran en mayor medida que la inevitable heterogeneidad entre las escuelas, y entre el propio alumnado respectivo, e incluso entre los objetivos de ambos colectivos, no va a sesgar los resultados finales de nuestra evaluación?

En el momento de enfrentarnos con la necesidad de dar una respuesta a las preguntas formuladas, existen varios aspectos básicos en los que prácticamente existe un acuerdo entre los investigadores y que nos facilitarán avanzar en nuestra reflexión:

a) La consideración del efecto del estatus socioeconómico y familiar del alumnado de cada escuela. Aunque éste es un aspecto que no está relacionado estrictamente con la medida del output, debemos recordar que la influencia de esta variable tiene un efecto definitivo sobre los resultados que posteriormente alcanzará cada escuela, reconocido incluso como el principal determinante del resultado escolar por encima de los propios recursos del centro. Por tanto, el modelo que se emplee en la evaluación debe incorporar esta influencia adecuadamente.

b) La necesidad de un mismo examen conjunto e igual a todos los centros evaluados. Si el investigador desea una consideración equitativa de los resultados conseguidos por los estudiantes en el examen que se vaya a tomar como referencia, dicho ejercicio debe ser idéntico para todas las escuelas evaluadas. Este aspecto volverá a ser discutido más adelante, cuando hagamos mención del potencial comportamiento estratégico de los centros educativos en la calificación de sus propias pruebas, dependiendo de la definición específica de su output que más les interese manejar en cada caso.

c) La inclusión en la medida del output escolar tanto de aspectos cuantitativos como cualitativos. La elección de un indicador de output educativo que ignore una de estas dos dimensiones sesgaría la evaluación. Parece coherente suponer que es tan importante el número o porcentaje de alumnos a los que se les ofrece una educación que supere determinado baremo (independientemente de cómo se defina este límite mínimo) como el nivel medio que alcanzan los alumnos que superen el baremo anterior. En el primer caso estaríamos hablando de cantidad de output, mientras que en el segundo estaríamos introduciendo en el análisis el grado de calidad del output educativo. Además, la consideración de ambas dimensiones evitaría una potencial sobrevaloración de aquellas escuelas que se especialicen sólo en uno de los dos posibles aspectos.

El escenario que hemos dibujado hasta el momento es perfectamente factible y por tanto sus requerimientos de información son asumibles sin despegarnos de la vocación práctica y empírica del presente estudio. De hecho así sucede en el caso español, con el empleo para este fin de los resultados de los alumnos de cada centro en las pruebas de acceso a la universidad,

información global y conjunta que por desgracia, en nuestro país ha sido suprimida muy recientemente. No obstante, el hecho de que basemos esta investigación en el contexto anterior a la supresión no invalida las conclusiones a que lleguemos. Bien porque puedan ser de aplicación en otro entorno geográfico, bien porque no se pueda descartar que en el futuro se reimplanten estos exámenes en nuestro país (todavía en el momento de redactar este documento ni siquiera está claro que realmente se lleve a cabo la supresión) o bien como punto de partida de posteriores discusiones sobre la misma problemática en el nuevo contexto legal español, no pensamos que las siguientes reflexiones sean una discusión estéril y sin consecuencias prácticas.

El siguiente paso en nuestra exposición consistirá en identificar qué posibles opciones de definición del output educativo, de las que han sido empleadas en la literatura con anterioridad², pueden ser descartadas de antemano basándonos para ello en la lógica y el sentido común. Respecto a la variable de cantidad, la opción más empleada y que a su vez parece más acertada consiste en emplear el porcentaje que representan aquellos alumnos que obtienen una calificación mínima en el examen respecto al total del alumnado (aunque como veremos, al menos en nuestro país tiene una importancia fundamental cómo se defina esta cantidad total de alumnos que sirva de base al cociente). Es habitual establecer el corte a estos efectos en el aprobado, puesto que además dicha nota suele servir de filtro para pasar al nivel educativo superior. Por tanto, podemos descartar el uso de medidas cuantitativas absolutas (como sería en este caso el número de alumnos que aprueben el examen), dado que renunciar al empleo de una medida relativa implicaría primar a aquellos centros con mayor número de alumnos. Lo mismo se puede decir respecto a aquellos estudios que directamente emplean el número absoluto de alumnos graduados (lo que además incumple varios de los supuestos anteriores de partida).

Otra posibilidad observada en ocasiones como medida de cantidad es la inversa del porcentaje de abandono escolar. Esta variable plantea varios problemas, el primero de ellos su definición específica. Si no se relaciona con el comentado examen homogéneo, su empleo quedaría invalidado en sí mismo. Pero aún conectándose con esa prueba, queda por ver qué se considera abandono, si no llegar a presentarse a dicho examen, no pasar los filtros previos que marque cada escuela, dejar de asistir a clase en el transcurso del año académico (lo que puede ser reflejo del entorno familiar y socioeconómico del alumno tanto o más que de la actuación de la escuela), etc. Y en todo caso, habría que conocer si en el entorno específico analizado el abandono de los estudios tiene una importancia numérica. Pero sobre todo, su mayor problema consiste en que no puede considerarse un sustitutivo de la medida relativa de aprobados en el examen conjunto, dado que hacen referencia a dos filtros o dimensiones cuantitativas distintas. Y aun asumiendo su empleo complementario con dicha opción, en ese caso es muy posible que

² Una discusión crítica sobre las variables empleadas al respecto en la literatura previa aquí comentadas, junto con una extensa bibliografía relacionada puede consultarse en Mancebón y Muñiz (2003).

ambas medidas de output tuvieran un grado elevado de correlación. Por tanto, nuestro juicio en este caso pasa por considerar excesivos los costes a asumir frente a los beneficios potenciales de añadir mayor información sobre los resultados de cada escuela. También podemos descartar de antemano el número de alumnos matriculados sin más, dado que no tiene ninguna relación con la conducta productiva de la escuela sino con su tamaño como unidad productiva.

Respecto a la variable a emplear como representante del nivel de calidad, el acuerdo es elevado respecto a la nota media obtenida por los alumnos de cada escuela en el examen conjunto. No obstante, es importante hacer mención de dos matices a este respecto. En primer lugar y dado que no es un error tan esporádico como pueda parecer, queremos recordar que si ya empleamos una medida cuantitativa que recoge el porcentaje de aprobados, la medida de calidad debería computarse teniendo en cuenta únicamente los resultados de los alumnos aprobados. De lo contrario estamos recogiendo el efecto medido en la variable cuantitativa también en la cualitativa, distorsionando por tanto la disociación en ambas dimensiones que pretendíamos. El segundo matiz hace referencia al uso de la desviación típica de las notas al tiempo que la media. La no consideración en este trabajo de su empleo se ha basado en la ausencia de argumentos concluyentes tanto respecto a su importancia como, en tal caso, en cuanto al signo del efecto positivo. Es decir, no sólo no está clara la relevancia de la dispersión de las notas, sino que en todo caso es igualmente defendible a priori (salvo que para un análisis específico así se asuma por unas metas empíricas muy concretas) tanto el beneficio de tener alumnos con muy buenas notas y elevada dispersión como el de tener una gran masa de alumnos de nivel medio.

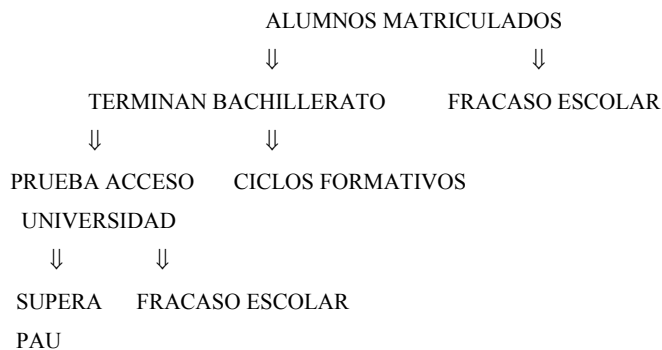
El siguiente paso consistirá en trasladar las conclusiones expresadas hasta el momento a las circunstancias específicas del sistema educativo en nuestro país, con lo que sentaremos las bases e hipótesis de partida del análisis empírico posterior.

3. LA MEDIDA DEL OUTPUT EDUCATIVO EN EL CASO ESPAÑOL: OPCIONES E HIPÓTESIS A CONTRASTAR.

Como es bien sabido, en España hemos contado hasta el momento con una prueba homogénea conjunta para todos los centros educativos de Enseñanza Secundaria, como es la Prueba de Acceso a la Universidad o PAU. No obstante, hay que tener en cuenta dos primeras prevenciones respecto a su utilización para nuestro objetivo. El contenido del examen es distinto para Comunidad Autónoma (lo que nos llevará a realizar análisis empíricos separados para los datos de cada CCAA). Pero sobre todo, que en la configuración presente del sistema educativo español la realización (y aprobado) de este examen por cada alumno individual no es una obligación, salvo que se desee seguir cursando estudios universitarios. Lo ideal en términos teóricos a estos efectos sería disponer de un examen que abarque a toda el alumnado de un determinado grado de cada escuela y que además, todos los alumnos tuviesen motivación por

obtener un buen resultado en ese examen (porque sea requisito para pasar de grado o por cualquier otro motivo). Sin embargo, este supuesto no se cumple en España con la PAU. En nuestra Educación Secundaria se ofertan dos itinerarios alternativos por los que puede optar libremente cada estudiante tras completar sus estudios de Bachillerato, o bien realizar un Ciclo Formativo de Grado Superior o bien iniciar estudios universitarios. A priori, no existen argumentos objetivos que impliquen la superioridad de una alternativa sobre la otra en cuanto a la formación que poseen sus alumnos, puesto que es una decisión en la que pesan tanto los gustos de cada alumno como la potencial presión familiar o incluso las previsiones de futuro laboral. La escuela es ajena a esta elección personal y por tanto, la medida que empleemos del output de cada escuela debe intentar tener en cuenta los resultados de la misma en ambos casos, es decir, la formación conseguida sobre ambos colectivos de estudiantes. Pero mientras que el examen PAU nos informa sobre la formación de los que optan por la Universidad, no existe un equivalente para los alumnos cuyo interés radica en los ciclos formativos.

Situados ante esta tesitura, se nos plantean dos posibilidades. La primera de ellas sería descartar el examen PAU por no informar sobre una parte de los estudiantes. En ese caso la única información que se podría manejar que abarque a todos los alumnos serían los exámenes propios de cada centro en Bachillerato, pero esta opción llevaría al investigador a incumplir el requisito de partida de pruebas idénticas para todas las escuelas. La segunda consiste en emplear la calificación en la PAU, pero en tal caso se debe intentar recoger en la investigación (sea matizando los resultados posteriores o en la misma definición de las variables de output) la heterogeneidad entre las escuelas en cuanto a la distinta proporción que representa en cada una de ellas el colectivo de alumnos que no tienen interés por presentarse a este examen. La problemática descrita para el caso español puede resumirse en el Esquema 1.



Esquema 1. Proceso educativo analizado en España.

No podemos cerrar esta descripción de la situación sin hacer referencia al papel de los posibles comportamientos estratégicos de las escuelas. En efecto, para poder presentarse a la

PAU cada alumno debe aprobar previamente los exámenes propios de cada escuela, lo que incorpora un sesgo potencial en cuanto al rasero que se emplee en cada una para dar por suficientes y completados los estudios de Bachillerato. La detección de un comportamiento estratégico en este sentido no implica una conducta oportunista premeditada de la dirección del centro puesto que, en la realidad que conocemos, estas actitudes suelen venir determinadas por elementos exógenos variados. Pero si contemplamos esta posibilidad, se podrían citar dos posibles estrategias de cara a ofrecer una buena imagen de su actividad escolar a los padres de los alumnos o incluso a la sociedad en general. La primera de ellas sería aplicar un bajo nivel de exigencia en sus propios exámenes, con lo que se podría vender la idea de que prácticamente todos sus alumnos completan el Bachillerato, sin hacer referencia a que después posiblemente obtengan en general unos malos resultados en la PAU. La segunda estrategia sería el caso contrario, elevar considerablemente el nivel de formación necesario para aprobar los exámenes de la propia escuela. En esta posibilidad, la idea a vender sería la de que todos los alumnos de su centro que acuden a la PAU obtienen unos grandes resultados, aunque en ese caso no se informaría sobre la posiblemente mayor tasa de suspensos en los exámenes previos. Sea en un caso o en otro, la definición que empleemos de output debe intentar ser neutral ante este tipo de conductas o al menos no verse influida de forma significativa.

Tras revisar la situación a la que nos enfrentamos en nuestro país, debemos estudiar la adaptación teórica a la misma de las alternativas de las que disponemos. Para ello hemos planteado las siguientes hipótesis de trabajo:

A) Sigue siendo ineludible el empleo de dos variables de output que midan **cantidad y calidad** del mismo, respectivamente. Esta idea se refuerza en nuestro caso debido a las consecuencias reales del examen PAU, que nos permite plantear dos argumentos en defensa del empleo simultáneo de ambas. El primero ya ha sido apuntado en cuanto a la heterogeneidad en la estrategia de los centros y su posible especialización en una sola dimensión. En el análisis empírico esta posibilidad se contrastará mediante el empleo de cada indicador de forma alternativa y por separado, lo que nos informará sobre **si se da o no realmente esta especialización** en una parte significativa de las escuelas (aquellas que presenten diferencias importantes en su clasificación). Además es fundamental tener presente que si en la práctica futura los centros fueran conscientes de ser evaluados a través de un solo indicador, la potencial especialización actual que podamos detectar se generalizaría en mayor medida.

Pero además esta hipótesis se refuerza con el hecho de que en la PAU cada una de ellas refleja dos objetivos igualmente decisivos para el futuro de cada estudiante. Dada la alta tasa de aprobados en el examen, la nota finalmente conseguida se torna un elemento fundamental en la probabilidad de que los escolares puedan acceder a la titulación universitaria que escojan como primera preferencia. Ello nos lleva a considerar que para la mayoría de los estudiantes que se

presentan a la PAU, tan importante es aprobar como conseguir una nota elevada, lo que legitima determinar ambos objetivos como parámetros de los resultados de las escuelas.

B) Mientras no existe alternativa en nuestra opinión a la definición de la variable de calidad (“nota media de los aprobados en el examen PAU”), la variable de cantidad presenta dos opciones, partiendo del empleo de un cociente que nos señale el porcentaje de aprobados en la prueba PAU. La dualidad radica en cómo cuantificar el denominador del cociente. La elección no será neutral, puesto que en cada caso se dará una solución distinta a la heterogeneidad de las escuelas en dos aspectos clave: el porcentaje de su alumnado que no tiene aspiraciones universitarias (y que por tanto no piensa presentarse al examen PAU) y el nivel de exigencia en la calificación de sus propias pruebas (como filtro previo para acceder a la PAU).

La primera posibilidad consiste en definir el porcentaje de aprobados **respecto a los alumnos presentados a la prueba PAU**. Su ventaja más importante radica en su simplicidad, puesto que se trata de evaluar a cada escuela en atención únicamente a sus resultados en este examen. Al tener en cuenta sólo a los presentados a la prueba, no se contabilizan ni se diferencia entre aquellos alumnos que hayan optado por acceder a un Ciclo Formativo (lo que no es resultado de la actuación de la escuela) y aquellos otros que hayan abandonado los estudios durante el curso o suspendido el Bachillerato (que sí debería ser indicador de mala actuación productiva). Por tanto, se evita que los resultados finales se vean sesgados por el porcentaje de alumnos en cada centro interesados por ciclos formativos, dado que no se incluyen en la definición de la variable. Como contrapartida, se renuncia a descontar una potencial conducta estratégica de los centros en el filtro de los alumnos que acceden al examen. Por tanto, se asume que no es significativo el número de escuelas que fija un filtro elevado en sus propios exámenes, que son las que se verían claramente beneficiadas por el empleo de este indicador. Una desventaja añadida de esta posibilidad radica en la baja dispersión de la tasa de aprobados en estas pruebas respecto a los presentados (suelen rondar en torno al 90% como media), con lo que queda poco margen a la variación de resultados entre las distintas escuelas evaluadas.

La hipótesis clave a contrastar en este caso como defensa de esta posibilidad sería la **no existencia de estos comportamientos estratégicos, sea en un sentido u otro**, en un número significativo de centros educativos. En el análisis emplearemos dos señales para la detección de una actuación estratégica (dependiendo del sentido de la misma). La primera será la obtención de una buena clasificación de la escuela en el ranking que emplee el volumen de presentados, unido a una mala posición en el ranking que considere el número de matriculados a inicios del curso como denominador del cociente. Ante un dato de este tipo sólo caben dos explicaciones: o la escuela aplica la estrategia de filtrar en exceso los alumnos que se presentan o el porcentaje de alumnos interesado por la PAU es muy pequeño. Por tanto una doble valoración como la anterior unida a un porcentaje alto o medio de alumnos con aspiraciones universitarias será la

primera señal lógica que empleemos de conducta estratégica. Y a la inversa, la segunda señal consistirá en la obtención de una mala clasificación según el número de presentados combinada con una buena valoración si se usa el volumen de matriculados. Las dos explicaciones a este hecho serían o una estrategia de filtro no exigente en la escuela para que casi todos los que lo deseen se presenten a la PAU o un porcentaje de alumnos interesados por estudios universitarios muy elevado. Por tanto, la doble clasificación anterior sumada a un porcentaje medio o bajo de alumnos con pretensiones universitarias será la segunda señal que emplearemos.

C) La segunda opción como definición de la variable de cantidad será el porcentaje de aprobados en la PAU **respecto a los matriculados a inicio del año académico en el curso previo a la misma**. Esta posibilidad se torna más complicada y ambiciosa a priori que la anterior, puesto que al situarnos en el inicio (matriculados) y final (aprobados PAU) del proceso temporal analizado, es la que potencialmente se verá afectada en mayor medida por la aparición de posibles sesgos en el transcurso del mismo. En este caso sí que se da una solución a la posibilidad de comportamientos estratégicos, puesto que su aparición no afectará a los resultados de la evaluación final al no considerarse el número de alumnos presentados a la prueba como referencia. Por tanto, se considera fracaso tanto que un alumno matriculado a inicio del curso no se presente a la PAU como que se presente y la suspenda, dejando sin influencia el potencial filtrado del centro de cara a la PAU. El contrapunto a conseguir la eliminación de este efecto será que se ignore la desigualdad entre los centros respecto al porcentaje de alumnos que no estén interesados por la PAU. Así, el hecho de que un alumno no aspire a estudios universitarios (y por tanto no se presente al examen aunque posea formación suficiente para superarlo) es asimilado a priori con los restantes casos de fracaso escolar (los que no aprueban el Bachillerato o los que suspenden la PAU). Asimismo, en esta posibilidad no existe la desventaja de baja dispersión de los resultados comentada en la opción anterior, dado que al ampliarse el período temporal analizado (y multiplicarse los efectos considerados) los valores conseguidos por las escuelas tienen un espectro mucho mayor de variación.

Por tanto, en este caso la hipótesis a contrastar en defensa de esta definición de la variable será la **no discriminación de aquellos centros con un porcentaje elevado a alumnos interesados en ciclos formativos**. El indicador empírico que emplearemos para la detección de esta infravaloración será comprobar si, de entre las escuelas con mayor porcentaje de alumnos sin aspiraciones universitarias, existe un número significativo de las mismas que obtengan una buena valoración en la clasificación que emplea el volumen de presentados combinado con una mala ordenación si se emplea el número de matriculados. Una valoración de este tipo sólo puede implicar o una estrategia de filtrado exigente en el acceso a la prueba (lo que no tendría influencia con el empleo de esta definición de output) o que la escuela posea un elevado

porcentaje de alumnos no interesados en la PAU, en cuyo caso sí tendremos indicios de que la discriminación anterior aparece con el empleo de esta variable.

4. EL ANÁLISIS EMPÍRICO. RESULTADOS.

Las unidades escolares evaluadas en esta investigación son 215 Institutos de Enseñanza Secundaria públicos localizados en las Comunidades Autónomas de Aragón (75), Asturias (55) y Extremadura (85) que ofertaban 2ª Bachillerato durante el curso académico 2001/2002. Como inputs controlables por las escuelas se han seleccionado los habitualmente empleados en la literatura, relacionados tanto con aspectos presupuestarios como de personal docente: GASTO, definida como gastos de funcionamiento durante el año 2001 por alumno matriculado (excluidas las partidas relativas a edificios y otras construcciones) y PROFESORES, definida como el número total de profesores del centro por cada 100 alumnos matriculados.

Respecto a los inputs no controlables por el centro, centrados en este caso en las características socioeconómicas y familiares de los estudiantes de cada escuela, la inexistencia de estadísticas oficiales al respecto obligó a los autores a realizar un trabajo de campo consistente en una encuesta entre los alumnos de último curso de todas las escuelas evaluadas, realizada durante los meses de marzo y abril de 2002 y que obtuvo una tasa de respuesta cercana al 80%. El cuestionario estaba formado por 54 preguntas relativas al estatus socioeconómico y familiar de cada estudiante, con cuya información se construyeron del orden de una veintena de variables, gran parte de ellas previsiblemente correlacionadas entre sí. De cara a seleccionar las que finalmente se emplearían en los análisis DEA, se aplicó el criterio de identificar aquellas variables que ofrecieran el máximo de información no redundante (es decir, correlacionadas significativamente con el resto de variables no seleccionadas pero no entre sí). Finalmente, los dos inputs no controlables seleccionados fueron BNOTAS (porcentaje de alumnos con nota media de notable o superior en el curso anterior) e INGRESOS (porcentaje de alumnos cuyos padres tienen unos ingresos mensuales superiores a 1.800 euros).

En cuanto a las variables de output a emplear y que son nuestro punto de atención en este documento, vamos a manejar las tres opciones comentadas en la exposición del problema: NOTA (nota media de los aprobados en la PAU), PRES (porcentaje de aprobados en la PAU respecto a los presentados a la prueba) y MATR (porcentaje de aprobados en la PAU respecto a los matriculados a inicio del curso académico en 2º de Bachillerato).

Además y completando la enumeración de los datos que utilizaremos, haremos referencia a los valores que alcanzan los alumnos de cada escuela en la pregunta de la encuesta relacionada con sus aspiraciones universitarias (ASPIR). Más concretamente, esta variable viene dada por el porcentaje de alumnos encuestados en cada escuela que responde afirmativamente a la pregunta “¿Deseas cursar estudios universitarios?”.

En cuanto a los aspectos metodológicos, se han respetado los supuestos básicos para la evaluación de la eficiencia de las escuelas que hemos venido repasando. De este modo, para considerar de forma adecuada en los DEA las características de los alumnos se ha empleado el modelo desarrollado por Muñiz (2002) para la inclusión de inputs no controlables en un análisis a través de tres etapas DEA. Se ha empleado la opción de minimización del input con rendimientos variables de escala (modelo BCC) y asimismo, para respetar la necesidad de un mismo examen para todas las escuelas evaluadas se han repetido, para cada una de las opciones alternativas de variables, los mismos análisis por separado para cada una de las tres CC. AA.. Esta aplicación de los distintos modelos por separado para cada zona geográfica nos permitirá además dotar de mayor consistencia a las conclusiones finales que se alcancen.

Se realizaron cinco DEAs para cada Comunidad Autónoma: tres con el empleo alternativo de una variable de output en cada uno (NOTA, MATR y PRES) y otros dos conjugando la variable NOTA con MATR o PRES, respectivamente. A las cinco clasificaciones obtenidas se suma la derivada de la pregunta sobre las aspiraciones universitarias de los alumnos (ASPIR), lo que nos ofrece en total seis clasificaciones para cada CC. AA.

Además para cada hipótesis y de forma previa al comentario de los resultados de los análisis DEA, puede resultar interesante comprobar si de la simple observación de los datos de output de cada escuela pueden extraerse conclusiones relativas a la importancia real de las hipótesis consideradas, es decir, si tienen sustento en la práctica educativa real. Esta exploración previa nos informará sobre la pertinencia de haber considerado que podían darse casos de especialización cantidad/calidad, estrategias en el filtro previo a la PAU, etc. o si por el contrario, dichas conductas son simples supuestos virtuales. De este modo, aseguramos un doble contraste de cada hipótesis: si existe el problema, y posteriormente, si tiene consecuencias en la evaluación DEA de la eficiencia de las escuelas. Como recordatorio previo, exponemos en el Esquema 2 las hipótesis a contrastar y las señales para su verificación y consecuencias.

La primera hipótesis a contrastar se basaba en la especialización potencial de los centros en una sola dimensión del output, cantidad o calidad. La observación de los ranking de outputs permite detectar la presencia real de especialización en un número significativo de centros, aunque con ligeras variaciones entre las tres CC. AA. En Aragón es donde esta conducta se detecta con un carácter más habitual, de tal modo que la mitad de los centros que obtienen una alta posición en NOTA obtienen clasificación baja en al menos una de las dos posibilidades de cantidad (MATR o PRES), lo que también sucede a la inversa para la mitad de escuelas con mala clasificación en NOTA. Respecto a los casos individuales, la especialización en calidad más consistente se da en los centros *Matarraña, Ramón J. Sender, Sobrarbe y Zurita*. Lo mismo se puede decir, pero especializados en cantidad, de los centros *Juan de Lanuza, Valle del Jiloca, Cinco Villas, Sierra del Guara y Mar de Aragón*. En Extremadura la especialización

alcanza a un tercio de las escuelas, ya sea en cantidad o en calidad. A nivel individual destacan los centros *Maestro Domingo Cáceres*, *Gregorio Marañón* y *Castelar* especializados consistentemente en calidad, al igual que *Francisco de Orellana* y *El Pomar* en cantidad. Finalmente, en Asturias es la región donde se detecta menor especialización, alcanzando aún así al 25% de las escuelas en uno u otro sentido. Los centros *Montevil* y *Luanco* destacan en su especialización en calidad, mientras que ninguna escuela mala en calidad destaca como buena en las dos clasificaciones de cantidad simultáneamente. Resumiendo de forma global, los porcentajes detectados permiten considerar la especialización en una sola dimensión del output como una conducta habitual en un número significativo de escuelas, siendo por tanto completamente pertinente el estudio de la primera hipótesis.

HIPÓTESIS	VARIABLE OUTPUT AFECTADA	SEÑALES EXISTENCIA PREVIA PROBLEMA	VERIFICACIÓN CONSECUENCIAS EN MODELOS DEA
A) ESPECIALIZACIÓN EN UNA SOLA DIMENSIÓN DEL OUPUT	EMPLEO AISLADO DE <i>NOTA</i> , <i>PRES</i> O <i>MATR</i>	ESCUELAS CON ALTA POSICIÓN EN <i>NOTA</i> Y BAJA EN <i>PRES</i> Y/O <i>MATR</i> (ESPECIALIZACIÓN CALIDAD) ESCUELAS CON BAJA POSICIÓN EN <i>NOTA</i> Y ALTA EN <i>PRES</i> Y/O <i>MATR</i> (ESPECIALIZACIÓN CANTIDAD)	ESCUELAS CON ALTA VARIACIÓN ÍNDICE EFICIENCIA ENTRE <i>NOTA</i> Y <i>MATR</i> Y/O <i>PRES</i>
B) EXISTENCIA COMPORTAMIENTO ESTRATÉGICO EN FILTRO PREVIO PAU	<i>NOTA+PRES</i>	ESCUELAS CON ALTA EN <i>PRES</i> , BAJA EN <i>MATR</i> Y ALTA-MEDIA EN <i>ASPIR</i> (FILTRRO ELEVADO) ESCUELAS CON BAJA EN <i>PRES</i> , ALTA EN <i>MATR</i> Y MEDIA-BAJA EN <i>ASPIR</i> (FILTRRO REDUCIDO)	ESCUELAS CON ALTA VARIACIÓN ÍNDICE EFICIENCIA ENTRE <i>NOTA+PRES</i> Y <i>NOTA+MATR</i> Y POSICIÓN RELEVANTE EN <i>ASPIR</i>
C) DISCRIMINACIÓN ESCUELAS ALUMNOS NO ASPIRAC. UNIV.	<i>NOTA+MATR</i>	ESCUELAS CON ALTA EN <i>PRES</i> , BAJA EN <i>MATR</i> Y BAJA EN <i>ASPIR</i> (DISCRIMINACIÓN SI USO <i>MATR</i>)	ESCUELAS CON ALTA VARIACIÓN POSITIVA ÍNDICE EFICIENCIA EN <i>NOTA+PRES</i> RESPECTO A <i>NOTA+MATR</i> Y POSICIÓN RELEVANTE EN <i>ASPIR</i>

Esquema 2. Hipótesis a contratar y señales para su potencial verificación.

(Nota: Para la señal de existencia del problema se considera una posición buena estar situado en el 33% de centros mejor clasificados, posición media en el tercio siguiente y así sucesivamente, y para la verificación de las consecuencias DEA una alta variación del índice de eficiencia la superior a +/- 10%)

Sin embargo, la conclusión a la que llegamos tras la comparación de las clasificaciones obtenidas en los tres DEAs realizados con una sola variable de output de forma alternativa es

que el efecto de la especialización anterior se diluye al analizarla en el contexto de una evaluación de la eficiencia, pero con una prevención que pasamos a comentar. Tanto en Aragón como en Extremadura son casos puntuales (en torno al 5% de los centros) los de aquellas escuelas que presentan una alteración superior al 10% en su índice de eficiencia dependiendo de que se considere solo cantidad o solo calidad. En cambio en Asturias, la mitad de las escuelas evaluadas presenta una alta variación de sus índices de eficiencia en tal caso, siendo además tal oscilación muy elevada para varias escuelas (sirva como ejemplo *Tineo*, índice de eficiencia del 100% con la variable de calidad y del 61% en los dos casos de cantidad). La explicación a este resultado, teniendo en cuenta que el grado de especialización en Asturias no era más elevado que en las restantes Comunidades, se puede encontrar en el número más reducido de escuelas evaluadas en el Principado de Asturias. Podemos concluir por tanto en relación a la primera hipótesis que la especialización en una dimensión, que sí aparece como conducta en un número importante de escuelas, no afecta de forma relevante a los índices de eficiencia siempre que se cuente con un número relativamente elevado de unidades productivas en el análisis. Con todo y dada la consistencia de los argumentos teóricos a favor de la consideración simultánea de las dos dimensiones, el empleo de una variable de cantidad y otra de calidad en el contexto evaluado es un aspecto recomendable, aunque su no observación no invalida de forma importante los resultados finales del estudio de eficiencia, siempre que se respete la condición relativa al número de unidades evaluadas.

La segunda hipótesis de nuestro estudio consistía en observar las actitudes estratégicas de las escuelas en relación con su propio filtro en el momento de decidir quien aprueba el Bachillerato. Observando en primera instancia los rankings de outputs y aspiraciones relevantes para esta hipótesis comprobamos como, tanto para Aragón como para Extremadura, una tercera parte de las escuelas con alta clasificación en PRES pasa a una mala ordenación en MATR. En cambio, en Asturias esta variabilidad es menor (apenas un 20% de las escuelas), debido quizá (y teniendo en cuenta que también en la primera hipótesis era la CC. AA. con menor presencia del problema analizado) al rasgo mencionado de que sea la Comunidad para la que contamos con menor número de unidades. Además un requisito añadido para detectar un comportamiento estratégico estaba relacionado con la posición de la escuela en el ranking de aspiraciones universitarias. En Aragón, de las escuelas claramente favorecidas en PRES, dos tercios alcanzan una calificación alta-media en ASPIR, lo que implica que son escuelas que fijan un filtro elevado para su alumnado. Destacan en este sentido los centros *Miguel de Molinos*, *Santiago Hernández*, *Grande Covián* y *Bajo Aragón*. Del mismo modo, la mitad de las escuelas beneficiadas en PRES en Extremadura cumple al tiempo el requisito en ASPIR que es sinónimo de filtro elevado, destacando las escuelas *Bembezar*, *Emérita Augusta*, *Sáez de Buruaga*, *Eugenio Hermoso* y *Cieza de León*. Y finalmente en Asturias, aunque ya se ha comentado que

es menor la presencia de escuelas con alta variación a favor de PRES, las tres cuartas partes de las mismas también cumplen el requisito en ASPIR de filtro exigente, destacando en ello *Emilio Alarcos, La Luz y Pravia*. Resumiendo la exploración previa y respecto a la presencia de comportamiento estratégico de filtro elevado, podemos concluir que del tercio de escuelas favorecidas claramente si se toma la variable PRES, la mayoría lo son por emplear un filtro riguroso en el acceso a la PAU y no por tener pocos alumnos con aspiraciones universitarias.

Un porcentaje prácticamente idéntico se detecta si repetimos los cálculos respecto a la presencia de estrategias de filtro reducido. En esta posibilidad las escuelas favorecidas relativamente en MATR comparadas con PRES representan de nuevo un tercio en Extremadura y Aragón y el 20% en Asturias. Pero en este caso es aún más claro que en el anterior que la posible causa para dicha posición barajada como alternativa a la estrategia (en este caso un alto porcentaje de alumnos con aspiraciones universitarias) tiene una presencia menor (25% de las escuelas en Aragón y Extremadura, aunque en Asturias y como excepción, sí sería significativa como causa predominante). Como ejemplos de escuelas que aplican filtros reducidos podemos destacar *Santa Eulalia, Reino Aftasí, La Serena y Luis Chamizo* en Extremadura, *Joaquín Costa, Río Gallego, Lucas Mallada y Andalán* en Aragón o *Monte Naranco* en Asturias.

Por tanto, queda demostrada una cierta presencia de comportamientos estratégicos en cuanto al filtro propio de cada escuela (sumando las dos posibles opciones de estrategia y generalizando para las tres CC. AA, estaríamos hablando de un proporción de escuelas “estratégicas” en torno al 25-30% del total) y con ello la pertinencia de la hipótesis. Pero si confrontamos los resultados de las evaluaciones DEA llevadas a cabo, de nuevo obtenemos una conclusión esperanzadora, como es que no se detecta, salvo casos aislados, indicios de que estos comportamientos hayan afectado a la evaluación realizada. En las tres regiones estudiadas, los casos en que hay una alta variación en los índices de eficiencia asociados a cada escuela entre emplear *NOTA+PRES* o *NOTA+MATR* no superan el 10% del total, no llegando en Aragón y Extremadura al 5%. En los casos aislados detectados de variación sí se observa que ésta se debe a comportamientos estratégicos (frente a la explicación alternativa relacionada con el porcentaje de aspiraciones universitarias en uno u otro sentido), pero no dejan de ser escuelas puntuales y como tal, podemos considerarlas como la excepción a la norma. Como tales podemos citar las oscilaciones detectadas en los centros *San Alberto Magno* en Aragón o *Colunga y Luarca* en Asturias.

Pasando ya a la última hipótesis enunciada, ésta consistía en que las escuelas que pasan de una alta clasificación en PRES a una mala en MATR deben dicha conducta a poseer un bajo porcentaje de alumnado con aspiraciones universitarias (lo que caso de cumplirse invalidaría el empleo de la variable MATR). El análisis de las clasificaciones previas muestra la escasa incidencia de dicha discriminación. Esto es lógico si se tiene en cuenta que una estrategia de

filtro elevado o pocos alumnos con pretensiones universitarias eran causas alternativas y excluyentes entre sí del mismo efecto en las clasificaciones. Y la presencia de las mencionadas estrategias ya se comprobó en la hipótesis anterior. Así y en consonancia con lo anterior, esta explicación de discriminación sólo está presente en un tercio de las escuelas favorecidas en PRES en Aragón, en la mitad de las mismas en Extremadura y apenas en la cuarta parte de las que están en esta situación en Asturias. A nivel individual, podemos mencionar como casos de escuelas que sí sufrirían esta discriminación los de *Los Enlaces* y *Santa Emerenciana* en Aragón, *Virgen del Soterraño*, *Valle del Ambroz* y *Agora* en Extremadura. Como es lógico y dado que la potencial discriminación no aparece en la práctica cotidiana de forma significativa, tampoco tiene efecto sobre los índices de eficiencia obtenidos en las evaluaciones DEA.

5. CONCLUSIONES.

En la presente investigación se han revisado los problemas relacionados con la medida empírica del output educativo, en el contexto de una evaluación de la eficiencia productiva de un conjunto de escuelas. Partiendo de las dificultades generales asociadas a este tipo de análisis, nos hemos centrado fundamentalmente en las cuestiones a resolver para el caso español. En nuestro país contamos hasta el momento con una sola opción de examen común para todos los centros, la Prueba de Acceso a la Universidad, cuyos resultados sin embargo no reflejan de forma global la formación conseguida sobre la totalidad de los alumnos de cada escuela, sino tan sólo sobre el colectivo interesado en iniciar estudios universitarios. Los resultados que se perseguía obtener con el estudio hacen referencia a dos cuestiones paralelas, detectar si en la práctica cotidiana de las escuelas aparecían una serie de conductas problemáticas, con el fin de contrastar la pertinencia de analizar sus consecuencias y comprobar la consistencia del Análisis Envoltante de Datos ante dichas conductas, para observar cómo distintas especificaciones alternativas del modelo podían o no dar solución a la aparición de estos comportamientos

Para ello se plantearon tres hipótesis de trabajo, cuyas conclusiones finales exponemos a continuación por separado:

A) La especialización de las escuelas en una sola dimensión, cantidad o calidad, del output educativo. Esta conducta se ha detectado en aproximadamente un tercio de las escuelas evaluadas, lo que exige que la evaluación DEA tenga en cuenta a priori en su desarrollo metodológico la posible especialización de los centros. Sin embargo, en los análisis llevados a cabo ignorando esta dualidad, los resultados de la evaluación no se ven afectados de forma significativa en dos de las tres CC. AA. consideradas. Sí se observa una fuerte variación de los índices de eficiencia en el caso de Asturias, aunque dado que las escuelas asturianas no presentaban una mayor especialización previa que las restantes, la explicación puede pasar por el menor número de unidades disponibles para la evaluación en esta región. Por tanto, se ha

demostrado la existencia de especialización cantidad/calidad en una parte significativa de las escuelas y con ello, la relevancia teórica de considerar ambos aspectos en la evaluación, aunque a nivel empírico los índices de eficiencia DEA no se verían sesgados por esta conducta salvo que se cuente con un número reducido de unidades en la evaluación.

B) La conducta estratégica de los centros educativos en sus propios exámenes de Bachillerato, considerados como un filtro interno previo a la Prueba de Acceso a la Universidad. La consecuencia de que pudiera aparecer tal conducta sería poner en duda la validez de que se cuantifique la variable de cantidad respecto a los presentados a la PAU en lugar de respecto a los matriculados a inicio del curso. De nuevo se detecta en los datos esta conducta en una cantidad significativa de escuelas (en torno a la cuarta parte de las mismas), bien sea en forma de filtro exigente o reducido. Sin embargo, los resultados de la evaluación DEA no se ven afectados por la existencia real de actitudes de este tipo, de modo que los índices de eficiencia asignados no presentan oscilaciones cuantitativamente importantes dependiendo de que se emplee o no esta posibilidad para definir la variable de cantidad.

C) La discriminación o infravaloración en la evaluación de las escuelas que posean un porcentaje bajo de alumnos con aspiraciones universitarias. Esta hipótesis se relacionaba con la conveniencia o no de definir la variable de cantidad respecto a los matriculados a inicio de curso. Los datos no muestran indicios empíricos de que se aparezca esta posibilidad, ni en el análisis previo de los outputs ni por tanto en los resultados de los modelos DEA, lo que llevaría a rechazar la crítica principal al uso específico de esta definición.

Resumiendo finalmente las conclusiones de la investigación, se ha justificado la pertinencia de emplear una variable de cantidad y otra de calidad para evitar un posible sesgo por la especialización real de las escuelas, aunque en términos empíricos la influencia sobre los índices DEA no es significativa salvo que se cuente con pocas unidades en la evaluación. En cuanto a la definición de la variable de cantidad respecto a los matriculados o respecto a los presentados a la PAU, es una dicotomía que no afecta a la consistencia de los resultados finales de la evaluación, dada la escasa variación entre los índices de eficiencia obtenidos en ambos casos. No obstante y dado que son alternativas excluyentes, la aparición real de conductas estratégicas de los centros, junto a la no identificación de discriminación en las escuelas con menor porcentaje de aspiraciones universitarias, inclinan la balanza a favor de la opción que tiene en cuenta el número de matriculados a inicio de curso. Es decir y ante la ausencia de sesgos empíricos significativos que nos decanten por una u otra opción, esta posibilidad da respuesta teórica a un problema detectado en la conducta real de las unidades evaluadas, mientras que la ventaja relativa más importante a priori de la opción alternativa hace referencia a una pauta discriminatoria que no se ha detectado en la práctica.

BIBLIOGRAFÍA

- Blaug, M. (1981). *Educación y empleo*, Instituto de Estudios Económicos, Madrid.
- Chubb, J.E. y Moe, T.M. (1990). *Politics, markets and american schools*, The Brooking Institution, Washington, D.C.
- Cohn, E. y Millman, S.D. (1975). *Input-output analysis in Public Education*, Cambridge, Ballinger Pub. Company.
- Glasman, N.S. y Biniaminov, S. (1981). "Input-output análisis of schools", *Review of Educational Research*, 51(4): 509-539.
- Hanushek, E.A. (1986). "The economics of schooling: production and efficiency in public schools", *Journal of Economic Literature*, 24 (septiembre): 1141-1177.
- Hanushek, E.A. (ed.) (2003). *The Economics of schooling and school quality*, Vols. I y II. Edward Elgar Publ. Ltd., Reino Unido.
- Madaus, G.F., Kellaghan, T. Rakow, E.A. y King, D.J. (1979). "The sensitivity of measures of school effectiveness", *Harvard Educational Review*, 49(2): 207-230.
- Mancebón, M.J. (1996). *La evaluación de la eficiencia de los centros educativos públicos*, Tesis doctoral de la Universidad de Zaragoza.
- Mancebón, M.J. y Muñiz, M.A. (2003). "Aspectos clave de la evaluación de la eficiencia productiva en la Educación Secundaria", *Papeles de Economía Española*, 95: 162-187.
- Muñiz, M.A. (2002). "Separating managerial inefficiency and external conditions in Data Envelopment Análisis", *European Journal of Operational Research*, 143(3): 625-643.