ANÁLISIS DE EFICIENCIA DE CENTROS EDUCATIVOS

Carlota M. Domínguez Gil

Memoria para optar al título de doctora

Programa de Doctorado en Administración y Dirección de Empresas

Universidad Pablo de Olavide

Noviembre, 2020

Director:
Dr. Ignacio Contreras Rubio

Agradecimientos

Quiero agradecer en primer lugar a mi director, el Dr. Ignacio Contreras, por su paciencia y apoyo en este largo proceso. Su optimismo y creatividad han sido una verdadera inspiración y me han mantenido en el camino cuando sentía tentaciones de abandonar. A la Dra. Manuela Segovia, compañera y co-autora de dos de los artículos, por su inestimable ayuda. A los Dres. Flor Guerrero y Miguel Ángel Hinojosa, quienes me dieron la oportunidad de trabajar en la Universidad Pablo de Olavide iniciando así el proceso que me ha llevado hasta aquí.

A mi marido, Miguel, y mis hijas, Carlota y Natalia, por su comprensión con el tiempo que tenía que sacrificar de estar con ellos por dedicarme a la tesis y por celebrar conmigo las publicaciones, aunque no entendieran muy bien de qué se trataba.

Y por último, a mis padres, su ilusión ha sido otro motivo para querer ser doctora. Ellos me enseñaron a valorar el conocimiento por sí mismo y no por su utilidad. Llenaron mi infancia de libros y conversaciones, que seguro están presentes de algún modo en esta tesis.

Índice general

Índice de tablas

Índice de figuras

Introducción

"La educación es el arma más potente que puedes usar para cambiar el mundo"

Nelson Mandela

La educación es una de las claves del desarrollo humano y económico de las sociedades. Mayores niveles de educación se identifican con incrementos en la calidad de vida de los ciudadanos. Así, este factor es tenido en cuenta en multitud de indicadores como el Índice de Desarrollo Humano de la Organización de Naciones Unidas.

Este hecho se ve agravado por el auge de las nuevas tecnologías y la automatización en lo que se ha dado en llamar la 4^a Revolución Industrial. Según el informe *El futuro de los trabajos 2018* (*The Future of Jobs 2018*), del Foro Económico Mundial, desaparecerán 75 millones de empleos y surgirán otros 133 millones de nuevos roles. Entre las profesiones condenadas a extinguirse encontramos aquellas con tareas repetitivas y que no requieren cualificación, por lo que la formación se vuelve indispensable para no quedar fuera del mercado laboral.

Este consenso acerca de su importancia no se traduce, sin embargo, en estrategias comunes ni en acuerdos generales sobre cómo alcanzar esas mayores cotas de formación. Esto se debe, entre otros factores, a su compleja naturaleza, su carácter acumulativo y el hecho de que el papel del individuo es crucial en su propio aprendizaje.

Ante esta complejidad, resulta necesario abordar la cuestión desde distintas perspectivas, de forma que cada enfoque aporte su propia visión. Una de ellas es la Economía de la Educación, que consiste en aplicar técnicas y perspectivas económicas a la educación. Así, podemos entender ésta como un proceso productivo en el que los estudiantes son transformados mediante su paso por el sistema educativo. En este proceso de transformación intervienen distintos agentes, lo que se corresponde con los distintos enfoques que se pueden adoptar, desde los sistemas educativos establecidos por los países, hasta llegar al individuo o agente involucrado en el proceso de aprendizaje.

Entre las ventajas de esta perspectiva se encuentra la objetividad de sus métodos. El uso de técnicas matemáticas y estadísticas, permite evaluar el desempeño de los distintos sistemas o centros educativos sin sesgos introducidos por el investigador. La posibilidad de aplicar esta técnica tan versátil al campo en el que he desarrollado mi labor profesional los últimos catorce años me pareció apasionante. Proveer a quienes han de tomar decisiones de datos fiables y objetivos en este campo, la educación, me parecía clave para la mejora y el desarrollo especialmente de nuestro país.

Recientemente, hemos asistido a la aprobación de una nueva Ley de Educación, la octava de nuestra democracia. Esto nos demuestra, más allá de cuestiones políticas o ideológicas, que se trata de un tema fundamental para nuestra sociedad, pero también que no se termina de conseguir un sistema educativo de calidad, eficiente y adecuado al siglo XXI. A ello se añade un persistente debate sobre la importancia de controlar el gasto público, iniciado en la crisis de 2006 y que continuará, sin duda, en la actual crisis. Este debate se extiende y centra en uno de los principales conceptos de gasto de las cuentas nacionales, como es el gasto en educación. Actualmente el gasto en educación, supone el cuadro rubro de las cuentas públicas en nuestro país (por detrás de el gasto en protección social, salud y servicios públicos generales), con un gasto superior a los 47.000 millones, lo que supone alrededor del 10 % del gasto público total y, aproximadamente, un 4.2 % del PIB.

Por un lado, parece clara la correlación entre el crecimiento y el desarrollo social con el nivel del *capital humano*. Ello supone, sin duda, un incentivo a la inversión en educación para las economías. Pero, por otro lado, se alzan voces respecto a la necesidad de reducir los crecientes déficits públicos en un entorno de crisis económica y sobre la necesidad de revisar cada unidad monetaria que se gasta en el sector público.

En este contexto, el concepto de *eficiencia de los sistemas educativos* se torna crucial. Se mantiene la exigencia de sistemas educativos de calidad, universales y que cumplan con los mínimos exigidos en economías avanzadas. Pero, por otro lado, se mantiene la exigencia a los políticos de minimizar o, al menos, optimizar los recursos económicos dedicados a tal fin. Se demandan resultados excelentes a la vez que se mantienen limitados, y en muchos casos se reducen, los recursos disponibles para tal fin.

De manera sintética, el estudio no puede centrarse exclusivamente en evaluar los resultados de las instituciones o sistemas educativos. Esta análisis ha de realizar se de manera paralela y vinculada con el análisis de los recursos y de uso que se hace de éstos, del desempeño de las unidades que intervienen en el proceso y del aprovechamiento que se realiza de dichos recursos.

El estudio de la Economía de la Educación ha recibido una atención creciente en las últimas décadas. Desde el punto de vista de la economía de la educación, la educación es considerada como un proceso de producción en la que se utilizan diferentes entradas (inputs) para producir diferentes resultados (outputs), para una tecnología dada. La base teórica de esta propuesta se basa en los estudios de Levin [18] y Hanushek [14], en la que proponen que para entidad i y para un período temporal dado t, puede definirse una función de producción compleja en la que las entradas se dividen en cinco categorías que tratan de medir el contexto familiar, los recursos de la institución, las características del alumnado y las posibles influencias externas.

La función puede representarse como:

$$A_i(t) = g\left[F_i(t), S_i(t), P_i(t), O_i(t), I_i(t)\right]; \tag{1}$$

donde

$A_i(t)$ representa el vector de salidas o resultados del proceso educativo.

$F_i(t)$ representa el vector características individuales y antecedentes familiares.

$S_i(t)$ representa el vector de entradas de la institución (recursos materiales, humanos y financieros).

$P_i(t)$ representa el vector de características de los estudiantes.

$O_i(t)$ representa el vector de otras influencias externas.

$I_i(t)$ representa el vector de recursos iniciales o innatos del estudiante.

A la vista de la función anterior, queda claro que cuantificar la educación recibida por un individuo no es una tarea simple. Muchos de los aspectos que deben considerarse son intangibles y, además, es necesario considerar diferentes años para cuantificar el resultado final. En [19], el autor pone de manifiesto como esta intangibilidad, el carácter acumulativo y el hecho de que sean los propios usuarios (estudiantes) los que desarrollan el proceso añaden una gran dificultad a la estimación, si lo comparamos con otros procesos productivos. Una de las consecuencias de estos elementos diferenciales ha hecho que, de manera consensuada, se haya terminado recurriendo al uso de pruebas estandarizadas para medir, de alguna manera, los resultados educativos.

En este contexto, las técnicas de análisis paramétricas, como pueden ser modelos de regresión que buscan estimar la función de producción, presentan ciertas limitaciones y una enorme dificultad para ser aplicadas en aplicaciones empíricas. Esta familia de procedimientos, parte de asumir una determinada forma para la función que se desea estudiar. A partir de ésta, se estudia cómo determinar los parámetros de dicha función con los valores observados con los que se cuenta. La primera decisión, asumir una u otra forma para la función que se desea aplicar, determinará por tanto el proceso completo. Y, en este caso particular, es difícil encontrar un modelo adecuado que conecte unas entradas y salidas con las características tan particulares como las descritas anteriormente.

Por contra, los modelos nos paramétricos no imponen una forma funcional concreta en la evaluación del proceso. Se parte de construir un conjunto de posibilidades de producción a partir de las observaciones, lo que permite incluir múltiples tipos de variables. Esta característica es especialmente para representar las particularidades del proceso de producción educativo. Dentro de esta familia de técnicas, los modelos basados en DEA ([3]; [2]) se han mostrados especialmente interesantes para medir su aplicación en el campo educativo.

El Análisis Envolvente de Datos (referenciado habitualmente por sus siglas en inglés, DEA), es una técnica de análisis no paramétrico desarrollado a partir del concepto de eficiencia económica de Farrell [11], que incialmente es concebido para la medición de la eficiencia de un conjunto de unidades. Los trabajos iniciales de Charnes et al. y Banker et al. ([3]; [2]) desarrollan modelos lineales que permiten identificar, basándose únicamente en los valores observados de entradas y salidas de cada unidad, a aquellas unidades con un mejor comportamiento. Dichas alternativas, denotadas como eficientes, constituyen la frontera eficiente y servirán como referencia al resto de alternativas (unidades ineficientes), con un peor desempeño. La distancia de cada unidad ineficiente a la frontera ofrece, además, una medida de la ineficiencia de cada unidad. La evolución de esta familia de modelos ha excedido, sin duda, la concepción original de los autores. Hoy día, los modelos inspirados en DEA se aplican a multitud de sectores y han dado lugar a un ingente número de metodologías para evaluar conjuntos de alternativas.

Los modelos desarrollados en DEA han sido utilizados ampliamente en el estudio del sector público. El hecho de que no sea necesario definir la forma de la función de producción, la posibilidad de considerar múltiples salidas u outputs del proceso productivo (tanto en número como en la forma en que se incluyen en el modelo) y que no se requiera información sobre los precios de éstos, hace que sean modelos especialmente adecuados para ser utilizados en el sector publico, como destacan Santin y Sicilia en [23].

En el caso particular de su aplicación al estudio de centros o sistemas educativos, Mancebón y Bandrés destacan en [19] tres razones fundamentales para su atractivo como técnica de medición en este tipo de estudios. En primer lugar, como modelo no paramétrico, no se busca ajustar los datos observados a una forma funcional preestablecida. En segundo lugar, son modelos que respetan las particularidades individuales de cada unidad. Si bien se asume que existe una tecnología común para el conjunto de unidades, los modelos DEA presentan un alto grado de flexibilidad local en tanto se resuelve un modelo para cada una de las unidades evaluadas. Y, en tercer lugar, los modelos DEA se ajustan muy bien a la naturaleza múltiple de las salidas del proceso educativo y a la carencia total de información sobre los precios de dichas salidas.

Las contribuciones a la literatura científica en este tópico arrancan casi con el origen de los modelos DEA. Los autores que dan origen a la metodología presentan un estudio aplicado al estudio de la eficiencia de los programas educacionales en Estados Unidos [4]. Desde entonces, son numerosos los trabajos que ha aportado al análisis de la eficiencia en el campo de la educación. Tal es así, que pueden encontrarse en la literatura varias revisiones detalladas de estas contribuciones. El lector interesado puede consultar, entre otros, los trabajos de Worthington [25], Johnes et al. [17] y De Witte y Lopez-Torres [10]. En este último, se incluye una muy detallada y actualizada revisión de las contribuciones que incluyen modelos DEA y educación.

En el caso particular de análisis aplicados a España, cabe citar algunos trabajo previos. En ellos, se trata de investigar la eficiencia diferenciada por tipo de institución (escuelas públicas versos privadas), la incidencia de la localización geográfica o el contexto socio-económico de los estudiantes. Pueden citarse, entre otros, los trabajos de Giménez et al. [13], Mancebón y

Muñiz [21], Mancebón et al. [20], Crespo-Cebada et al. [9] y Aparicio et al. [1].

Este gran número de de trabajos que aplican modelos DEA al campo de la educación supone un doble condicionante. Sin duda, tiene una componente positiva en tanto es reflejo del interés del tema, del hecho de que la aplicación de esta familia de modelos al campo particular de la educación está ciertamente justificado y aceptado por la comunidad científica. Además, esta amplitud de referencias permitirá corroborar cualquier nueva propuesta por comparación con anteriores trabajos. Pero sin duda supone también un hándicap para cualquier investigador que quiera iniciarse en el tema, en tanto es complicado aportar nuevas metodologías o propuestas en un campo tan trabajado.

En la presente memoria se presentan tres trabajos en los que se ha profundizado en el estudio de la eficiencia en el contexto de la Economía de la Educación. En particular, hemos analizado la aplicación de modelos basados en el Análisis Envolvente de Datos (DEA) para el análisis de diferentes aspectos relacionados con la educación, tanto en la evaluación de las instituciones educativas como de los sistemas nacionales de educación.

Con los trabajos que se presentan aquí, hemos pretendido analizar diferentes aspectos relacionados con la economía de la educación a través de la propuesta de modelos basados en DEA. Como se podrá ver en los tres trabajos que se presentan, en ningún caso se propone la aplicación sin más de los modelos clásicos de esta metodología. Esto es, las propuestas no pasan por medir sin mas la eficiencia en la actuación de instituciones, países o cualquier agente que intervenga en el proceso con la aplicación de modelos clásicos. En todos ellos, se ha buscado desarrollar nuevos propuestas metodológicas. Todas cuentan con el componente común de proponer modelos de valoración en la que los pesos de las diferentes variables se determinan de manera libre, como caracteriza a los modelos DEA: la determinación de los vectores de pesos se produce de manera endógena, como variables del propio modelo.

En el desarrollo de la misma se ha utilizado como fuente básica de información la publicada en los informes PISA (Programme for International Student Assessment). El informe PISA, publicado originalmente en el año 2000, ha sido elaborado desde ese momento de manera trienal y supone una importante fuente de información para los estudios que pretenden realizar comparaciones entre países. El estudio ha ido incrementando el número tanto de países como de estudiantes encuestados en cada una de sus sucesivas ediciones. Desde los 32 países y 265.000 estudiantes del primer estudio, se ha pasado a evaluar 80 países y más de 500.000 estudiantes en estudio realizado en 2018. Tal y como destaca el propio estudio, su principal objetivo es evaluar las competencias y habilidades de estudiantes de 15 años en los ámbitos de conocimientos de Matemáticas, Ciencias y Lectura. Desde el estudio de 2012, además, se incluyen como opción los conocimientos financieros. La relevancia de los informes reside en que, aparte de la gran cantidad de información sobre los resultados de los test con los que se evalúan a los estudiantes, contiene una vasta base de datos de información sobre cada estudiantes: contexto socioeconómico, familiar, condiciones de la escuela,... que permite evaluar cada caso desde una perspectiva más amplia que la de los resultados obtenidos en un test. Una descripción detallada del proyecto PISA puede verse, entre otros, en [22].

Aunque los datos publicados en los informes son ampliamente utilizados tanto por investigadores como por decisores políticos, PISA no está exento de críticas en ambos ámbitos. Se cuestiona tanto la metodología estadística utilizada en el tratamiento de datos como la base ideológica sobre la que se construyen los informes. Claramente, el que se trate de un proyecto ambicioso, que pretende cubrir un amplísimo campo (tanto en el número de países evaluados como de aspectos concretos de cada estudiante), hace que se trate de un estudio más susceptible de críticas que, por ejemplo, estudios concretos que se limitan a medir o evaluar una parte concreta del conocimiento de los alumnos o que se limitan a un contexto geográfico particular.

Con respecto a la ideología inherente el estudio, diferentes autores han criticado el pragmatismo de las pruebas, la perspectiva utilitarista del conocimiento en la que se basa el estudio. Los informes tratan de evaluar la habilidad de los estudiantes para aplicar los conocimientos adquiridos en la escuela para resolver problemas del día a día [22]. Esta concepción de la educación como herramienta para ser aplicada ha sido discutida por varios autores que entienden que el objetivo de la educación debe ir más allá de dotar herramientas aplicadas a la vida real a los alumnos. Además, el que muchos países hayan iniciado una *carrera* en la mejora de las posiciones que ocupan en el ranking que inducen las puntuaciones del informe, como uno de los objetivos clave de sus políticas educativas (entendido como un fin, no como un resultado de las buenas prácticas), supone un conjunto de restricciones que empeoran las decisiones en la política educativa. Algunos estudios críticos con el proyecto PISA que pueden consultarse son, entre otros, [26] y [24].

Con respecto a los aspectos técnicos, la base estadística que se ha utilizado para la encuesta y algunos de los procedimientos estadísticos para la generación de índices también han sido objeto de crítica. en este punto, es importante destacar que muchas de las críticas realizadas a lo largo de la vida del proyecto han sido consideradas por los directores del mismo, sirviendo como base para mejoras de los resultados publicados en informes sucesivos. Una discusión detallada de los aspectos técnicos más discutidos puede verse, entre otros, en [12]

No obstante, conocidas todas las limitaciones destacadas en los anteriores párrafos, PISA aparece como una de las bases de datos más notables para investigadores y responsables políticos. Hopman justifica en [16] como PISA es una referencia fundamental para cualquier estudio comparativo entre países. En particular, la orientación aplicada del estudio lo destaca como un punto positivo del mismo, en tanto los resultados de las pruebas serán independientes de la escuela a la que pertenece el alumno. E en cuanto a las limitaciones metodológicas, muchas de ellas han sido sucesivamente corregidas y, además, el que se publique los microdatos permite a los investigadores salvar las limitaciones de los índices construidos por los responsables del estudio o construir, a partir del dato individual de cada estudiante, los índices en los que está interesado. Prueba de su amplia aceptación como fuente de información es el número de trabajos basados en el mismo. Según [15], hasta 2016 más de 650 trabajos científicos publicados estaban basado en la información de los informes PISA.

La presente memoria se trata de una tesis doctoral por compendio, en el que se incluyen tres trabajos de investigación ya publicados con un hilo conductor común: la aplicación de modelos basados en DEA al campo de la educación. Buscamos estudiar la utilidad de esta familia de procedimientos en diferentes aspectos, con diferentes aproximaciones a una realidad compleja como es el estudio de los sistemas educativos tanto desde una perspectiva nacional como en comparaciones entre diferentes países.

Como se verá a continuación, cada trabajo, que presentamos por orden temporal de realización, ha tratado de aportar un enfoque diferente al estudio de los sistemas educativos. Hemos considerado diferentes enfoques tanto en la metodología utilizada, como en los objetivos que se persiguen con dicho trabajo como en la perspectiva geográfica y temporal de las aplicaciones empíricas que se incluyen. Como podrá comprobarse, en todos ellos se incluye una aportación metodológica, no se busca aplicar de manera directa modelos ya conocidos de la metodología, y una aplicación empírica que permita ver la utilidad real del modelo propuesto.

El trabajo que se presenta en el primer capítulo, *An assessment of the efficiency of Spanish schools: evaluating the influence of the geographical, managerial, and socioeconomic features*, es el que propone una metodología más cercana a los modelos clásicos de medición de eficiencia No obstante, se propone una aplicación diferente de los mismos, lo que permitirá, como podrá verse, obtener resultados más completos[1].

De manera sintética, el trabajo propone el estudio de las colegios de secundaria en España para el año 2012. A partir de los datos del informe PISA para ese año, proponemos la agregación de los colegios según tres criterios: tipología (diferenciando entre privados, públicos y concertados), localización geográfica (diferenciando 15 regiones recogidas en PISA) y nivel de formación de los padres (bajo, medio y alto). Estos tres criterios, nos permiten construir 135 perfiles o agrupaciones de colegios que comparten las tres características descritas.

Dichos perfiles actúan como unidades de decisión o alternativas a las que estudiamos su eficiencia operativa. Tomando los recursos tradicionales propuestos en la literatura (recursos físicos, financieros y humanos) y los resultados medios de los alumnos en las pruebas realizadas por PISA, medimos la eficiencia de cada tipo de institución. En este punto, se introducen algunas mejoras respecto a los modelos clásicos, como es la reducción de incidencia de outliers mediante la computación de muestras generadas de manera aleatoria.

La aplicación de los modelos de medición de eficiencia permiten evaluar cada alternativa (grupo de colegios en este caso) desde la perspectiva de su desempeño. No se evalúa únicamente si los resultados de los test han sido mayores o menores. Se evalúan dichos resultados a la vista de los inputs, de la dotación de recursos con los que contaba cada alternativa.

La aplicación de técnicas multivariantes sobre los resultados obtenidos de esta primera fase permite obtener diferentes conclusiones. En particular, estamos interesados en detectar si alguno de los criterios utilizado para la construcción de los perfiles es determinante en el valor de su eficiencia. Esto es, cómo de determinante es la tipología de colegio, su ubicación geográfica o el nivel socioeconómico de las familias en la bondad de su desempeño. Como podrá verse, el estudio concluye con que tanto la tipología de colegio como el nivel de formación de los padres son variables determinantes de la eficiencia, mientras que la localización geográfica no aparece como determinante.

El segundo trabajo *A DEA-inspired model to evaluatethe efficiency of education in OECD countries*[2], surge en el desarrollo del trabajo anterior.

---

[1]Segovia-Gonzalez, M.M., Dominguez, C. Contreras, I. (2020) *An assessment of the efficiency of Spanish schools: evaluating the influence of the geographical, managerial, and socioeconomic features.* International Transactions in Operational Research 27: 1845–1868 DOI: 10.1111/itor.12711 (**JCR Q2**, Operations Research and Management Science)

[2]Domínguez, C., Contreras, I. (2020) *A DEA-inspired model to evaluate the efficiency of education in OECD countries*. Revista de Métodos Cuantitativos para la Economía y la Empresa. En prensa. (**SJR Q3**, Business, Management and Accounting)

Sin duda, la información más destacada que se publica en cada informe PISA es el valor de las calificaciones obtenidas por los alumnos en los tests. Generalmente, esta información aparece de manera resumida para cada disciplina (matemáticas, ciencia y lectura) como medias por países, que en muchos casos sirve como base para la construcción de rankings, o diferenciando por regiones dentro de cada país.

Este valor único calculado por PISA no es de lejos la única información contenida en el informe. Y su construcción implica un procedimiento estadístico complejo basado en unos niveles de desempeño o competencia (*proficiency levels*). Cada uno de los 7 niveles incluidos en el informe, ordenados de menor a mayor, tiene un significado en sí mismo. Se entiende, por ejemplo, que el nivel mínimo de competencias se alcanza superando el segundo nivel, o que aquellos alumnos que alcanzan el nivel cinco o superior pueden considerarse excelentes en esa materia. Esto induce a pensar que trabajar directamente con estos niveles, y no con el valor único obtenido como media en el informe, puede resultar de interés. sin embargo, incluir la información recogida en estos niveles en modelos DEA no es directo.

Es por ello que en este segundo trabajo desarrollamos un modelo de medición de eficiencia, basado en el modelo Aditivo DEA, para incluir la información recogida en los niveles de competencia. En la información incluida en DEA aparecen los porcentajes de estudiantes que alcanzan cada uno de los siete niveles. Esto obliga a una doble modificación en los modelos clásicos:

- Cada una de los niveles deben incorporarse al modelo como una variable ordinal, en la que la valoración de estar situado en un nivel (categoría) superior debe valorarse más de estar situado en una inferior.

- Tanto el conjunto de posibilidades de producción como, especialmente, los valores de referencia deben adaptarse a las características particulares de los datos que consideramos. Los valores representan porcentajes y tanto las observaciones como los valores proyectados en la frontera deben sumar 100.

Se articula un modelo en el que la mejora de una unidad implica necesariamente el trasvase de alumnos desde los niveles menos valorados hasta niveles más altos, mejor valorados. A través de la información adicional incluida, se modula el *esfuerzo* que implica el trasvase de una a otra categoría.

Al igual que en el resto de trabajos, en el artículo se incluye una aplicación empírica para evaluar los sistemas educativos de los países OCDE.

En el tercer trabajo *A multiplicative composite indicator to evaluate educational systems in OECD countries*[3], a diferencia de los dos anteriores, la utilización de la metodología DEA no es directa, si no que se utiliza como herramienta complementaria, y no como base del estudio.

En el trabajo proponemos la construcción de un indicador sintético para la evaluación de los sistemas educativos nacionales. Un indicador compuesto, supone de manera sintética como una agregación de una conjunto de indicadores simples o sub-indicadores a través de una función matemática, en la que el vector de pesos que pondera la importancia de cada indicador es un elemento determinante.

Los modelos DEA se han mostrado como una herramienta de gran utilidad para determinar el vector de ponderaciones en los procedimientos de construcción de indicadores compuestos. Esta idea, propuesta inicialmente en [5] y [6] ha dado origen a una filosofía para la generación de pesos conocida como *Beneficio de la Duda*. Las principales ventajas de esta forma de determinar los pesos reside en dos aspectos:

- El vector de ponderaciones se determina de manera autónoma, como parte del procedimiento, y no viene impuesto a partir de información o decisiones subjetivas.

- Cada unidad tiene la oportunidad de seleccionar los pesos en la mejor situación posible. De esta forma, en caso de no recibir una buena evaluación no podrá atribuirla a una elección arbitraria del vector de pesos.

En el trabajo se diseña una metodología completa para la construcción de un indicador compuesto y para la explotación de los resultados. De manera sintética, los puntos más destacables del trabajo son los siguientes.

- Se diseña un panel de indicadores más amplio del utilizado tradicionalmente para evaluar un sistema educativo. Considerando que el objetivo de un sistema educativo va más allá de maximizar los resultados académicos de los alumnos, proponemos un sistema más complejo de indicadores con tres dimensiones.

[3]Domínguez, C., Segovia-González, M.M., Contreras, I. (2020) *A multiplicative composite indicator to evaluate educational systems in OECD countries*. Compare: A Journal of Comparative and International Education. En prensa. (**JCR Q2** Education and Educational Research)

  - Una dimensión académica en la que se evalúa tanto el resultado medio de los estudiantes como que el máximo de éstos alcancen un mínimo de competencias y que se optimice la excelencia.
  - Una dimensión social, en la que se intenta evaluar la educación como instrumento para reducir las desigualdades sociales.
  - Una dimensión personal del alumno, que intenta evaluar cómo influyen los años escolares en el desarrollo y bienestar personal de los alumnos.

- Se propone un esquema de agregación multiplicativo, en el que se pretende minimizar la posible compensación entre indicadores. Este esquema, además, una mejora de las evaluaciones requerirá un esfuerzo de mejora individual de aquellos indicadores en los que peores resultados se ha obtenido.

- La determinación de los pesos se realiza a través de un modelo inspirado en DEA. Siguiendo la idea original propuesta en [7], desarrollamos un nuevo modelo con elección libre de pesos que permite determinar el vector de ponderación de cada unidad.

- El que se utilice un esquema de agregación multiplicativo permite que, cuando se realizan comparaciones de diferentes períodos, las tasas de variación de uno a otro período puedan descomponerse. De esta manera, puede identificarse y aislarse el efecto que genera la variación de las propias observaciones, la selección de un vector de pesos particular y la evaluación del valor base o de referencia.

Al igual que en los otros dos artículos, se presenta una aplicación empírica para el modelo teórico propuesto. En este caso, se evalúan los sistemas educativos de los países OCDE para los años 2012 y 2015. De la comparación de los resultados se concluye que, aunque existe cierta estabilidad en los resultados globales y los rankings de países inducidos a partir del indicador compuesto, se pueden identificar algunas variaciones importantes, derivadas fundamentalmente del cambio en los propios valores observados de los indicadores simples y de una reducción generalizada de los valores medios (que son utilizados como valores de referencia).

Bibliografía

[1] Aparicio, J., Crespo-Cebada, E., Pedraja-Chaparro, F., Santín, D. (2017) *Comparing school ownership performance using a pseudo-panel database: A Malmquist-type index approach*. European Journal of Operational Research 256(2): 533-542.

[2] Banker, R.D., Charnes, A., Cooper,W.W. (1984) *Some models for estimating technical and scale inefficiency in data envelopment analysis*. Management Science 30: 1078-1092.

[3] Charnes, A., Cooper, W.W., Rhodes, E. (1978) *Measuring the efficiency of decision making units*. European Journal of Operations Research 2: 429-444.

[4] Charnes, A., Cooper, W.W., Rhodes, E. (1981) *Evaluating program and managerial efficiency: an application of Data Envelopment Analysis to Program Follow Through*. Management Science 27: 668-697.

[5] Cherchye, L., Moesen, W., and Puyenbroeck, T. (2003) *Legitimately Diverse, yet comparable: On synthesising social inclusion performance in the EU.* CES Discussion paper series 03.01, KU Leuven.

[6] Cherchye, L., Moesen, W., Rogge, N., and Puyenbroeck, T. (2007) *An Introduction to 'Benefit of the Doubt' composite indicators.* Social Indicators Research 82 (1): 111-145.

[7] Contreras, I., and Hinojosa, M.A. (2019) *A note on 'The cross-efficiency in the Optimistic-pessimistic framework'.* Operational Research International Journal. Advance online publication. doi: 10.1007/s12351-019-00484-2.

[8] Cordero-Ferrera J.M., Crespo-Cebada, E., Pedraja-Chaparro, F., Santin-Gonzalez, D. (2011) *Exploring educational efficiency divergences across Spanish regions in PISA 2006*. Revista de Economia Aplicada 19(57): 117-145.

[9] Crespo-Cebada, E., Pedraja-Chaparro, F., Santín, D. (2014) *Does school ownership matter? An unbiased efficiency comparison for regions of Spain*. Journal of Productivity Analysis 41 (1): 153-172.

[10] De Witte, K., Lopez-Torres, L. (2017) *Efficiency in education: a review of literature and a way forward*. Journal of the Operational Research Society 68(4): 339-363.

[11] Farrell, M.J. (1957) *The measurement of the efficiency*. Journal of the Royal Statistical Society Series A 120(3): 253-290.

[12] Fernandez-Cano, A. (2016) *A Methodological Critique of the PISA Evaluations.* e-Journal of Educational Research, Assessment and Evaluation 22 (1), art. M15. doi: http://dx.doi.org/10.7203/relieve.22.1.8806.

[13] Giménez, V., Prior, D., Theme, C. (2007) *Technical efficiency, managerial efficiency and objective setting in the educational system: An international comparison*. Journal of the Operational Research Society 58: 996-1007.

[14] Hanushek, E. (1979) *Conceptual and empirical issues in the estimation of educational production functions*. Journal of Human Resources 14(3): 351-288.

[15] Hopfenbeck, T.N., Lenkeit, J. El Masri, Y., Cantrell,K., Ryan, J., and Baird, J.A. (2018) *Lessons Learned from PISA: A Systematic Review of Peer-Reviewed Articles on the Programme for International Student Assessment, Scandinavian.* Journal of Educational Research 62 (3): 333-353.

[16] Hopmann, S., Brinek, G., and Retzl, M. (2009) *PISA according to PISA-Does PISA keep what is promised.* University of Vienna Press. Vienna.

[17] Johnes, J., Portela, M., Thanassoulis, E. (2017) *Efficiency in education*. Journal of the Operational Research Society 68(4): 331-338.

[18] Levin, H. (1974) *Measuring the efficiency in educational production*. Public Finance Quarterly 2: 3-24.

[19] Mancebon M.J., Bandres, E. (1999) *Efficiency evaluation in secondary schools: The key role of model specification and of ex post analysis of results*. Education Economics 7(2): 131-152.

[20] Mancebon, M.J., Calero, J., Choi, A., Ximenez-De-Embun, D.P. (2012) *The efficiency of public and publicly subsidized high schools in Spain: Evidence from PISA-2006*. Journal of the Operational Research Society 63(11): 1516-1533.

[21] Mancebon, M.J., Muñiz, M.A. (2008) *Private versus public high schools in Spain: Disentangling managerial and programme efficiencies*. Journal of the Operational Research Society 59 (7): 892-901.

[22] OECD (2012) *PISA 2009 Technical Report*. PISA, OECD Publishing. http://dx.doi.org/10.1787/9789264167872-en.

[23] Santin, D., Sicilia, G. (2015) *Measuring the efficiency of public schools in Uruguay: main drivers and policy implications*. Latin America Economic Review 24(5): 1-28.

[24] Sjøberg, S. (2015) *PISA and Global Educational Governance - A Critique of the Project, its Uses and Implications.* Eurasia Journal of Mathematics, Science and Technology Education 11 (1): 111-127.

[25] Worthington, A.C. (2001) *An empirical survey of frontier efficiency measurement techniques education*. Education Economics 9(3): 245-268.

[26] Zhao, Y. (2020) *Two decades of havoc: A synthesis of criticism against PISA.* Journal of Educational Change 21: 245-266.

CAPÍTULO 1

An assessment of the efficiency of Spanish schools: evaluating the influence of the geographical, managerial and socio-economic features.

The aim of the present paper[1] is to explore the efficiency of Spanish schools while simultaneously considering Data Envelopment Analysis (DEA) and multivariate analysis. Test scores from the PISA reports are used as outputs while the resources of each institution are considered as inputs of the analysis. The methodology utilized determines the DEA efficiencies under various input/output combinations and the results are interpreted through the application of factor analysis and property-fitting techniques. The objective of the study is to identify the strengths and weaknesses of each type of school and the connections with the way in which the efficiency is obtained. In the light of the results, the study concludes that there exist differences related with two of the criteria considered: the type of management of the schools; and the parental socio-economic level of the students. However, no differences appear when the geographical location of the institutions is considered to characterize the entities.

[1]Segovia-Gonzalez, M.M., Dominguez, C. Contreras, I. (2020) *An assessment of the efficiency of Spanish schools: evaluating the influence of the geographical, managerial, and socioeconomic features.* International Transactions in Operational Research 27: 1845–1868 DOI: 10.1111/itor.12711

1.1. Introduction

There is a recent and escalating debate in developed countries regarding the importance of controlling public expenses in education. On one hand, based on the correlation between economic growth and social development with the level of human capital, there is a clear incentive for an increase in investment in education [27]. On the other hand, the economic crisis and public deficit in almost all countries worldwide impose the necessity to make the best use of every penny invested in the educational system.

In this context, the concept of efficiency of educational systems becomes equally crucial, that is, the government is required to provide educational services while minimizing the amount of public resources devoted to said services. Equivalently, good results are demanded in terms of educational output with the limited resources available.

From the point of view of the economics of education, education is seen as a production process in which diverse inputs are employed to obtain multiple outputs for a given production technology. The theoretical approach of linking resources to educational outcomes at school level is based on the production function proposed in [30] and [24]. For a particular school i and period t, the following knowledge production function is considered:

$$A(t)_i = g\left[F_i(t), S_i(t), P_i(t), O_i(t), I_i(t)\right]; \tag{1.1}$$

where

$A_i(t)$: vector of educational outcomes.

$F_i(t)$: vector of individual and family background characteristics.

$S_i(t)$: vector of school inputs.

$P_i(t)$: vector of peer or fellow student characteristics.

$O_i(t)$: vector of other external influences.

$I_i(t)$: vector of initial or innate endowments of the student.

It is clear that it is no easy task to quantify the education received by an individual, due to its inherent intangibility and to the necessity for the quality of the education to be taken into account over several years of study. There is, however, a consensus in the literature regarding a standardized test whose results or outcomes are considered as educational outcomes(see, among others, [20]). These results are difficult to forge and can be taken into account by policy makers and families when making decisions in education.

In (1.1), the inputs are divided into five categories, which strive to measure the student's family background, the educational resources assigned to schools (including raw material, and physical and human capital), the characteristics of the students, and possible external influences. Most of these variables interacts with each other. In [25], a complete revision on the specification of educational production functions and the relationships between variables can be consulted.

Nevertheless, unlike other industries, education presents certain characteristics that hinder the estimation of a production function. [32] stress the intangible and multiple nature of the output of education, the time-lag in achieving its results, its cumulative nature, and the fact that the educational process is carried out by the customers themselves.

In this work, we are interesting in the study of the efficiency in education, as a relative measure of the outcomes with respect to total input. Although the objective is not the specification of the education production function, this requires a quantification of the mapping of inputs on outputs. This task can be achieved using either parametric or non-parametric approaches. In [9], a complete discussion about the advantages of non-parametric methods over parametric procedures is given. In brief, in both cases an estimation of the production frontier is carried out. When parametric methods, like Stochastic Frontier Analysis (SFA), are considered, the specification of a particular functional form is required. . This assumption enables the parameters to be estimated, by determining whether the effect is positive or negative, and whether the parameters are statistically significant. The main handicap lies in the difficulty in the function specification and the consequences of a misspecification [31].

On the other hand, considering non-parametric methods such as DEA, misspecification problems are obviated since no assumptions regarding the functional form of the production function is required ([36]) and therefore the consideration of multiple outputs is also possible. A piecewise linear frontier that envelops the data is constructed directly from the observed values. This characteristic enables each entity to be different, by permitting each unit to have local flexibility [16]. These features permit multiple data to be included which represents the particularities of the educational production process. In [50], the authors point out that the freedom in the relation between inputs and outputs is yet another reason for the wide application of non-parametric methodologies. Furthermore, in the educational context in which some usual axioms of productivity may breakdown. This is why non-parametric techniques, especially those from Data Envelopment Analysis (DEA) ([11]), are so convenient for the measurement of the efficiency in this context.

The DEA methodology has been widely used to analyse efficiency in several areas of public expenditure. The main reason for its widespread application is its flexibility and the fact that DEA accounts for multiple outputs, the uncertainty regarding true production technology, and the lack of price information, thereby making it well suited to the peculiarities of the public sector [42].

In [32], three reasons are specified which explain the attractiveness for this methodology to be employed when making the efficiency estimation of education centres. First, DEA (as non-parametric methods) does not oblige the data to adapt itself to an arbitrary functional form. Second, DEA models respect the individual productive practices of each entity, that is, their local flexibility to be considered in the evaluation since an individual model is computed to evaluate each unit. Finally, DEA fits very well with the multiple nature of the education production process and the absence of prices.

In the present paper, an analysis of the efficiency of the Spanish educational system is carried out. The data of the schools is analysed using DEA and multivariate statistics in order to identify relevant variables to explain the inefficiency of the institutions. The main target of this work is to extract conclusions from the discussion of the results. We propose the evaluation of the efficiency of the schools (not the performance, we propose measuring the results with respect to the resources of each institution). The originality of the paper is the study of the incidence of certain characteristics (geographical location, type of management, and parental socio-economic level), which have been proved relevant in previous work, with an alternative approach. The conclusions obtained can be valuable for determining future directions of the sector.

The rest of the paper is organized as follows. Section 1.2 includes a brief review of the literature on DEA applications into the educational context. Section 1.3 introduces the DEA methodology. In Section 1.4, the problem of measuring the efficiency of educational institutions is introduced and the results of a two-step procedure are discussed. Section 1.5 is devoted to the conclusions.

1.2. Literature review

Data Envelopment Analysis is a statistical technique implemented for the evaluation of the relative efficiency of a set of units, developed in [11]. By using linear programming, a frontier of best-practice units is constructed based on observed data. The efficient frontier is used as a benchmark against which the performance of less efficient units can be assessed. The estimated frontier encompasses all the available observations, and each deviation from that frontier is interpreted as a measure of the inefficiency of the units.

In DEA, efficiency is defined in its technical sense, that is to say, as the ability to transform inputs into outputs for a given technology. The literature on efficiency analysis through the use of DEA models in the educational context covers a wide range. The concept of efficiency was first contextualized in the field of education by [30] and has been widely used in the literature to evaluate efficiency in education. Although a complete literature review thereon would require a dedicated research paper, several of the previous studies on the efficiency in education should be cited. A more detailed revision can be seen in [50] and more recently in [27] and [20], in which a complete and recent revision of the contributions in this field are developed.

This family of studies starts with [12], where the authors of the DEA methodology investigate the efficiency of an educational programme in the USA. Since then, several papers have continued the study of efficiency in the field of education. Several papers, such as those by [3], [4] and [49], consider international data for the assessment of a comparison across countries. More recently, [28] study the impact of public and private schooling separating the effect of rural and urban location. Examples of studies for a particular country include [7], [8], [34] and [1]; in particular, [33] and [18] developed studies on the different types of schools across the regions in Spain.

Work, such as that by [14] and [2], applies DEA for the study of efficiency by placing the emphasis on education spending. Other related papers introduced new elements into the analysis. This is the case of [41], who analysed the efficiency of English secondary schools by decomposing them into either the efficiency depending on the centre or the efficiency depending on the individual students themselves.

In a similar way, several studies on the educational efficiency in Spain have been developed in an attempt to investigate the effects on efficiency of the type of school (differentiating between state and private institutions), the geographical location, and the background of the students. Studies that deserve mention include the work of [23], [35], [33] and [19].

In the aforementioned studies, diverse inputs are considered: measures of school resources, such as expenditure per student, articulated in subcategories; ratios student/teacher; facilities; and contextual variables to measure the student's family background.

With respect to the outputs, although different measures can approximate the results of the educational process (including success rates and grades assigned by teachers), there exists a consensus on the use of indicators derived from standardized test scores since these are homogeneous, comparable across countries, and are less difficult to manipulate. In this regard, the Programme for International Assessment (PISA), launched in 2000 and carried out every three years, constitutes an important source of information in the study of competencies acquired by the students and in carrying out comparisons across economies.

The PISA programme has increased in the number of participating schools and countries. In the first edition of the programme, 265,000 students from 32 countries were evaluated. The last edition of this report, in 2015, covered 540,000 students from 72 countries. The main objective of the programme is to evaluate educational systems worldwide by testing the skills and knowledge of 15-year-old students in Mathematics, Science, and Reading skills (and, since 2012, also in financial literacy as an option for each country).

In addition to data on academic achievements, and to statistics covering the results of test on different topics, the PISA database contains a vast amount of information on students, their households, and on the schools they attend. Furthermore it contains synthetic indexes created by OECD experts, and cluster responses to related questions provided by students and school authorities (see [38] for a detailed discussion).

1.3. Data Envelopment Analysis

Data Envelopment Analysis (DEA) is a technique originally proposed in [11] as a methodology for the evaluation of the relative efficiency of a set of units, referred to as Decision-Making Units (DMUs) in DEA terminology, which are involved in a production process or in public services. This methodology formalizes the original ideas proposed in [22] for the measurement of the efficiency of production. In DEA models, the technical efficiency is defined as the relative ability of each DMU to produce outputs from several inputs. The term relative means that each unit is evaluated with respect to the other homogeneous units (as opposed to an absolute evaluation of the performance of each DMU).

The basic efficiency of each unit is evaluated through the ratio of outputs over inputs, that is to say, the measurement of efficiency is defined as a ratio of weighted outputs over weighted inputs. Consider a set of n DMUs to be evaluated. Each DMU consumes m inputs to produce s outputs. The amount consumed of input $i(i = 1, \dots, m)$ and the amount produced of output $r(r = 1, \dots, s)$ by the jth DMU (with $j = 1, \dots, n$) are denoted by $x_i j$ and $y_r j$, respectively. The efficiency of unit j is defined as follows:

$$\text{Efficiency of unit } j = \frac{\sum_{r=1}^{s} v_r \cdot y_{rj}}{\sum_{i=1}^{m} u_i \cdot x_{ij}}; \tag{1.2}$$

where v_r and u_i denotes the weights assigned to output r and input i, respectively.

In DEA models, each unit can freely select the weighting vector, (i.e., each DMU can select its own vectors of weights u and v, so that its own efficiency measurement is optimized), with a common set of constraints that limit this value across the complete set of units, usually by unity. Therefore, each DMU can select its own vector of weights to optimize its individual efficiency measurement. Hence, if a unit fails to achieve the maximum value of efficiency, this failure cannot be attributed to an arbitrary or subjective selection of the weighting factors.

Mathematically, the evaluation of unit o is determined as the solution of the following model:

$$\begin{array}{lll} Max & \theta_o = \frac{\sum_{r=1}^{s} v_r \cdot y_{ro}}{\sum_{i=1}^{m} u_i \cdot x_{io}} & \\ s.t. & \theta_j = \frac{\sum_{r=1}^{s} v_r \cdot y_{rj}}{\sum_{i=1}^{m} u_i \cdot x_{ij}} \leq 1 & j = 1, ..., n \\ & u_i, v_r \geq 0 & i = 1, \dots, m;\ r = 1, \dots, s. \end{array} \tag{1.3}$$

Note that model (1.3) determines the efficiency of unit o, with its own vector of weights (those that maximize the efficiency ratio) subject to a common set of constraints such that the efficiency score is no greater than unity. Model (1.3) must be computed n times, once for each DMU. An efficient unit is characterized by an efficiency score (θ_o) equal to unity. The remaining units, which achieve a value lower than unity, are considered inefficient.

Model (1.3) can be transformed into a linear programming model with certain algebraic transformations, ([11]). The previous model is equivalent to the following expression

$$\begin{array}{ll} Max & \sum_{r=1}^{s} v_r \cdot y_{ro} \\ s.t. & \sum_{i=1}^{m} u_i \cdot x_{io} = 1 \\ & \sum_{r=1}^{s} v_r \cdot y_{rj} - \sum_{i=1}^{m} u_i \cdot x_{ij} \leq 0, \quad j = 1, \ldots, n \\ & u_i, v_r \geq 0 \qquad\qquad\qquad\qquad\quad i = 1, \ldots, m;\ r = 1, \ldots, s. \end{array} \tag{1.4}$$

Model (1.4) is referred to as the CCR model (in reference to the initials of its authors: Charnes, Cooper and Rhodes) in multiplicative form. The dual formulation of (1.4), usually referred to as enveloped form, is simpler to solve and has a useful interpretation:

$$\begin{array}{lll} Min & \theta_o & \\ s.t. & \sum_{j=1}^{n} \lambda_j \cdot y_{rj} \geq y_{jo}, & r = 1, \ldots, s \\ & \sum_{j=1}^{n} \lambda_j \cdot x_{ij} \leq \theta_o \cdot x_{io} & i = 1, \ldots, m \\ & \lambda_j \geq 0 & j = 1, \ldots, n \end{array} \tag{1.5}$$

In (1.5), the variables λ_j represent weights on units. The model determines the maximum inputs that each unit should use to attain its observed output. In this specification, referred to as the input-oriented model, the objective of the model is to determine the maximum radial (proportional) reduction of inputs such that the unit under evaluation is included in the production possibility set, constructed as a linear hull of the observed values of the n DMUs. Efficient units, since they are located at the efficiency frontier, admit no reduction in the vector of inputs, which is reflected by an efficiency score equal to unity. In contrast, output-oriented models determine the maximum expansion of DMU_o such that the unit is in the production possibility set:

$$\begin{array}{lll} Max & \theta_o & \\ s.t. & \sum_{j=1}^{n} \lambda_j \cdot y_{rj} \geq \theta_o y_{jo}, & r = 1, \ldots, s \\ & \sum_{j=1}^{n} \lambda_j \cdot x_{ij} \leq \cdot x_{io} & i = 1, \ldots, m \\ & \lambda_j \geq 0 & j = 1, \ldots, n \end{array} \tag{1.6}$$

DEA models can deal with both constant returns to scale (CRS) and variable returns to scale (VRS). Model (1.5) considers that all the units operate under CRS. In [6], the model with a VRS assumption is proposed. This model is obtained by adding an additional constraint to the dual such that $\sum_{j=1}^{n} \lambda_j = 1$.

Interested readers can find a more extensive explanation regarding the DEA methodology in [13] and in [17], among others. Nevertheless, the application of DEA and the development of models have vastly exceeded their initial objectives, and have generated a wide selection of models and procedures. Certain procedures involve completing the efficiency study with statistical tools (see, among others, [15]. This is the line of research proposed in the present paper, in which the information obtained from DEA models is explored through the application of multivariate techniques.

1.4. An assessment of the efficiency of Spanish schools: DEA and multivariate statistics

The aim of the present work is to study the efficiency of Spanish schools through the data contained in the PISA database. To this end, we propose a two-stage model developed in [45] and in [44] that combines the DEA methodology and statistical tools. In the first stage, DEA efficiency is computed in order to estimate an efficiency score. Instead of considering a unique specification of the model, several combinations of inputs and outputs are considered, which results in the computation of a set of models. In a second stage, the information obtained from the previous step is analysed using multivariate statistical analysis. The objective of the process is to explore sources of inefficiency of Spanish schools and to identify relevant variables, in order either to provide valuable information for the decision-making process, or to identify practices that could improve the current values.

The following subsections include: the description of the dataset (the set of units under evaluation, input and output variables); the efficiency analysis by computing DEA models; and the exploration of those results by means of multivariate statistical tools.

1.4.1. Description of the dataset

In this subsection, the dataset is described through the definition of the units and variables of the efficiency analysis. The data on which the study is based is obtained from the PISA report of 2012 ([39], [40]). We are interested in the performance of the students, hence the individual records of Spanish students are extracted, which amounts to 373,691 values. This data represents the values of 12-year-old students from 902 schools in Spain.

The aim of the work is to study the strengths and weaknesses of each institution and to determine whether any connection exists between the characteristics of the school and the way in which the efficiency is obtained.

Instead of considering each institution individually, the data has been clustered by constructing several profiles on the basis of relevant variables or criteria that can determine the class of school being addressed. We have considered three variables (geographical location in terms of which region is the school is located, ownership or type of school, and educational level of the parents) which enable the set of profiles or group of schools of interest to be constructed. The aim of the study is to determine the relevance of each of these criteria in the efficiency-measurement of the institutions.

It should be highlighted that, in Spain, the competences in educational policy are the concern of the autonomous communities and not of the central government; among other decisions, the quantity of the educational budget and its distribution remain the responsibility of the regions. This analysis enables the evaluation of potential efficiency divergences between regions within the same country, that is, regarding geographical criteria, it permits us to study whether the efficiency is concentrated in certain areas, which is valuable since it could reveal better policies or administrations. With respect to geographical localization, 15 regions have been considered: 14 regions are Autonomous Communities while 1 aggregated all the data from the remaining regions. In this last group, the records from Castile-La Mancha, Ceuta, Melilla and the Canary Island are aggregated. The small number of registers in these regions discourages the separate consideration of each region, and the PISA report presents the aggregated data. Table 1 summarizes the number of records, represented by the number of students and schools that correspond to each region. In parentheses, the percentages of the total number of students and schools are included.

The representation of each of the regions is similar with respect to the number of institutions (approximately 5 % of the national total), except for the case of the Basque Country which accumulates close to 20 % of the total number of schools evaluated. More differences exist between regions with respect to the number of students.

The second criterion considered for the stratification is the type of management. Three types of school are considered: State, private, and government-dependent private schools. The description is included in the PISA Report. State schools are those managed by a public education authority or agency. Private schools are managed by a non-government organisation, such as a church, a trade union, or a private institution. Private schools can be either government-dependent or independent of the government. Government-dependent private schools are managed independently but receive more than 50 % of their core funding from government agencies.

Table 1.1: Geographical distribution of students and institutions

Regions	**Students**		**Schools**	
All other Regions	79,452	(21.26 %)	38	(4.21 %)
Andalusia	75,553	(20.22 %)	52	(5.76 %)
Aragon	9,988	(2.67 %)	51	(5.65 %)
Asturias	7,125	(1.91 %)	56	(6.21 %)
Balearic Islands	8,385	(2.24 %)	54	(5.99 %)
Cantabria	4,334	(1.16 %)	54	(5.99 %)
Castile and Leon	18,422	(4.93 %)	55	(6.10 %)
Catalonia	55,833	(14.94 %)	51	(5.65 %)
Extremadura	10,399	(2.78 %)	53	(5.88 %)
Galicia	18,287	(4.89 %)	56	(6.21 %)
La Rioja	2,566	(0.69 %)	54	(5.99 %)
Madrid	48,845	(13.07 %)	51	(5.65 %)
Murcia	13,115	(3.51 %)	52	(5.76 %)
Navarra	5,245	(1.40 %)	51	(5.65 %)
Basque Country	16,143	(4.32 %)	174	(19.29 %)
Total	373,692	(100 %)	902	(100 %)

Private schools that are independent of the government are similarly managed, but less than 50 % of their core funding comes from government agencies. PISA defines private schools as those that are managed locally, without regard to funding sources. Table 1.2 summarizes the number of schools and students in each category. In this respect, any advantages that have been derived from the management concept can be therefore identified.

Table 1.2: Distributions of the data according to the type of management

Type of School	**Students**		**Schools**	
Private	27,112	(7.39 %)	43	(4.90 %)
Government-dependent	89,685	(24.44 %)	284	(32.38 %)
Public	250,235	(68.18 %)	550	(62.71 %)
Total	367,032		877	

Finally, the schools are classified with respect to the socio-economic status of their students. The average of the highest level of education held by the parents of the students is considered, measured in years, and then the complete list of schools is put into order in terms of this value. The schools are separated into three categories: Low, Medium, and High. We consider a school to be categorized as having Low parental education if its average value is included in the first quartile. The values in the second and third quartiles are specified and denoted as Medium, and those values gin the fourth quartile are categorized as High. Table 1.3 presents the values with respect to this variable.

Table 1.3: Distribution of the data according to the level of parental education

Socio-economic Level	**Students**		**Schools**	
Low	113,626	(30.41 %)	225	(24.94 %)
Medium	175,757	(47.03 %)	452	(50.11 %)
High	84,307	(22.56 %)	225	(24.94 %)
Total	373,690		902	

It is interesting to note how the total number of students varies between Tables 1.1, 1.2 and 1.3 as a consequence of the missing values. This design permits 135 feasible profiles or groups of schools ($15 \times 3 \times 3$) to be constructed which share these three characteristics (for instance, Private schools with low socio-economic level in Andalusia, Public schools with medium socio-economic

Table 1.4: Number of profiles with respect to geographical criterion

Region	Profiles	
All other Regions	6	(6.06 %)
Andalusia	8	(8.08 %)
Aragon	6	(6.06 %)
Asturias	6	(6.06 %)
Balearic Islands	7	(7.07 %)
Cantabria	6	(6.06 %)
Castile and Leon	8	(8.08 %)
Catalonia	8	(8.08 %)
Extremadura	7	(7.07 %)
Galicia	7	(7.07 %)
La Rioja	6	(6.06 %)
Madrid	8	(8.08 %)
Murcia	5	(5.05 %)
Navarra	5	(5.05 %)
Basque Country	6	(6.06 %)

Table 1.5: Number of profiles with respect to management criterion

Type of School	Profiles	
Private	16	(16.16 %)
Government-Dependent	39	(39.39 %)
Public	44	(44.44 %)

level in Aragon,...). We have considered only those cases in which the number of records merit the inclusion. Hence, only 99 profiles have been studied. The following tables (Tables 1.4 to 1.6) summarize the number of profiles considered for each of the criteria referred to above.

The set of profiles constitute the DMUs of the study since we are interested in whether the efficiency or inefficiency of Spanish schools is linked with any of the variables considered for the clustering of the information. In the majority of the previous studies that applied the DEA model in the education context, the input-output combination strives to represent the resources of the educational institutions and the results obtained by the students. In the case of inputs, the variables should measure the resources consumed by the institutions in

Table 1.6: Number of profiles with respect to the criterion of parental socio-economic level

Socio-Economic Level	**Profiles**	
Low	25	(25.25 %)
Medium	36	(36.36 %)
High	38	(38.38 %)

order to perform their activities. In this regard, we consider, as does the PISA report itself, four types of resources needed for learning: financial resources, human resources, material resources, and resources of time. Other factors intervening in the learning process, such as the characteristics of the students, their parental-education levels, and the characteristics of the school, have been considered in the clustering of the data. Note that we strive to evaluate profiles or groups of institutions, and the methodology considered treats each group as an entity. This requires the aggregation of individual observations into a single value for each unit.

The first input considered is the inverse of the size of the classes: this input is denoted as I_A. We consider that this variable can approximate the resources of the institutions and that there exists a inverse relationship between the size of the classes and academic performance.

The third kind of input PISA identifies in the learning process is that of material resources. Schools need certain resources, such as classrooms, heating, and books. Currently, many countries are also making a special effort to provide students with technological material, such as access to the Internet and computers. Technological material is used herein as a proxy for material resources. Specifically, we use the number of computers available for educational purposes in the school divided by the number of students. This variable is labelled as I_B and is used as out second input.

With respect to human resources, teachers represent the most significant part, and hence we use the student-teacher ratio. Although PISA provides the average number of students per teacher in every country, in order to use it as an input in the DEA model, the inverse of this ratio is calculated, that is, the number of teachers divided by the number of students. This third input is denoted as I_C.

The outputs must reflect the results obtained by the students in the learning process. The mean test-scores in Mathematics, Language, and Science for each

Table 1.7: Descriptive statistics of the input/output values

	I_A	I_B	I_C	O_1	O_2	O_3
Max	0.063	2.265	0.224	584.198	603.529	587.357
Min	0.021	0.231	0.047	408.572	406.152	422.762
Average	0.043	0.760	0.090	499.781	501.353	510.705
Std Deviation	0.007	0.324	0.033	31.231	33.619	30.203

type of centre constitute the set of outputs (denoted, respectively, by O_1, O_2, and O_3). Although the efficiency of an education process is not solely dependent on academic results (it is clear that is not easy to quantify the education received by an individual), these values may reflect an efficient use of the resources by the main actors of the process. This is justified by the fact that almost all research that has included studies of efficiency in education using the PISA database has considered the test scores as the outputs of the process (see, among others, Worthington, 2001). Therefore, a consensus exists in the literature regarding consideration of the results from a standardized test as the outputs of the educational process. In this work, the three aforementioned outputs are considered: average results in Mathematics (O_1), Language (O_2), and Science (O_3). Hence, the dataset is composed of a matrix of 99 units and six variables (three inputs and three outputs). Table 1.7 summarizes the statistics for the values of inputs and outputs.

1.4.2. DEA-model specifications and efficiency analysis

In order to evaluate the efficiency of the previously described profiles (DMUs), a model with variable returns of scale (BCC model) is considered. Under the BCC model, an increase in an input would result in a non-proportional increase in the outputs. In our case, the consideration of the BCC model is equivalent to not imposing any assumption regarding the returns of scale, that is, we are interested in the measure of the efficiency of the profiles net to scale effect. The selection of the BCC model is based on the results of [26]. In their paper, a specific discussion is provide on the use of the BCC model when ratios rather than absolute numbers are for inputs and outputs, as in our case. The authors study the appropriateness of using BCC formulation in the presence of ratios, and conclude that, in that case, the reference to VRS is misleading.

An important decision in DEA modelling involves the selection of inputs and outputs to be included in the evaluation of the units. Note that a particular

DMU may or may not be deemed efficient depending on the selection of a particular combination of variables. In this regard, we consider an alternative approach in order to decide which combination of variables the model should contain. Following the procedure proposed in [45], [44], and more recently, in [46], all the possible combinations are considered and, in a second stage, the results obtained are analysed.

The standard way to analyse efficiency is to compute a unique model that includes the complete set of inputs and outputs, which. in our case, are the three inputs (labelled A, B, and C) and three outputs (labelled 1, 2, and 3) described in the previous section. We denote this model as ABC123. It is also interesting, however, to study other combinations of inputs and outputs (for instance, models that only consider the results in the Mathematics test (output 1) and the complete set of inputs). This model is denoted as ABC1. Other combinations can be denoted along the same lines.

There are two main reasons to compute all the combinations of variables: in our case, 49 different models. First, since the efficiency of a unit depends on the mix of inputs and outputs, we can study how the efficiency of each unit varies from one model to another. This may reveal the strengths and weaknesses of each DMU. Second, it can be determined which combinations of inputs and outputs are equivalent in the sense that they produce equivalent results in terms of efficiency ([45]). In order to reduce any redundant information, statistical tools are considered in a second stage of the procedure. The efficiency scores of the 49 models have been adjusted based on bootstrapping methods. To mitigate the sensitivity of efficiency scores to the influence of outliers and measurement errors, following the insights in [10], each model has computed 2.000 rounds in each with a subsample of 90 randomly selected profiles.

The results of computing the DEA model for the 49 specifications are summarized in Table 1.10 (see Appendix). The large number of DMUs (99) and models computed (49) prohibits the inclusion of the complete table, and hence only a summary of the efficiency scores has been represented.

It can be observed that the efficiency of a particular unit depends on the specification considered. For instance, unit 96 High-State-La Rioja (State schools located in La Rioja which have a high level of parental education) is efficient in almost all the models. Its efficiency varies from 100 % in models ABC123 and ABC12, to 78.26 % in model C2 (in which only the third input and the second output are included). The average efficiency, included in the last column, is 98.46 %. At first sight, the efficiency of this unit is based on the first output since the unit is not efficient in all the models which contain only output 2 of about 3 (models AC2, AC3, A2, A3,...).

When the saturated model ABC123 is considered, it can be observed that only ten units are fully efficient. This set of profiles is composed of one profile with a low level (Low-Gov. Dependent-Others), one with a medium level (Medium-Private-Catalonia), and eight profiles with a high parental socio-economic level. Among these high-level units, three are with private or government-dependent institutions (in Andalusia, Catalonia, and Galicia) and five have public management (in Aragon, Balearic Islands, Extremadura, La Rioja, and Madrid). The average efficiency of each model is included in the last row of the Table.

In order to comprehensively reveal the full features of the information contained in the data of Table 1.10 in the Appendix, multivariate techniques are applied in the following section, in an effort to find sources of efficiency in both the variables considered by clustering the institutions and input/output variables.

1.4.3. Multivariate analysis

The data in Table 1.10 (see the Appendix) summarizes the efficiency scores of the 49 models that can be constructed by combining the inputs and outputs considered in the study. This dataset can be treated as a multivariate dataset, in which the model specifications are the variables and the DMUs (the profiles of the schools) are the observations. Hence, we are faced with a table with 49 variables, and 99 observations (corresponding to the profiles).

In order to explore the relationships between the variables, the data-reduction procedures of Factor Analysis (FA) and Principal Component Analysis (PCA) are applied. The application of PCA is carried out in order to determine the number of factors to be estimated ([21]). The idea here is to reveal the similarities and differences between profiles by using PCA, and to explore the reasons for these similarities and differences by employing regression analysis. The study is completed with several hypothesis tests to evaluate the advantages of schools on the basis of the variables considered. The purpose of the procedure of constructing the factors based on FA and PCA is, in brief, to find which models (constructed as combinations of inputs and outputs) offer a similar efficiency measurement. The objective is to summarize the information by representing groups of models with a single value.

First, we proceed with a PCA to remove redundant information from Table 1.10 As stated in [21], PCA enables redundant information to be removed,

Table 1.8: Principal Component Analysis results

Component	Eigenvector	Variance (%)	Cumulative (%)
PC1	34.34	70.08	70.08
PC2	7.86	16.05	86.13
PC3	3.90	7.95	94.08

hidden features of the data set to be highlighted, and it visualizes the main relationships that exist between observations. Three principal components are associated with eigenvalues greater than 0.8, which is the cut-off value proposed in [29]. The first principal component (PC1) explains 70.08 % of the total variance. The second component (PC2) accounts for 70.08 % while the third (PC3) supposes 7.95 %. Eigenvalues, percentage of variance, and the cumulative variance are all summarized in Table 1.8. This is interpreted as the data set and can be described by means of three factors.

Factor Analysis is performed on the data contained in Table 1.10 using the PCA approach, whereby the factors are not rotated. The traditional way of interpreting each factor is to study the correlations between the factors and the original variables. These values are represented in Table A.2 (see the Appendix). For each factor, we present the value of the correlations and their sign. The values have been ranked with respect to the correlations of the models with the first factor.

All the models have a positive sign in the first factor, and the highest values correspond to the complete model ABC123, followed by the models with all three inputs (ABC23, ABC13, ...). The first factor (labelled as F1) can be interpreted as an overall measure of efficiency or overall efficiency as suggested by [44], since the models which include most of the variables, inputs and outputs, have a higher correlation.

In order to interpret the second factor (F2), not only do the numeric values have to be considered, but also the sign of the scores. Models that simultaneously contain inputs A (size of the classrooms) and B (computers per student) have positive values in F2, as do those models in which the only input is B (whereby the maximum is achieved in models B1 and B13). On the other hand, negative scores appear in models that contain input C (number of teachers per student). The highest values are those in which only input C is present (whereby the maximum is achieved in model C2). Therefore, F2 discriminates between models of efficiency that concentrate on input C (with a negative sign) and models of efficiency that concentrate on the other two inputs. We label this component as being efficiency orientated towards human resources vs. material resources.

The third factor is correlated positively with models in which only input B is included, and is correlated negatively with models in which only input A is included, whereby the maximum in each case is given by the model with output 1 (scores in Mathematics). Since only 7.95 % of the variance is represented by this component, the interpretation of the results is focused on the scores of the first and second factor. In a way, the source of the efficiency of each entity can be studied by analysing the scores with respect to the factors. Profiles with similar scores in the factors obtain an efficiency measurement though a similar handling of the resources.

The following Figure (Figure 1.1) shows the plot of the first and second factors. The directional vectors of the efficiency are indicated, and can be interpreted in the same way as can a geographical map. The vector associated with model ABC123, the saturated model, almost coincides with F1 and represents overall efficiency. On the right-hand side of Figure 1, we find units with high overall efficiency, while on the left-hand side there is a concentration of those units with a low level of overall efficiency.

The profiles that are located on the extreme right-hand side of the graphic are labelled HPAN (High level of parental education, private school in Andalusia), HPULR (High level of parental education, state school in La Rioja), HPC (High level of parental education, private school in Catalonia), and MPC (Medium level of parental education, private school in Catalonia).

In a similar way, the direction is given by models which only contain input B and the remaining models, represented by models with input C. The profiles which achieve the highest values in F2 are HPUB (High level of parental education, state school in Baleares), HPUA (High level of parental education, state school in Aragon), and HPUM (High level of parental education, state school in Madrid), whose efficiency is explained by an optimal use of the material resources.

In contrast, the following profiles located at the bottom of the graphic are labelled LGDAN (Low level of parental education, Government-Dependent school in Andalusia), LGDE (Low level of parental education, Government-Dependent school in Extremadura), and MGDE (Medium level of parental education, Government-Dependent school in Extremadura).

It should be pointed out that models that contain inputs of B and combinations of A and B measure efficiency in terms of the use of material

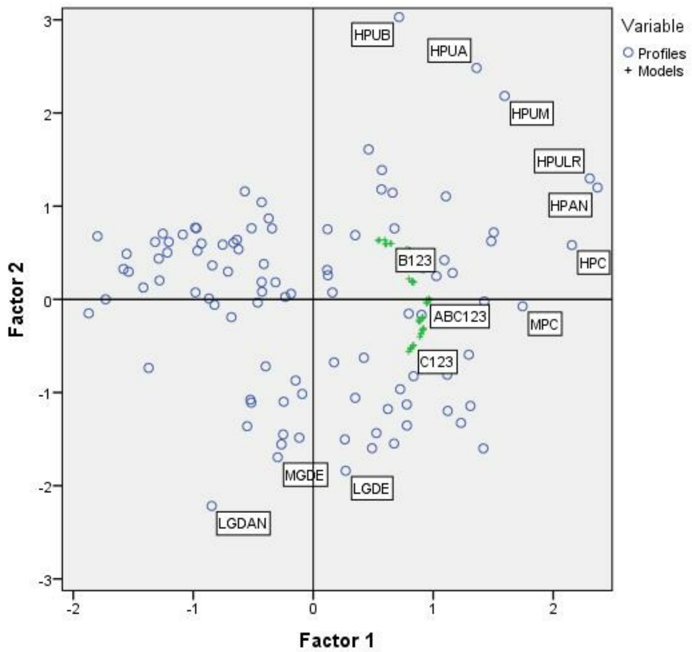

Figure 1.1: Graphical representations of the factor scores

resources. In general terms, state schools present the best results in this family of models. With respect to the models that include only input C (human resources), the best results are achieved mostly by Government-Dependent schools.

The relationship between factors and models can be displayed graphically by using the Property Fitting (Pro-Fit) technique. In this technique, the direction of each DEA-model represents the way in which the efficiency increases. The angle between any two vectors is therefore related with the correlation between the efficiency generated by the considered models. In order to determine the direction of each vector, a regression analysis is carried out in which the dependent variable is the efficiency score of each model and the independent variables are the scores of the first three factors ([43]). The regression coefficients have been normalized for their representation. In Table A.3 (see the Appendix), the values of the analysis including the adjusted regression coefficient are summarized, and the normalized regression coefficient is shown with the contrast statistics in order to test the significance. Note that the minimum value of R^2 is equal to 0.865.

The set of vectors is represented in Figure 1.1: for each of the 49 models, a direction line can be plotted, which indicates the direction in which the efficiency increases for the model (with respect to Factors 1 and 2). Most of the lines overlap, and hence, for a better interpretation, only three of these direction lines are included in the figure. In the upper right-hand-side models, only those that include input B are found (B123, B23, B12, AB123, AB23). Closely related with the direction of Factor 1 are those of models ABC123, ABC23, ABC13 and ABC12. Finally, on the lower right-hand side, models C123, C12, and C23 are represented, among others.

The profiles with a high efficiency in the saturated model ABC123, are located on the right of the horizontal axis. Those profiles which obtain a high efficiency score though an effective handling of input C but not of input A and B will be located in lower-right-hand corner of Figure 1.1 (in the direction specified by the model C123). In a way, the source of the efficiency of each entity can be identified by studying the scores with respect to the factors. Profiles with similar scores in the factor, obtain an efficiency measurement though a similar handling of the resources.

In Figure 1.2, the profiles are represented with regards to the type of management. When the profiles are ranked in terms of the values in F1, then only 5 of the first 30 profiles correspond to state schools. In order to test whether significant differences exist between the efficiency measured by factors 1 and 2 according to the type of ownership of the school (private, state or government-dependent), a non-parametric Kruskal-Wallis test is computed (p-value<0.01). When the Dunn test is considered, significant differences are found. In particular, there are differences between private and state schools (p-value<0.01) and between government-dependent versus state schools (p-value<0.01), but no difference exists between private and government-dependent schools, for factor 1 or factor 2, in favour of private and government-dependent schools in both factors.

In Figure 1.3, the profiles are each labelled with the corresponding parental-education level. Graphically, it can be seen that most of the profiles with a high level are located on the right-hand side. If the profiles are ranked in terms of the values in F1, only 4 of the first 50 profiles contain a low level of parental education. When the Kruskal-Wallis test is computed, it is found that statistically significant differences (p-value$<$ 0,01) exist with respect to the overall efficiency (Factor 1). In contrast, no differences exist when Factor 2 is considered (p-value=0.290). For the overall efficiency, the Dunn test is computed in order to evaluate the differences by pairs. We find that differences exist between profiles with a high level of parental education with respect to those of a low level and also with respect to medium level (p-value$<$ 0,01 in both cases) and between those of a medium level of parental education with

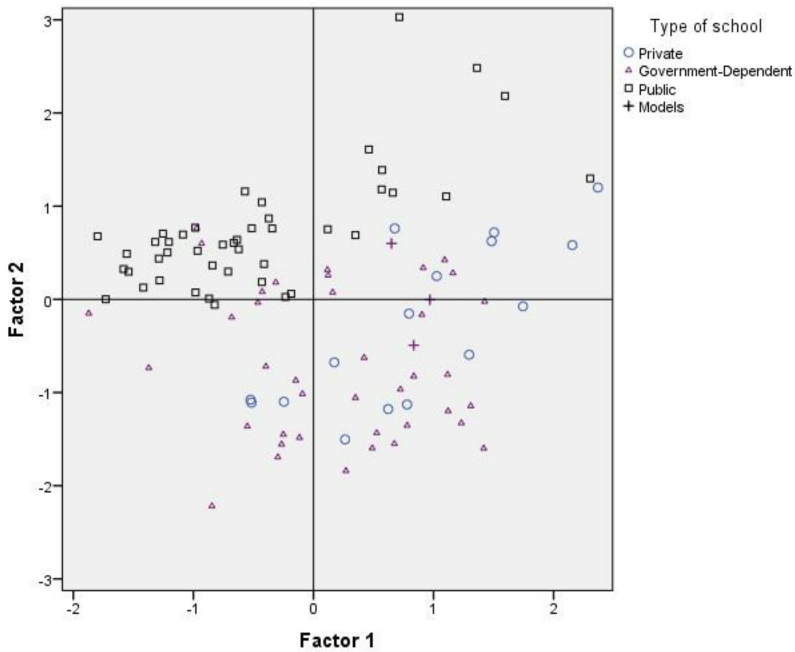

Figure 1.2: Type of management with respect to the first and second factor

respect to those of a low level (p-value$< 0{,}05$).

Regarding geographical location, there is no significant differentiation between regions (Figure 1.4). In this case, an ANOVA analysis is considered in order to support the idea obtained from the graphical analysis: No differences exist across the geographical regions.

We pointed out that efficiency depends on the type of management and the level of parental education but efficiency does not depend on the geographical location of the institution. In order to corroborate this idea, a variance analysis is carried out. Instead of the classic ANOVA of two factors, we have considered an equivalent technique: a regression with dummy variables ([37]). The dependent variable is the efficiency under the saturated model (model ABC123), whereas independent variables are the dummy variables; a set of two dummy variables indicate the level of parental education, and a set of two dummy variables indicate the type of school. When working with dummy

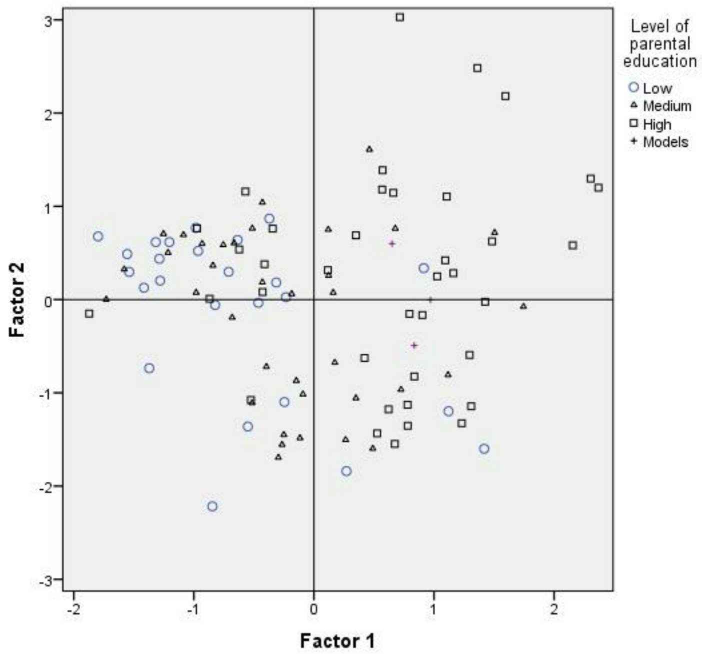

Figure 1.3: Level of parental education with respect to the first and second factors

variables, a group is taken to be the base line, and the effects of changing circumstances with respect to this group is assessed. A low level of parental education and a state school is taken as the basis for comparison. The model includes the interactions between the socio-economic level and the type of school. However, the model including interactions is rejected at a significance level of 5 % when considering the F Fisher test.

In order to correct the bias and according to [47], [48], we construct bootstrap confidence intervals for the regression coefficients. The number of bootstrap samples have been 2.000. The results of the model without interactions is summarized in Table 1.9.

The first and second largest increases in efficiency are obtained when the school is private and the level of education of the parents is high. In a similar way, higher efficiency is found in private and government-dependent schools with respect to state schools.

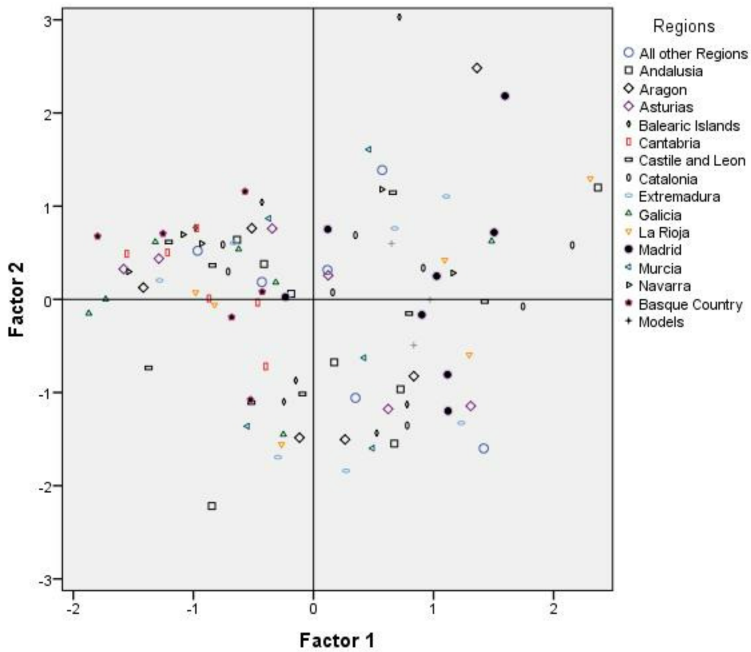

Figure 1.4: Geographical distribution with respect to the first and second factors

1.5. Concluding remarks

In the present paper, we have studied the efficiency of Spanish educational institutions and have striven to measure how school resources influence the academic results of the students. For this task, the information published in the PISA report for Spanish schools has been applied while considering material and human resources and the normalized test scores as variables for the measurement of the success of the educational process.

First, a set of profiles or types of schools has been constructed, which considers three relevant aspects: type of management (state, private, and government-dependent schools), the average level of parental education of the students, and the geographical ubication. A combination of Data Envelopment Analysis and Multivariate Statistical Analysis has been utilized in an effort to determine not only a measure of efficiency of each type of institution but also to determine relevant aspects (from both the inputs and outputs considered in the efficiency analysis and from the separation variables) that explain these results.

With respect to the sources of efficiency, we identify three vectors of

Table 1.9: Bootstrapping regression analysis

	Bootstrap Confidence interval (95 %)		Unstandarized coefficients		
	Lower	Upper	B	Std. Error	t
Constant	0.612	0.697	0.650	0.022	25.339*
Medium level of parental education	-0.029	0.084	0.031	0.028	0.981
High level of parental education	0.072	0.205	0.142	0.034	4.540*
Government-Dependent schools	0.018	0.128	0.075	0.026	2.859*
Private schools	0.061	0.202	0.135	0.035	3.801*

* Significant at the 0.01 level

efficiency: one associated to the saturated model (the vector with all inputs and outputs), and two associated with the use of a specific input (material resources versus human resources). As for the importance of the variables under consideration for the construction of the profiles, we observe that there are significant differences with respect to the type of management (worse results appear in the state schools with respect to the other two management styles), and with respect to the level of parental education (better results of efficiency are related with a higher socio-economic level of the parents). In contrast with the previous two variables under consideration for the construction of the profiles, no differences appear when the differentiation is carried out in terms of the geographical location of the institutions.

With respect to future implications, we should differentiate between managerial and policy implications. With respect to managerial implication, we find differences between public schools and private and government-dependent schools. The managerial style seems to be important in the efficiency-performance in favour of those institutions which management teams do not depend from the governments. Some practices from these institutions should be adopted by the public schools. With repect to policy implications, we considered the geographical location as a criterion since, in Spain; the competences in educational policy are the concern of the autonomous communities and not of the central government. No differences are identified with respect to geographical location; the performance of the institutions does not seem to be related with the regional political activity. Nevertheless, the influence of the parental socio-economic level has been found. The policy initiatives should include development programmes in order to reduce differences and increase the perception of how important the formation is for future generations.

Future lines are such that to carry out intertemporal analysis, the evaluation of the data through the time and the study of convergence or divergence between the classes of institutions we have studied as well as the consideration of alternative methodology as metafrontier analysis.

Bibliografía

[1] Agasisti, T. (2013) *The efficiency of Italian secondary schools and the potential role of competition: a data envelopment analysis using OECD-PISA 2006 data*. Education Economics 21(5): 520-544 .

[2] Agasisti, T. (2014) *The efficiency of public spending on education: An empirical comparison of EU countries*. European Journal of Education 49(4): 543-557.

[3] Afonso, A., St. Aubyn, M. (2005) *Non-parametric approaches to education and health efficiency in OECD countries*. Journal of Applied Economics 8(2): 227-246.

[4] Afonso, A., St. Aubyn, M. (2006) *Cross-country efficiency in secondary education provision: a semi-parametric analysis with non-discretionary inputs*. Economic Modelling 23(3): 476-491.

[5] Aparicio, J., Crespo-Cebada, E., Pedraja-Chaparro, F., Santín, D. (2017) *Comparing school ownership performance using a pseudo-panel database: A Malmquist-type index approach*. European Journal of Operational Research 256(2): 533-542.

[6] Banker, R.D., Charnes, A., Cooper,W.W. (1984) *Some models for estimating technical and scale inefficiency in data envelopment analysis*. Management Science 30: 1078-1092.

[7] Bessent, A.M., Bessent, E.W. (1980) *Determining the comparative efficiency of schools trough data envelopment analysis*. Educational Administration Quarterly 16(2): 57-75.

[8] Bessent, A.M., Bessent, E.W., Kennington, J., Reagan, B. (1982) *An application of mathematical programming to assess productivity in the Houston independent school district*. Management Science 28: 1355-1367.

[9] Burney, N.A., Johnes, J., Al-Enezi, M., Al-Musallam, M. (2013) *The efficiency of public schools: the case of Kuwait*. Education Economics 21(4): 360-379.

[10] Cazals, C., Florens, J.P., Simar, L. (2002) *Nonparametric frontier estimation: A robust approach*. Journal of Econometrics 106(1): 1-25.

[11] Charnes, A., Cooper, W.W., Rhodes, E. (1978) *Measuring the efficiency of decision making units*. European Journal of Operations Research 2: 429-444.

[12] Charnes, A., Cooper, W.W., Rhodes, E. (1981) *Evaluating program and managerial efficiency: an application of Data Envelopment Analysis to Program Follow Through*. Management Science 27: 668-697.

[13] Charnes, A., Cooper, W.W., Lewin, A.Y., Seiford, L.M. (Eds.) (1994) *Data Envelopment Analysis: Theory, methodology, and applications*. Springer Netherlands.

[14] Clemens, B. (2002) *How efficient is education spending in Europe?* European Review of Economics and Finance 1: 3-26.

[15] Coelli, T.J., Rao, D.S.P., O'Donnell, C.J., Battese, G.E. (1998) *An Introduction to Efficiency and Productivity Analysis*. Springer-Verlag.

[16] Cohn, E., Rossmiller, R.A. (1987) *Research on effective schools: Implications for less developed countries*. Comparative Education Review 31(3): 377-399.

[17] Cooper, W.W., Seiford, L.M., Tone, K. (2007) *Data Envelopment Analysis: A comprehensive text with models, applications, references and DEA-solver software*. Springer-Verlag US.

[18] Cordero-Ferrera J.M., Crespo-Cebada, E., Pedraja-Chaparro, F., Santin-Gonzalez, D. (2011) *Exploring educational efficiency divergences across Spanish regions in PISA 2006*. Revista de Economia Aplicada 19(57): 117-145.

[19] Crespo-Cebada, E., Pedraja-Chaparro, F., Santín, D. (2014) *Does school ownership matter? An unbiased efficiency comparison for regions of Spain*. Journal of Productivity Analysis 41 (1): 153-172.

[20] De Witte, K., Lopez-Torres, L. (2017) *Efficiency in education: a review of literature and a way forward*. Journal of the Operational Research Society 68(4): 339-363.

[21] Dunteman, G.H. (1989) *Principal Component Analysis*. Sage, London.

[22] Farrell, M.J. (1957) *The measurement of the efficiency*. Journal of the Royal Statistical Society Series A 120(3): 253-290.

[23] Giménez, V., Prior, D., Theme, C. (2007) *Technical efficiency, managerial efficiency and objective setting in the educational system: An international comparison*. Journal of the Operational Research Society 58: 996-1007.

[24] Hanushek, E. (1979) *Conceptual and empirical issues in the estimation of educational production functions*. Journal of Human Resources 14(3): 351-288.

[25] Hanushek, E. (1986) *The economics of schooling: Production and efficiency in public schools*. Journal of Economic Literature 24(3): 1141-1177.

[26] Hollingsworth, B., Smith, P. (2003) *Use of ratios in data envelopment analysis*. Applied Economic Letters 10(1): 733-735.

[27] Johnes, J., Portela, M., Thanassoulis, E. (2017) *Efficiency in education*. Journal of the Operational Research Society 68(4): 331-338.

[28] Johnes, G., Virmani, S. (2019) *The efficiency of private and public schools in urban and rural areas: moving beyond the development goals*. International Transactions in Operational Research https://doi.org/10.1111/itor.12658

[29] Joliffe, I.T. (1972) *Discarding variables in Principal Components Analysis*. Applied Statistical 21: 160-173.

[30] Levin, H. (1974) *Measuring the efficiency in educational production*. Public Finance Quarterly 2: 3-24.

[31] Lovell, C.A.K. (1993) *Production frontiers and productive efficiency*. In Ed. Fried, H.O., Lovell, C.A.K., Schmidt, S.S. *The measurement of productive efficiency*, 3-67. Oxford, Oxford University Press.

[32] Mancebon M.J., Bandres, E. (1999) *Efficiency evaluation in secondary schools: The key role of model specification and of ex post analysis of results*. Education Economics 7(2): 131-152.

[33] Mancebon, M.J., Calero, J., Choi, A., Ximenez-De-Embun, D.P. (2012) *The efficiency of public and publicly subsidized high schools in Spain: Evidence from PISA-2006*. Journal of the Operational Research Society 63(11): 1516-1533.

[34] Mancebon, M.J., Mar-Molinero, C. (2000) *Performance in Primary Schools*. Journal of the Operational Research Society 51: 843-854.

[35] Mancebon, M.J., Muñiz, M.A. (2008) *Private versus public high schools in Spain: Disentangling managerial and programme efficiencies*. Journal of the Operational Research Society 59 (7): 892-901.

[36] Mante, B. (2001) *Measuring the performance of the state secondary schools in Victoria: An application of data envelopment analysis.* Education Research and Perspectives 28(1):105-132.

[37] Neter, J., Kutner, M. H., Nachtsheim, C. J., Wasserman, W. (1996). *Applied Linear Statistical Models* 4th Edition, Chicago, Irwin.

[38] OECD (2012) *PISA 2009 Technical Report*. PISA, OECD Publishing. http://dx.doi.org/10.1787/9789264167872-en.

[39] OECD (2013) *PISA 2012 Results: What Makes Schools Successful? Resources, Policies and Practices (Volume IV)*. PISA, OECD Publishing. http://dx.doi.org/10.1787/9789264201156-en.

[40] OECD (2014) *PISA 2012 Results: What Students Know and Can Do? Student Performance in Mathematics, Reading and Science (Volume I, Revised edition, February 2014)*. PISA, OECD Publishing. http://dx.doi.org/10.1787/9789264201118-en.

[41] Portela, M.C.A.S., and Thanassoulis, E. (2001) *Decomposing school and school-type efficiency*. European Journal of Operational Research 132(2): 114-130.

[42] Santin, D., Sicilia, G. (2015) *Measuring the efficiency of public schools in Uruguay: main drivers and policy implications*. Latin America Economic Review 24(5): 1-28.

[43] Schiffman, SS., Reynolds, ML., Young, FW. (1981). *Introduction to Multidimensional Scaling: Theory, Methods and Applications*. Academic Press: London.

[44] Serrano-Cinca, C., Fuertes-Callen, Y., Mar-Molinero, C. (2005) *Measuring DEA efficiency in Internet companies*. Decision Support Systems 38: 557-573.

[45] Serrano-Cinca, C., Mar-Molinero, C. (2004) *Selecting DEA specifications and ranking units via PCA*. Journal of the Operational Research Society 55: 521-528.

[46] Serrano-Cinca, C., Mar-Molinero, C., Fuertes, Y. (2016) *Input and output search in DEA: the case of financial* . In *Handbook of operations analytics using Data Envelopment Analysis*. Hwang, S.N., Lee, H.S., Zhu, J. (Editors). Springer. New York.

[47] Simar, L., Wilson, P.W. (2007) *Estimation and inference in two-stage, semi-parametric models of production processes*. Journal of Econometrics 136: 31-64.

[48] Simar, L., Wilson, P.W. (2011) *Two-stage DEA: caveat emptor*. Journal of Productivity Analysis 36: 205-218.

[49] Sutherland, D., Price, R., Gonand, F. (2009) *Improving public spending efficiency in primary and secondary education*. OECD Journal of Economic Studies 1: 1-30.

[50] Worthington, A.C. (2001) *An empirical survey of frontier efficiency measurement techniques education*. Education Economics 9(3): 245-268.

Table 1.10: Summary of efficiencies per profile

Profile	ABC123	ABC12	ABC23	ABC13	ABC1	…	C23	C13	C1	C2	C3	Average
Low-Private-Balearic Islands (LPB)	74.21 %	74.21 %	74.20 %	74.21 %	74.21 %	…	70.94 %	72.74 %	72.74 %	72.74 %	70.93 %	61.79 %
Low-Gov.Dep.-Others (LGDOR)	100.00 %	100.00 %	100.00 %	100.00 %	100.00 %	…	100.00 %	100.00 %	100.00 %	100.00 %	100.00 %	83.40 %
Low-Gov.Dep.-Andalusia (LGDAN)	69.79 %	69.79 %	69.79 %	69.79 %	69.79 %	…	69.79 %	69.79 %	69.79 %	69.79 %	69.79 %	52.58 %
Low-Gov.Dep-Cantabria (LGDC)	66.66 %	66.66 %	66.66 %	66.66 %	66.66 %	…	58.11 %	58.11 %	58.11 %	58.11 %	58.11 %	58.32 %
Low-Gov.Dep-Castile and Lion (LGDCL)	54.18 %	54.18 %	54.18 %	50.78 %	50.77 %	…	50.37 %	46.23 %	46.23 %	50.37 %	46.23 %	45.04 %
Low-Gov.Dep.-Catalonia (LGDC)	88.79 %	88.79 %	88.79 %	88.74 %	88.47 %	…	69.49 %	69.73 %	69.73 %	68.49 %	68.48 %	75.74 %
Low-Gov.Dep.-Extremadura (LGDE)	84.42 %	84.42 %	84.42 %	84.42 %	84.42 %	…	83.47 %	83.47 %	83.47 %	83.47 %	83.47 %	67.56 %
Low-Gov.Dep.-Galicia (LGDG)	72.32 %	72.32 %	72.32 %	72.32 %	72.32 %	…	60.20 %	60.19 %	60.19 %	60.20 %	60.19 %	61.75 %
Low-Gov.Dep-Madrid (LGDM)	95.25 %	95.25 %	95.25 %	95.25 %	95.25 %	…	92.33 %	92.33 %	92.33 %	92.33 %	92.33 %	79.73 %
Low-Gov. Dep.-Murcia (LGDMU)	69.67 %	69.67 %	69.67 %	69.67 %	69.67 %	…	67.49 %	67.49 %	67.49 %	67.49 %	67.49 %	56.53 %
⋮	⋮	⋮	⋮	⋮	⋮	⋱	⋮	⋮	⋮	⋮	⋮	⋮
High-State-Galicia (HPUG)	64.22 %	64.18 %	64.22 %	64.21 %	63.93 %	…	47.39 %	47.27 %	47.26 %	47.34 %	47.08 %	56.13 %
High-State-La Rioja (HPULR)	100.00 %	100.00 %	100.00 %	100.00 %	100.00 %	…	100.00 %	100.00 %	100.00 %	79.84 %	89.53 %	98.64 %
High-State-Madrid (HPUM)	100.00 %	100.00 %	100.00 %	100.00 %	100.00 %	…	62.17 %	63.02 %	63.01 %	61.71 %	62.14 %	88.35 %
High-State-Navarra (HPUN)	97.03 %	97.03 %	91.61 %	97.03 %	97.03 %	…	33.88 %	34.97 %	34.97 %	33.88 %	33.87 %	67.57 %
High-State-Basque Country (HPUBC)	68.08 %	68.08 %	68.08 %	68.08 %	68.08 %	…	38.80 %	39.92 %	39.92 %	38.80 %	38.79 %	57.44 %
Average	76.71 %	76.14 %	76.50 %	76.24 %	75.48 %	…	62.50 %	62.49 %	61.73 %	61.52 %	61.30 %	

Table 1.11: Factor analysis: correlations between factors and variables

Model	F1	F2	F3	Model	F1	F2	F3
ABC123	0.970	- 0.004	0.030	A123	0.834	0.184	- 0.484
ABC23	0.969	- 0.009	0.053	AB23	0.831	0.510	- 0.136
ABC13	0.965	0.006	0.035	AB123	0.830	0.507	- 0.157
ABC12	0.952	- 0.034	0.030	A3	0.830	0.188	- 0.475
ABC3	0.947	- 0.004	0.073	C23	0.828	- 0.502	0.100
ABC2	0.946	- 0.041	0.061	A13	0.827	0.191	- 0.492
ABC1	0.942	- 0.029	0.044	A12	0.825	0.196	- 0.484
AC13	0.918	- 0.311	- 0.171	A2	0.822	0.188	- 0.490
AC23	0.917	- 0.325	- 0.155	C12	0.815	- 0.529	0.098
AC123	0.917	- 0.317	- 0.170	C3	0.813	- 0.524	0.111
BC123	0.914	- 0.195	0.312	C1	0.812	- 0.534	0.101
BC23	0.909	- 0.198	0.324	AB12	0.811	0.515	- 0.157
AC3	0.908	- 0.333	- 0.149	AB13	0.807	0.526	- 0.161
BC13	0.908	- 0.199	0.321	AB2	0.802	0.524	- 0.125
AC12	0.901	- 0.368	- 0.150	A1	0.798	0.220	- 0.504
AC1	0.898	- 0.370	- 0.140	AB3	0.797	0.530	- 0.128
BC12	0.892	- 0.229	0.324	C2	0.795	- 0.560	0.113
AC2	0.890	- 0.401	- 0.130	AB1	0.781	0.541	- 0.151
BC3	0.885	- 0.213	0.343	B123	0.650	0.599	0.351
BC2	0.885	- 0.231	0.341	B23	0.641	0.599	0.361
BC1	0.881	- 0.240	0.334	B12	0.611	0.593	0.394
A23	0.838	0.183	- 0.471	B2	0.600	0.593	0.407
C13	0.836	- 0.492	0.092	B13	0.597	0.636	0.386
C123	0.835	- 0.493	0.090	B1	0.552	0.636	0.431
				B3	0.540	0.631	0.423

Table 1.12: Pro-Fit analysis: linear regression results 1/2

Model	ϕ_1	ϕ_2	ϕ_3	F	**Adj** R^2	Model	ϕ_1	ϕ_2	ϕ_3	F	**Adj** R^2
ABC123	0.970	-0.004	0.030	511.736	0.940	**BC1**	0.881	-0.240	0.334	556.020	0.944
	39.163**	-0.171	1.193				37.009**	-10.078**	14.029**		
ABC12	0.952	-0.034	0.030	317.509	0.906	**BC2**	0.885	-0.231	0.341	632.693	0.951
	30.828**	-1.110	0.975				39.495**	-10.335	15.213		
ABC23	0.969	-0.009	0.053	514.178	0.940	**BC3**	0.885	-0.213	0.343	560.766	0.945
	39.216**	-0.368	2.130*				37.307**	-8.989**	14.479**		
ABC13	0.965	0.006	0.035	437.803	0.930	**A123**	0.834	0.184	-0.484	832.201	0.962
	36.216**	0.237	1.321				42.460**	9.342**	-24.627**		
ABC1	0.942	-0.029	0.044	254.084	0.886	**A12**	0.825	0.196	-0.484	643.537	0.952
	27.567**	-0.840	1.274				37.112**	8.834**	-21.800**		
ABC2	0.946	-0.041	0.061	287.269	0.898	**A23**	0.838	0.183	-0.471	704.520	0.956
	29.268**	-1.254	1.899				39.374**	8.597**	-22.121**		
ABC3	0.947	-0.004	0.073	295.057	0.900	**A13**	0.827	0.191	-0.492	819.588	0.962
	29.663**	-0.117	2.298*				41.812**	9.667**	-24.842**		
AB123	0.830	0.507	-0.157	1060.711	0.970	**A1**	0.798	0.220	-0.504	488.544	0.937
	47.522**	29.033**	-8.996**				31.515**	8.686**	-19.925**		
AB12	0.811	0.515	-0.157	575.143	0.946	**A2**	0.822	0.188	-0.490	618.092	0.950
	34.598**	21.989**	-6.698**				36.301**	8.311**	-21.620**		
AB23	0.831	0.510	-0.136	953.555	0.967	**A3**	0.830	0.188	-0.475	592.384	0.948
	45.152**	27.701**	-7.389**				35.904**	8.120**	-20.546**		
AB13	0.807	0.526	-0.161	662.750	0.953	**B123**	0.650	0.599	0.351	296.820	0.901
	36.938**	24.023**	-7.357**				20.392**	18.791**	11.025**		
AB1	0.781	0.541	-0.151	390.499	0.923	**B12**	0.611	0.593	0.394	231.961	0.876
	27.781**	19.252**	-5.390**				17.177**	16.666**	11.094**		
AB2	0.802	0.524	-0.125	451.488	0.932	**B23**	0.641	0.599	0.361	286.971	0.897
	30.552**	19.962**	-4.748**				19.817**	18.530**	11.173**		

* Significant at the 0.05 level
** Significant at the 0.01 level

Table 1.12: Pro-Fit analysis: linear regression results 2/2

Model	ϕ_1	ϕ_2	ϕ_3	F	**Adj** R^2	Model	ϕ_1	ϕ_2	ϕ_3	F	**Adj** R^2
AB3	0.797	0.530	-0.128	434.529	0.930	**B13**	0.597	0.636	0.386	322.036	0.908
	29.802**	19.813**	-4.784**				19.456**	20.718**	12.584**		
AC123	0.917	-0.317	-0.170	1040.768	0.970	**B1**	0.552	0.636	0.431	270.537	0.892
	52.022**	-17.959**	-9.669**				16.619**	19.165**	12.966**		
AC12	0.901	-0.368	-0.150	1003.348	0.968	**B2**	0.600	0.593	0.407	224.962	0.873
	50.206**	-20.488**	-8.344**				16.644**	16.447**	11.286**		
AC23	0.917	-0.325	-0.155	1080.822	0.971	**B3**	0.540	0.631	0.423	210.027	0.865
	53.002**	-18.798**	-8.941**				14.548**	17.000**	11.378**		
AC13	0.918	-0.311	-0.171	975.831	0.968	**C123**	0.835	-0.493	0.090	581.447	0.947
	50.459**	-17.117**	-9.401**				35.817**	-21.131**	3.863**		
AC1	0.898	-0.370	-0.140	811.315	0.961	**C12**	0.815	-0.529	0.098	655.092	0.952
	45.144**	-18.616**	-7.031**				37.010**	-23.991**	4.466**		
AC2	0.890	-0.401	-0.130	993.578	0.968	**C23**	0.828	-0.502	0.100	583.990	0.947
	49.334**	-22.245**	-7.213**				35.604**	-21.584**	4.295**		
AC3	0.908	-0.333	-0.149	722.812	0.957	**C13**	0.836	-0.492	0.092	580.701	0.947
	43.217**	-15.830**	-7.083**				35.813**	-21.070**	3.949**		
BC123	0.914	-0.195	0.312	1090.194	0.971	**C1**	0.812	-0.534	0.101	657.391	0.953
	53.052**	-11.297**	-18.123**				36.908**	-24.267**	4.590**		
BC12	0.892	-0.229	0.324	657.674	0.953	**C2**	0.795	-0.560	0.113	720.366	0.957
	40.583**	-10.417**	14.750**				37.759**	-26.577**	5.387**		
BC23	0.909	-0.198	0.324	1033.665	0.969	**C3**	0.813	-0.524	0.111	568.787	0.946
	51.384**	-11.208**	18.305**				34.488**	-22.240**	4.720**		
BC13	0.908	-0.199	0.321	942.012	0.966						
	49.093	-10.478**	17.332**								

* Significant at the 0.05 level
** Significant at the 0.01 level

CAPÍTULO 2

A DEA-inspired model to evaluate the efficiency of education in OECD countries

In this paper[1] empirical application to the study about the efficiency of the performance of the educational systems across countries is developed. With the information published in the PISA 2015, Data Envelopment Analysis methodology is considered to evaluate the efficiency in the use of the resources devoted to education by OECD countries. Similar to previous studies, the main resources needed for learning, financial, human resources, material and time have been considered. Alternatively to previous proposals, the mean scores have not been included as the output of the process. Instead of that, to quantify the results of the learning process, the percentages of students in each proficiency level of the PISA test have been computed.

An ad hoc model based on the Additive DEA-model is proposed, adapting the formulation to the particular features of the vector of outputs considered. Considering that the the aggregate value of output is fixed and that the relative weight of the outputs differs, inefficient units improve their performance by reallocating that fixed value among different outputs, moving units from the less valued to the most valued ones.

[1]Domínguez, C., Contreras, I. (2020) *A DEA-inspired model to evaluate the efficiency of education in OECD countries*. Revista de Métodos Cuantitativos para la Economía y la Empresa. En prensa.

2.1. Introduction

There is a recent and increasing debate at the developed countries about the relevance of controlling public expenses in education. On the one hand, based on the correlation between the economic growth and social development with the level of human capital, there is a clear incentive for an increasing investment in education. On the other hand, the economic crisis and public deficit in almost all countries impose the necessity of a best use of every coin invested in the educational system.

In this context, the concept of efficiency of educational systems becomes crucial. That is, governments are required to provide educational services by minimizing the amount of public resources (money) devoted to them. Or equivalently, they are required to obtain good results in terms of educational outputs with the available (fixed) resources.

From the point of view of the economics of education, education is seen as a production process in which diverse inputs are used to obtain multiple outputs for a given production technology. The theoretical approach of linking resources to educational outcomes at school level is based on the production function proposed in [28] and [26]. For a particular school s the function is defined as follows

$$A_s = f(B_s, S_s); \tag{2.1}$$

where A_s represents the educational output, normally measured through scores on standardised tests. It is clear that it is not an easy task to quantify the education received by an individual, due to its inherent intangibility and necessity to consider the quality beyond several years of study. However there is a consensus in the literature about considering the results from a standardised test as educational outputs. They are difficult to forge and they are taken into account by policy makers and families when making decisions in education.

In (2.1) the inputs are divided into B_s and S_s, which denote the average student's family background and the educational resources assigned to school s respectively. Classically, they consider the main inputs required to carry out the learning process: raw material, physical and human capital.

Nevertheless, unlike other industries, education presents certain characteristics that hinder the estimation of a production function. [31] stress the intangible and multiple nature of the output, the time-lag in achieving its results, its cumulative nature and that the educational process is carried out by the customers themselves. This is why non-parametric techniques such as Data Envelopment Analysis (DEA) are so convenient to measure the efficiency in this context. They allow the assessment of the efficiency of the different units without having to estimate a production function.

DEA is a statistical technique used to evaluate the relative efficiency of a set of units developed in [13]. By using linear programming a frontier of best-practice units is constructed based in observed data. The efficient frontier is used as a benchmark against which the performance of less efficiency units can be assessed. The estimated frontier envelops all the available observations, and each deviation from that frontier is interpreted as a measure of the inefficiency of the units. The DEA methodology has been widely used to analyse efficiency in several areas of public expenditure. The main reason for its widespread application is its flexibility, the fact that it accounts for multiple outputs, the uncertainty about true production technology and the lack of price information; making it well suited to the peculiarities of the public sector [38].

In DEA, efficiency is defined in a technical sense. That is to say, as the ability of transforming inputs into outputs for a given technology. The concept of efficiency was first contextualized in the field of education by [28] and has been widely used in the literature to evaluate efficiency in education. Although a complete literature review would require a specific research paper, some of the previous studies about the efficiency in education must be cited. In any case, a more detailed revision can be seen in [44] and [27].

This family of studies starts with [14], where the authors of the DEA methodology investigate the efficiency of an educational program in the USA. Since them, several work have continued the study of efficiency in the field of education. [4], [5], [40] or [1], among others, considered international data to asses a comparison across countries. Examples of studies for a particular country are, for instance, [10], [11] or [2]; in particular, [32] or [21] developed studies of the different types of school across the regions in Spain.

Works like [17], [7] or [3] apply DEA for the study the efficiency placing the emphasis on the educational spendings. Other related papers, introduced new elements into the analysis. That is the case of [36], which analysed the efficiency of English secondary schools by decomposing them into the efficiency depending on the centre and on the individual students themselves. In a similar way [32] studied the results for Spain, in an attempt to differentiate between the effects of the type of school, the school, and the students in the efficiency; and [24] which introduced the concept of managerial efficiency.

In the aforementioned studies, diverse inputs are considered: measures of schools' resources like expenditure per student, eventually articulated in subcategories, student/teacher ratios, facilities, contextual variables to measure the student-family's background,...

With respect to the outputs, although different measures can approximate the results of the educational process (success rates, grades assigned by teachers,...), there exists a consensus about the use of indicators derived by standardised test scores as they homogeneous, comparable across countries and more difficult to manipulate. In this point, the Programme for International Assessment (PISA programme), launched in 2000 and carried out every three years, constitutes an important source of information to study the competencies acquired by the students and to make comparison across economies.

The PISA programme, initiated in 2000 and carried out every three years, has experimented a constant increase in the number of participating schools and countries. In the first edition of the programme, 265.000 students from 32 countries were evaluated. The last edition of this report in 2015 covered 540.000 students from 72 countries. The main target of the programme is to evaluate educational systems worldwide by testing the skills and knowledge of 15-year-old students in mathematics, science and reading (and, since 2012 in financial literacy as an option for each country).

In addition to academic achievement data, summarizing the results on the test about different topics, the PISA database contains a vast amount of information about students, their households and the schools they attend; as well as synthetic indexes, elaborated by OECD experts, by clustering responses to related questions provided by students and school authorities.

In this paper an alternative DEA-inspired model is proposed in order to assess the efficiency of the educational systems in the OECD (The Organisation for Economic Co-operation and Development) countries, using the information included in the PISA database. In particular, we are interested in the consideration of the number of students that achieve each proficiency level as the output of the system. To this end, an innovative model based in DEA methodology is developed.

The rest of the paper is organized as follows. Section 2.2 introduces DEA methodology and a new model for the evaluation of the efficiency, in a situation in which the output represents percentages of different categories is studied. In Section 2.3 the problem of measuring the efficiency of the educational systems across economies through PISA dataset is introduced and the dataset is described. Section 2.4 contains the discussion of results and Section 2.5 is devoted to the conclusions.

2.2. Methodology: Data Envelopment Analysis

Data Envelopment Analysis (DEA) is a technique originally proposed in [13] as a methodology to evaluate the relative efficiency of a set of units, referred to as Decision Making Units (DMUs) in DEA terminology, involved in a production process or in public services. This methodology formalizes the original ideas proposed in [23] of measuring efficiency of the production. In DEA models, the technical efficiency is defined as the relative ability of each DMU to produce outputs from several inputs.

The basic efficiency of each unit is evaluated through the ratio of outputs over inputs. That is to say, the measurement of efficiency is defined as a ratio of weighted outputs over weighted inputs. Consider a set of n DMUs to be evaluated. Each DMU consumes m inputs to produce s outputs. By x_{ij} and y_{rj} are denoted, respectively, the amount consumed of input i ($i = 1, \dots, m$) and the amount produced of output r ($r = 1, \dots, s$) by the jth DMU (with $j = 1, \dots, n$). The efficiency of unit j is defined as follows

$$\text{Efficiency of unit } j = \frac{\sum_{r=1}^{s} v_r \cdot y_{rj}}{\sum_{i=1}^{m} u_i \cdot x_{ij}}; \tag{2.2}$$

where v_r and u_i denotes the weights assigned to output r and input i respectively.

DEA models determine those DMUs that constitute the efficiency frontier (efficient units) and the distance of the remaining DMUs (inefficient units) from the frontier. This distance, which represents a measure of the inefficiency of the units, will depend on the DEA model considered. The main characteristic of DEA methodology is that each unit can freely select the weighting vector, (i.e., each DMU can select their own vectors of weights u and v so that its own efficiency measurement is optimized), with a common set of constraints that limit this value for the complete set of units, usually equal to or lower than unity. Therefore, each DMU can select its own vector of weights to optimize its individual efficiency measurement. Hence, if a unit fails to achieve the maximum value of efficiency, this failure cannot be attributed to an arbitrary selection of the weighting factors.

Mathematically, the evaluation of unit o is determined as the solution of the following model,

$$\begin{array}{lll} Max & \theta_o = \frac{\sum_{r=1}^{s} v_r \cdot y_{ro}}{\sum_{i=1}^{m} u_i \cdot x_{io}} & \\ s.t. & \theta_j = \frac{\sum_{r=1}^{s} v_r \cdot y_{rj}}{\sum_{i=1}^{m} u_i \cdot x_{ij}} \leq 1 & j = 1, ..., n \\ & u_i, v_r \geq 0 & i = 1, \dots, m;\ r = 1, \dots, s. \end{array} \tag{2.3}$$

Note that model (2.3) determines the efficiency of unit o, with its own vector of weights (these ones that maximizes the efficiency ratio) subject to a common set of constraints such that the efficiency score is not greater than the unity for the n DMUs. Model (2.3) must be computed n times, one for each DMU. An efficient unit is characterized by an efficiency score (θ_o) equal to the unity. The remaining units, which achieve a value lower that the unity, are considered inefficient.

Model (2.3) can be transformed into a linear programming model with some algebraical transformations [13]. The previous model is equivalent to the following expression

$$\begin{array}{lll} Max & \sum_{r=1}^{s} v_r \cdot y_{ro} & \\ s.t. & \sum_{i=1}^{m} u_i \cdot x_{io} = 1 & \\ & \sum_{r=1}^{s} v_r \cdot y_{rj} - \sum_{i=1}^{m} u_i \cdot x_{ij} \leq 0 & j = 1, \ldots, n \\ & u_i, v_r \geq 0 & i = 1, \ldots, m;\ r = 1, \ldots, s. \end{array} \tag{2.4}$$

Model (2.4) is referred as CCR-model (in reference to the initial of its authors: Charnes, Cooper and Rhodes).

Two different specifications of DEA models can be considered: output-oriented, in which each units tries to maximizes its vector of output for a given amount of input; and input-oriented, in which the units tries to optimizes the amount of consumed inputs to produce a given amount of output. Note that the objective of the model implies respectively the determination of the maximum radial (proportional) reduction of inputs and the expansion of the outputs, such that the unit under evaluation is included in the production possibility set, constructed as a linear hull of the observed values of the n DMUs. Efficient units, since they are located at the efficiency frontier, do not admit any reduction of the vector of inputs, which is reflected by an efficiency score equal to the unity.

DEA-models can deal with both constant returns to scale (CRS) and variable returns to scale (VRS). Model CCR considers that all the units operate under constant return of scale. In [8] the model with VRS assumption is proposed (commonly referred as BCC model). The model includes a convexity condition in the construction of the production possibility set. An interested reader can find a more extended explanation about the DEA methodology in [16] or [20] among others.

Nevertheless, the application of DEA and the development of models has vastly exceeded its initial objectives, by generating a wide number of models and procedures, all of which are characterized by an endogenous determination of weights. That is, the weighting vectors are determined as a variable of the problem and are not externally fixed by the decision makers.

Those extensions includes the development of alternative models and the inclusion of variables which initially do not fit with the methodology. Among the models proposed as an alternative to the radial measures, one of the most applied is the additive model. This model was initially proposed in [15]. In contrast to standard CCR and BCC models, which consider a radial measure to compute the distance to the efficient frontier, the additive model considers the maximization of the distance to the efficient frontier to evaluate the performance of each DMU. The basic expression of the additive model with VRS is

$$\begin{array}{lll} Max & \sum_{i=1}^{m} s_{io}^{-} + \sum_{r=1}^{s} s_{ro}^{+} & \\ s.t. & \sum_{j=1}^{n} \lambda_j \cdot x_{ij} \leq x_{io} - s_{io}^{-} & i = 1, \ldots, m \\ & \sum_{j=1}^{n} \lambda_j \cdot y_{rj} \geq y_{ro} + s_{ro}^{+} & r = 1, \ldots, s \\ & \sum_{j=1}^{n} \lambda_j = 1 & \\ & \lambda_j, s_r^{+}, s_i^{-} \geq 0 & \end{array} \tag{2.5}$$

This family of models deals directly with input excesses and output shortfalls (proposing a slack-based efficiency measure). Although this model can discriminate between efficient and inefficient DMUs by the existence of slacks, it has no means of gauging the depth of inefficiency, as can the efficiency measure in the CCR and BCC models. For a detailed discussion of the features of additive models, see, for instance, [42] and [29]. This last paper is particularly interesting for the sake of this paper as the authors develop a model in which weights to differentiate between the factors (inputs and outputs) are included.

Note that model (2.5) is a non-oriented model; both inputs and outputs can be modified by inefficient units to reach the efficient frontier. The projections of observed values (denoted respectively by to the efficient frontier are determined as

$$\begin{array}{l} \hat{y}_{ro} = y_{ro} + s_r^{*+} \\ \hat{x}_{io} = x_{io} - s_i^{*-}; \end{array} \tag{2.6}$$

where s_r^{*+} and s_i^{*-} denote the optimal values of the slacks determined when model (2.5) is computed. These quantities represent the differences between the observed values and the corresponding reference point. The projected efficient point is reached by reducing inputs and/or increasing outputs so as to maximize the sum of the slacks in the objective function (this is why models the objective functions of this class of models are also referred to as slack-based measures). The original non-oriented model can alternatively be transformed to an input-oriented or output-oriented model whereby only the corresponding slack variable is considered in the objective function. Also the CRS model can be considered just eliminating the convexity constraint $\sum_{j=1}^{n} \lambda_j = 1$.

In [41], a weighted additive model is proposed. The model includes a vector of weights for the slacks in the objective, respectively $g^+ = (g_1^+, \dots, g_s^+)$ and $g^- = (g_1^-, \dots, g_m^-)$ for outputs and inputs slacks, which may be determined either subjectively or objectively in a separate procedure. The model with the assumption of VRS is transformed into the following,

$$\begin{array}{lll} Max & \sum_{i=1}^{m} g_i^- \cdot s_{io}^- + \sum_{r=1}^{s} g_r^+ \cdot s_{ro}^+ & \\ s.t. & \sum_{j=1}^{n} \lambda_j \cdot x_{ij} \leq x_{io} - s_{io}^- & i = 1, \dots, m \\ & \sum_{j=1}^{n} \lambda_j \cdot y_{rj} \geq y_{ro} + s_{ro}^+ & r = 1, \dots, s \\ & \sum_{j=1}^{n} \lambda_j = 1 & \\ & \lambda_j, s_r^+, s_i^- \geq 0 & \end{array} \tag{2.7}$$

These weighting factors can be utilized in order to ensure that the units of measure associated with the slack variables do not affect the optimal solution. Note that the original additive model fails to satisfy the property of unit invariance. That is, the projections of the inefficient units on the efficient frontier depend on the scales used to measure each variable, which implies that the efficient measure does not have an intuitive interpretation [25] since the objective is a sum of incommensurable slacks. It is necessary, therefore, to pre-standardize the original dataset when the variables are measured in diverse units. In contrast, the additive model is translation invariant, which renders it an optimal option to handle with negative values (since they can be transformed to positive values by adding an adequate positive quantity).

From an economic point of view, these weighting factors represent the marginal worth of the corresponding slack. Weights are associated with unit cost and unit prices of excess and shortfall slack variables. Hence the sum of weighted slack represents an approximation of the total cost of inefficiencies [9].

For both radial and additive models, standard DEA-models assume certain basic features. Among others, must be cited the consideration of positive real values for variables (inputs and outputs); that all the outputs are desirable (in the sense that more is always preferred to less); the assumption that all the variables are controllable by DMUs (i.e. all variables, inputs and outputs can be modified by the units to achieve the efficient frontier); and that, once the efficient frontier is identified, inefficient DMUs reach this frontier by increasing the observed output values, decreasing the observed input values or by simultaneously modifying both variables. This depends on the orientation of the model: output-oriented, input-oriented, or non-oriented models, respectively.

However, many real-world situations can be found in which these assumptions are not verified. For those situations, a number of variations over original DEA models have been developed. Among others, for those cases in which real values do not fit the data available, several proposals can be found. See, among others, [18] and [19], where the inclusion of ordinal data and data on categorical variables is studied; [30] where integer values are considered or [22] and [39] in which the inclusion of undesirable outputs is studied.

In this paper, a new model for the evaluation of the efficiency is proposed, which takes ideas from additive models and the consideration of non-standard variables. In particular, we consider situations in which only the redistribution of the observed output values will be permitted for the efficiency to be attained, and not the incorporation of new units to increase the value of the output vector.

2.2.1. A DEA-inspired model to evaluate the efficiency in the presence of percentages

In this section, a variation of additive model that permit to include percentages as values is developed. Le consider that the outputs represents percentages of categories of the same variable. This supposes that in every case, for both observed and projected values, the sum is equal to 100. We consider that the categories are ranked from the less to most valued ones.

Both features have important implications for the benchmarks and the way in which the inefficient units are projected to the efficient frontier. Necessarily, the improvement of the observed value of outputs must be carried out by a reallocation of the units from the less valued categories to the most ones. This is the unique alternative to improve the value of the outputs since increasing the value of the observed output (without reducing any other) is not a feasible option.

For this task, we propose a model inspired in the additive model described previously. Consider a set of n DMUs which are being evaluated with respect to the m inputs and one output separated in s categories. It is interesting to bear in mind that this supposes to consider in practice s outputs (each y_{rj} represents the values observed for DMU j in category r, with $r = 1, \ldots, s$).

Starting with the weighted additive model (2.7), consider a weighted output-oriented model. The evaluation of the DMU o is carried out by computing

$$\begin{array}{ll} Max & \sum_{r=1}^{s} g_j^+ \cdot s_r^+ \\ s.t. & \sum_{j=1}^{n} \lambda_j \cdot x_{ij} \leq x_{io} - s_i^- \quad i = 1, \ldots, m \\ & \sum_{j=1}^{n} \lambda_j \cdot y_{rj} \geq y_{ro} + s_r^+ \quad r = 1, \ldots, s \\ & \sum_{j=1}^{n} \lambda_j = 1 \\ & \lambda_j, s_r^+, s_i^- \geq 0. \end{array} \tag{2.8}$$

In the context described above, the only way to improve the efficiency for an inefficient unit is to reallocate units across categories. That is to say, if one output is increased in one unit then it necessarily implies a reduction by the same amount in one or more than one of the remaining outputs. We propose the following variation regarding the output-oriented weighted additive model

$$\begin{array}{ll} Max & \sum_{r=1}^{s} g_r \cdot (s_r^{++} - s_r^{+-}) \\ s.t. & \sum_{j=1}^{n} \lambda_j \cdot x_{ij} \leq x_{io} - s_i^- \quad i = 1, \ldots, m \\ & \sum_{j=1}^{n} \lambda_j \cdot y_{rj} \geq y_{ro} + (s_r^{++} - s_r^{+-}) \quad r = 1, \ldots, s \\ & \sum_{j=1}^{n} \lambda_j = 1 \\ & \sum_{r=1}^{s} s_r^{++} - \sum_{r=1}^{s} s_r^{+-} = 0 \\ & \lambda_j, s_i^-, s_r^{++}, s_r^{+-}, g_r \geq 0; \end{array} \tag{2.9}$$

where g_r represents the weighting factor assigned to the the rth category. The vector $g = (g_1, \ldots, g_s)$ has to be constructed in order to assure that the relative importance of the categories are well represented. This can be a set of incomplete information, represented by a set of constraints with g_r variables (in this case, model (2.9) is not a linear model) or it contains a numerical value, objectively or subjectively determined. In that case, it is easy to see that model (2.9) is a linear programming model. In any case, considering that latter levels are better that prior ones, the relation between components of vector g must hold: $g_r \leq g_{r+1}$, for every $r = 1, \ldots, s-1$.

It is interesting to note that the slack variables of the output have been divided into two separated variables denoted by s_r^{++} and s_r^{+-}. The outputs represent percentages so both observed and projected values must verify that the sum is equal to 100. This implies that any modification of the observed value must be carried out by a reallocation. That is to say, if one output increase (this raise is measured by variable s_r^{++}) this necessarily implies that other(s) is(are) reduced (denoted by variables s_r^{+-}) in order to assure that the sum of the s outputs is equal to 100.

The objective function of model (2.9) implies that the projected efficient point is reached by increasing certain levels (the most valued ones) and reducing others (the least valued), obtained from the maximization of augmentations and reductions through the objective function of (2.9). Note that the projections only affect the observed output values (output-oriented model), such that $\hat{y}_{ro} = y_{ro} + s_{ro}^{++} - s_{ro}^{+-}$.

The first and second set of constraints includes the classic DEA production structure and therefore all the units have to verify that $\sum_{j=1}^{n} \lambda_j x_{ij} \leq x_{io}$ and that $\sum_{j=1}^{n} \lambda_j y_{rj} \geq y_{ro}$. Equivalent to said models, the condition of efficiency for the DMU o under model (7) is that the value of all slack variables is zero. That is to say, efficient units lead the constraints to the equality, and hence modifications are not possible. In any case, the observed values y_{ro} plus the optimum increase s_{ro}^{++} or minus the optimum decrease s_{ro}^{+-} will be compatible with the possibility production set. The output-orientation supposes that the modifications of inputs are not valuable. In this case, the target of the DMUs is the optimization of the observed outputs values (performance in mathematics test) for a given vector of inputs (resources assigned to the educational system).

The restriction $\sum_{j=1}^{n} \lambda_j = 1$ is included in order to consider the VRS assumptions. By deleting that constraint, a model under constant return of scale would be constructed. Note that alternative assumptions over the returns of scale structure are also feasible.

The starting point of the DMU under evaluation o is its observed value of inputs and outputs. It is easy to see that a solution such that $\lambda_o = 1$ and $\lambda_j = 0$ for every $j \neq o$; and $s_r^{++} = s_r^{+-} = s_i^{-} = 0$ for every i, r always exists, therefore the model is feasible.

The model proposes a modification of the output vector only if it involves a positive value of the objective function of (2.9). This is equivalent to a new distribution of the values y_{or} which implies the movement of units from the less valued outputs to the most valued outputs. The improvement of the output is measured through the weighted sum of the differences $(s_r^{++} - s_r^{+-})$. It is important to highlight how constraint $\sum_{r=1}^{s} s_{1ro}^{+} - \sum_{r=1}^{s} s_{1ro}^{-} = 0$ guarantees that only reallocations of units across the s outputs are permitted, and not an increment of total output, to improve the efficiency of DMU o is feasible.

By considering a vector of weights such that $g_r \leq g_{r+1}$, an inverse distribution (in which the worst categories are globally increased at the expense of the best categories) is not considered by the model. Bearing in mind that if one level is increased in a unit, this necessarily implies that any other (considering the simplest case) decreases by the same quantity this modification only holds in those cases in which the objective function is positive (which only occurs if the difference between better and worse levels is positive). Otherwise, the result of the objective is negative and does not improve the initial valuation of the unit.

Note that if no modifications on the outputs are carried out, then the value computed by the model is zero. Movements across outputs will only be carried out if the vector of outputs is improved (the aggregated measure) such that the values of better variables increase at the expense of a decrease in the values of the worse variables.

2.3. Evaluating the efficiency of educational systems of OECD countries

In this Section, a proposal for the evaluation of the efficiency of the educational systems of OECD countries is studied. Similar to the main papers revised in Section 1, this study is based on the information of the PISA programme. Several studies about efficiency in education are based on the information contained in the PISA database. Some of these studies were referred in the literature review in Section 1. We consider a set of 34 OECD countries (all the OECD countries included in PISA except Greece, since the data of one of the input considered is not available).

With regard to the inputs variables, although each proposal consider a particular set of variables, most of them try to include the classical division of inputs: raw material, physical and human capital. In this paper, we consider, as does the PISA report itself, four types of resources needed for learning: financial resources, human resources, material resources, and resources of time.

As an indicator of the intensity of financial resources invested by each country in education, we use the cumulative expenditure by educational institutions per student from 6 to 15 years old measured in equivalent USD converted using purchasing-power parities (this input is labelled as I_1). We consider it a very convenient proxy for the financial inputs as it takes into account the long-term nature of the learning process. Moreover, it uses a converted unit that enables various countries to be compared regardless of their cost of living.

With respect of human resources, teachers represent the most important part, and hence we use the student-teacher ratio. PISA provides the average number of students per teacher in every country. In order to use it as an input in the DEA model, the inverse of this ratio is calculated, that is, the number of teachers divided by the number of students (I_2).

The third kind of input PISA identifies in the learning process is that of material resources. Schools need certain resources such as facilities, classrooms, heating,... Currently, countries are also making a special effort to provide students with technological material, such as access to the Internet and computers. Following [1], technological material is used here as a proxy for the material resources. Specifically, we use the number of computers available for educational purposes in the school divided by the number of students (I_3).

The last type of resource that education requires is time.This variable measures the time per week spent in school in regular mathematics lessons, expressed in hours (I_4). It is important to highlight that the selection of the time in mathematics is justified by the selection of the outputs. The evaluation is focused is the performance in this topic.

The following table (Table 2.1) summarizes the main characteristics of the variables described above.

Table 2.1: Descriptive statistics of input variables

	I_1	I_2	I_3	I_4
Max	187,458.81	0.14	1.52	7.20
Min	27,848.44	0.42	0.16	2.40
Average	90,293.98	0.27	0.78	3.66
Std. Dev.	34,585.41	1.20	0.31	0.84

Most authors, when choosing the output for DEA-models, use the mean scores of the topics evaluated in PISA. These scores are determined based on the so-called plausible values. These are found within the probability distribution estimated for a student's score in each test. Therefore, for every student's test, PISA provides five plausible values, where these are the probabilities for the student to obtain each of the values.

The PISA mean scores are based on the Rash model, see [37] and [43], which uses plausible values instead a particular mean value for each student's knowledge. These values are random values obtained from the distribution function of the results estimate from the results obtained in each test. They can be interpreted as a representation of the ability range of each student [45]. The determination of plausible values can be seen in detail in [33].

The main reason stated in the report for using the plausible values is the necessity for the transformation of a continuous variable (e.g. student's ability) into a discrete variable, (e.g. the scores). In this process, the plausible values have proved themselves as unbiased measures for the variable. They reduce the errors both from measuring and from the omission of underlying aspects that have not been considered specifically in the test.

The computation of these plausible values, however, presents numerous disadvantages for researchers since it is necessary to calculate any given statistic, e. g. the mean, for every plausible value and then to compute the average for every individual student, which renders this method cumbersome. If the investigator were to omit this procedure, then the results could be biased.

In order to avoid all these problems, we propose an alternative vector of outputs: the percentage of students of each country in the different proficiency levels. As an alternative way of measuring the results in every subject, specifically mathematics, the PISA report classifies the students depending on their achievement in seven categories, called proficiency levels. The way these proficiency levels are constructed take into account not only the abilities of the students but also the difficulty of the items, thereby constituting a scale of literacy. In doing so, every proficiency level can be described as a group of abilities we can expect from the students contained within this level. According to the PISA report, its aim is to provide useful information for decision-making and predictions about education policies. This is why, in a complementary way, various related reports published by the OECD provide the percentage of students in each level. Working with these results enables any problems regarding plausible values to be avoided.

To consider the vector of proficiency levels as the output of the model, a modified efficiency evaluation model is required. It is important to bear in mind the characteristics of these values and to adapt the existing models to these particularities.

It is interesting to point out how these values can be easier for policy makers to evaluate and interpret. With these variables, the benchmark of the efficiency model are represented by the percentages of students that must be in each level for an observed vector of resources. The improvement is measured through the number of students that must achieve a particular level of proficiency in the test.

Therefore, the results in each topic in the PISA are standardized with a mean of 500 and a standard deviation of 100. Seven proficiency levels are constructed in which the students are allocated depending on their results in the topic (see [34]). In this paper, the results obtained in Mathematics have been considered.

The first level comprises the students with scores below 357.77 points. The following levels includes students with scores included in the following intervals: second level from 357.77 to 420.07 points, third level from 420.07 to

482.38, fourth level 482.38 to 544.68 points, fifth from 606.99 to 669.3 points. The last level, the most valued one, includes the students with more than 669.3 points.The data considered here includes the percentages of students that achieve each proficiency level, as a means of reflecting the performance of the educational system. The outputs vector contain the seven level described above, labelled from O_1 (percentage of students with scores below 355.77 points) to O_7 (percentage of students with scores over 669.3 points). The data of the statistics of the seven outputs considered (proficiency levels) are summarized in the following table (Table 2.2).

Table 2.2: Descriptive statistics of output variables.

	O_1	O_2	O_3	O_4	O_5	O_6	O_7
Max	25.53	31.11	28.27	29.98	25.92	15.05	6.58
Min	2.22	7.75	17.23	12.89	3.20	0.31	0.02
Average	8.47	14.89	22.55	24.81	18.60	8.37	2.31
Stand. Dev.	5.68	5.11	2.55	3.59	5.11	3.46	1.44

The special characteristics of these proficiency levels require an adaptation of the model for the evaluation of the efficiency. The model developed in Section 2.2.1, model (9), is fitted to these particular features. The consideration of an additive model (versus a radial model) is based on the characteristics of the feasible variations for the outputs. In order to reach the frontier by modifying the percentages in each level, these outputs could increase in different quantities and not radially. Note that levels denote different importance; obviously higher scores imply a larger importance. An increase of the number of students in the latter categories requires, from the educative system, a greater effort than an increase of the number of students in the previous ones. From this point of view, the efficient countries would be those that have larger percentages of students in the better proficiency levels and smaller percentages in the worse categories. It is important to bear in mind that the outputs represent percentages; consequently the sum for each DMU has to be equal to 100 not only for the observed values but also for the projection in the efficiency frontier. That is, if the country achieves more students in better categories this is because it has fewer students in the worse categories.

To compute the model, a set of 34 units (OECD countries,all the OECD countries included in PISA except Greece), are being evaluated with respect to the four inputs previously and the vector of output which represents the seven proficiency levels. To mitigate the effect of outliers and/or the existence of errors, the models has been robustified using the concepts proposed in [12]. To this end, 2,000 computation rounds of each model are obtained with a sub-sample of 28 randomly selected units.

2.3.1. Discussion of results

Table 2.3 summarizes the results obtained for a weighting vector such that $g = (1, 2, 4, 8, 16, 32, 64)$. Note that the particular value assigned to g could be done in several ways, in a subjective way (like the one we propose) or by means of an additional procedure that measure the relative importance of each level. Note that each component g_r of vector g tries to approximate the marginal worth of the corresponding slack (in relative terms). Thus, the objective function of model (9) approximate the total cost of the inefficiency on the unit. The selection of g can proceed from a political decision in order to emphasise the relative importance or effort in the reallocation of one unit from one category to the other. Or alternatively, the determination of g may result from a technical analysis. In any case, the consideration of alternative values for vector g affects to inefficient units since the sum of slacks are weighted in a different way. Therefore, the construction of a ranking of unit based on the optimal value of the model would be affected. But note that those unit characterized as efficient are not affected by any modification in vector g.

In Table 2.3, the most relevant results of the application of our model are shown. For every country, the net (positive or negative) variation for each proficiency level is provided. It represents the amount by which that specific country must increase or decrease the percentage of students in that category to become efficient, calculated as the difference between the s_r^{++} and the s_r^{+-} variables. Those countries that lie on the efficiency frontier show a 0 in all the slacks.

From this analysis, the countries can be classified into two different groups, efficient (denoted in bold) and inefficient. In the first group, we find Austria, Chile, Czech Republic, Denmark, Estonia, Finland, Hungary, Ireland, Israel, Japan, Korea, Mexico, Netherlands, Poland, Slovak Republic, Slovenia, Switzerland and Turkey. The way in which these countries achieve efficiency differs greatly. Certain countries, such as Chile, Turkey and Mexico, despite their low results, have an efficient educational system, because their investment in education is comparatively smaller.

The results in Table 2.3 must be interpreted as follows. Each value represents the percentage of the net variation of the corresponding level. Let consider the case of Spain. With the resources considered the system is

Table 2.3: Main results: Slack variables results

DMU	O_1	O_2	O_3	O_4	O_5	O_6	O_7
Australia	-3.68	-4.35	-2.69	1.31	4.57	3.48	1.36
Austria	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Belgium	-2.85	-2.87	0.16	2.25	1.65	0.53	1.13
Canada	1.57	0.07	-2.18	-2.86	-0.32	2.02	1.71
Chile	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Czech Republic	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Denmark	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Estonia	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Finland	0.00	0.00	0.00	0.00	0.00	0.00	0.00
France	-5.18	-4.97	-0.72	3.06	2.97	2.69	2.14
Germany	-0.17	-0.37	-0.31	0.02	0.25	0.32	0.26
Hungary	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Iceland	-4.69	-6.64	-6.40	0.38	7.35	6.59	3.42
Ireland	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Israel	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Italy	-2.80	-2.89	-2.27	1.26	2.85	2.55	1.30
Japan	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Korea	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Latvia	-0.22	-4.02	-6.23	-1.69	5.01	5.23	1.92
Luxembourg	-3.67	-6.88	-4.69	0.63	5.04	6.06	3.52
Mexico	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Netherlands	0.00	0.00	0.00	0.00	0.00	0.00	0.00
New Zealand	-1.22	-2.66	-1.46	0.83	2.07	1.87	0.57
Norway	0.01	-1.72	-4.94	-3.02	2.22	4.53	2.92
Poland	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Portugal	-3.09	-4.69	-3.84	0.07	3.53	4.78	3.24
Slovak Republic	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Slovenia	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Spain	-3.62	-5.07	-4.04	0.09	5.17	5.10	2.38
Sweden	-2.38	-3.20	-3.90	-0.51	3.61	3.95	2.27
Switzerland	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Turkey	0.00	0.00	0.00	0.00	0.00	0.00	0.00
United Kingdom	-2.52	-4.01	-4.98	-1.83	4.17	5.60	3.56
United States	-5.07	-6.70	-5.30	1.87	6.94	5.98	2.28

characterized as inefficient. The improvement proposed implies the raise of the outputs from O_4 (propose an increase of the percentages of students with scores between 482.28 and 544.68 of 0.48 points) to O_7 (the feasible increase in the percentage of students with a score over 669.30 is 2.44 percentage points) at expenses a reduction of the remaining ones. The excess in the first level, students under 357.77 points, is 4.25 percentage points. For the following two levels, the reduction is 5.55 and 4.15 respectively. It is clear that this reallocation of students from the worst valued level to the best ones would suppose a improvement in the aggregated value of the output and in the results of the systems (better students' results with a given amount of resources). But also the model guarantees that the proposed reference value is feasible in the sense that in included in the possibility production set constructed with the observed units. This feature explains movements like the one proposed for Norway, in which a raise in the first level (the one with student with lowest scores) is proposed. This is explained by the requirement of the classic constraints of DEA for being enveloped by the efficiency frontier. Even so, the aggregated value of the projected output vector would increase.

A subset of countries found in the first group involve those systems that obtain good results in PISA but need to invest resources above the mean. Austria and Finland are found in this subset. Finally, there are countries that achieve great levels of proficiency but employ fewer resources than the rest of the members of the OECD. This is the case of Estonia, Korea, Netherlands and Poland. The inefficient countries are those which, given their available resources, should obtain better results in PISA. Among these, we can mention Portugal, Sweden and Italy as the countries that are farthest from the efficiency frontier.

Another important result provided by the model is the units of reference for each country. In order to become efficient, the inefficient units have to increase their outputs until they reach the efficient frontier. The inefficient countries should modify their outputs until they reach the levels of those efficient countries that have a similar structure of inputs. In Table 2.4 (see Annex), the inefficient countries can be seen in the first column and the countries which they should imitate appear in the following columns with their corresponding lambda value. In this case, the reference sets have been obtained by computing model (9) for the complete set of units.

From this point of view, and given that these countries are efficient, we can consider the best countries in terms of educational efficiency to be those which constitute a reference for other countries. Since these units not only are located in the efficiency frontier but also there are certain units with a similar combination of inputs and outputs that are revealed as less valued. In this respect, Korean, as the reference for 14 countries, Switzerland for 9 and Estonia, Ireland and Netherlands for 5 constitute the reference set for the inefficient countries.

On the other hand, we found efficient countries which do not constitute the reference set for any inefficient unit. This is the case of Austria, Czech Republic, Israel, Mexico and Turkey. In brief, this set of countries constitutes *extreme* cases, since their combination of inputs and outputs are characterized as efficient, they are quite different of the remaining countries under evaluation. That is, the observed values of inputs and output of these units are quite different of the other ones and this could be the cause to be part of the efficiency frontier (and not a good performance).

2.4. Concluding remarks

In this paper, an study on the efficiency of the educational systems of the OECD countries has been developed. The study is based on the application of the Data Envelopment Analysis (DEA) methodology and the dataset provided by the PISA report. The analysis has been done considering the resources of each system and results of each economy in the mathematical test in PISA 2015.

The PISA report assesses the learning achievement of the students and classifies them into seven level of proficiency, depending on their abilities. Therefore, for each country, the percentage of students in every proficiency level is available. We propose to use these values instead of the mean score on a particular topic to evaluate each country. Using this variable as an output permits to avoid the consideration of plausible values and a straight interpretation of the benchmarks but requires a specific model, such as that developed in this paper. It is easy to see that the total amount of the different outputs cannot increase, but can only be reallocated, since we are dealing with percentages.

A variation of the weighted additive DEA model to reallocate outputs has been proposed. Contrary to radial models, the strategy to achieve efficiency of additive models allows each variable (inputs and outputs) to be modified by a particular quantity. By including a vector of weights, the relative importance of each variable can be suitably represented. This feature allows us to take into account the differences in cost or the effort the units must exert to increase or reduce, respectively, the diverse outputs and inputs.

We develop a model for this particular context. We consider a situation in which the aggregate value of the output is a fixed value and the strategy to improve the performance is a reallocation of units across the outputs. Increasing a particular output necessarily implies a reduction of any other output (in order to maintain the aggregate value constant and equal to 100) and the DMUs are interested in moving units from the least to the most valued level.

As a result, the countries have been classified into two different groups: the efficient and inefficient units. The first group is identified by null values in all the slacks. Additionally, the model provides an efficient reference country with similar input structure for every inefficient country in order to improve their results. The countries that serve as a reference for the greatest number of educational systems are Korea and Japan. For the inefficient units, the values of slacks and projected outputs can serve as an accurate guide to political actors. These values represents a target for the number of students that achieve each proficiency level, and have been obtained considering other countries with a similar structure of inputs and outputs.

Future lines of research could carry out an in-depth study into the causes that provoke the differences between the educational systems, and could analyse how to make reforms that would solve the problems of the inefficient countries, since DEA models can only offer general guidelines. Improvements in the theoretical model are also possible, among others, the inclusion of additional information (complete or incomplete) or the consideration of additional procedures in order to determine the vector of weights associated to each level.

Bibliografía

[1] Agasisti, T. (2011) *Performances and spending efficiency in higher education: A European comparison through non-parametric approaches*. Education Economics 19(2): 199-224.

[2] Agasisti, T.,(2013) *The efficiency of Italian secondary schools and the potential role of competition: a data envelopment analysis using OECD-PISA 2006 data*. Education Economics 21(5): 520-544 .

[3] Agasisti, T. (2014) *The efficiency of public spending on education: An empirical comparison of EU countries*. European Journal of Education 49(4): 543-557.

[4] Afonso, A., St. Aubyn, M. (2006) *Non-parametric approaches to education and health efficiency in OECD countries*. Journal of Applied Economics 8(2): 227-246.

[5] Afonso, A., St. Aubyn, M. (2006) *Cross-country efficiency in secondary education provision: a semi-parametric analysis with non-discretionary inputs*. Economic Modelling 23(3): 476-491.

[6] Ali, A.I., and Seiford, L.M. *The mathematical programming approach to efficiency measurement*, in: H. Fried, K. Lovell and S. Schmidt (eds.), The Measurement of Productive Efficiency: Techniques and Applications, (1993) Oxford University Press, Oxford.

[7] Aristovnik, A. (2011) *An analysis of the efficiency of education spending in central and eastern Europe*. EconPapers http://EconPapers.repec.org/RePEc:isv:mklp11:277-286.

[8] Banker, R.D., Charnes, A., Cooper,W.W. (1984) *Some models for estimating technical and scale inefficiency in data envelopment analysis*. Management Science 30: 1078-1092.

[9] Bardhan, I., Bowlin, W.F., Cooper, W.W., Sueyoshi, T. (1996) *Models for efficiency dominance in data envelopment analysis. Part I: Additive models and MED measures*. Journal of the Operational Research Society of Japan 39: 322-332.

[10] Bessent, A.M., Bessent, E.W. (1980) *Determining the comparative efficiency of schools trough data envelopment analysis*. Educational Administration Quarterly 16(2): 57-75.

[11] Bessent, A.M., Bessent, E.W., Kennington, J., Reagan, B. (1982) *An application of mathematical programming to asses productivity in the Houston independent school district*. Management Science 28: 1355-1367.

[12] Cazals, C., Florens, J.P., Simar, L. (2002) *Nonparametric frontier estimation: A robust approach*. Journal of Econometrics 106(1), 1-25.

[13] Charnes, A., Cooper, W.W., Rhodes, E. (1978) *Measuring the efficiency of decision making units*. European Journal of Operations Research 2: 429-444.

[14] Charnes, A., Cooper, W.W., Rhodes, E. (1981) *Evaluating program and managerial efficiency: an application of Data Envelopment Analysis to Program Follow Through*. Management Science 27: 668-697.

[15] Charnes, A., Cooper, W.W., Golany, B., Seiford, L. (1985) *Foundations of Data Envelopment Analysis for Pareto-Koopmans efficient empirical production functions*. Journal of Econometrics 30(1): 91-107

[16] Charnes, A., Cooper, W.W., Lewin, A.Y., Seiford, L.M. (Eds.) (1994) *Data Envelopment Analysis: Theory, methodology, and applications*. Springer Netherlands.

[17] Clemens, B. (2002) *How efficient is education spending in Europe?*. European Review of Economics and Finance 1: 3-26.

[18] Cook, W.D., Kress, M., Seiford, L.M. (1993) *On the use of ordinal data in Data Envelopment Analysis*. Journal of the Operational Research Society 44 (2): 133-140.

[19] Cook, W.D., Kress, M., Seiford, L.M. (1996) *Data Envelopment Analysis in the presence of both quantitative and qualitative factors*. Journal of the Operational Research Society 47: 945-953.

[20] Cooper, W.W., Seiford, L.M., Tone, K. (2007) *Data Envelopment Analysis: A comprehensive text with models, applications, references and DEA-solver software*. Springer-Verlag US.

[21] Cordero-Ferrera J.M., Crespo-Cebada, E., Pedraja-Chaparro, F., Santin-Gonzalez, D. (2011) *Exploring educational efficiency divergences across Spanish regions in PISA 2006*. Revista de Economia Aplicada 19(57): 117-145.

[22] Färe, R., Grosskopf, S., Lovell, C.A.K., Pasurka, C. (1989) *Multilateral productivity comparison when outputs are undesirable: A non-parametric approach*. The Review of Economic and Statistics 71: 90-98.

[23] Farrell, M.J. (1957) *The measurement of the efficiency*. Journal of the Royal Statistical Society Series A 120(3): 253-290.

[24] Gimenez, V., Prior, D., Theme, C. (2007) *Technical efficiency, managerial efficiency and objective setting in the educational system: An international comparison*. Journal of the Operational Research Society 58: 996-1007.

[25] Gouveia, M.C., Dias, L.C., Antunes, C.H. (2008) *Additive DEA based on MCDA with imprecise information*. Journal of the Operational Research Society 59(1): 54-63.

[26] Hanushek, E. (1979) *Conceptual and empirical issues in the estimation of educational production functions*. Journal of Human Resources 14(3): 351-288.

[27] Johnes, J. (2006) *Data Envelopment Analysis and its application to to the measurement of efficiency in higher education*. Economics of Educations Review 25(3): 273-288.

[28] Levin, H. (1974) *Measuring the efficiency in educational production*. Public Finance Quarterly 2: 3-24.

[29] Lovell, C.A.K.. Pastor, J.T. (1995) *Units invariant and translation invariant DEA models*. Operations Research Letters 18: 147-151.

[30] Lozano, S., Villa, G. (2006) *Data Envelopment Analysis of integer-values of inputs and outputs*. Computers and Operational Research 33: 3004-3014

[31] Mancebon M.J., Bandres, E. (1999) *Efficiency evaluation in secondary schools: The key role of model specification and of ex post analysis of results*. Education Economics 7(2): 131-152.

[32] Mancebón, M.J., Calero, J., Choi, A. y XimÈnez-De-Emb˙n, D.P. (2012) *The efficiency of public and publicly subsidized high schools in Spain: Evidence from PISA-2006*. Journal of the Operational Research Society 63(11): 1516-1533.

[33] OECD (2012) *PISA 2009 Technical Report*. PISA, OECD Publishing. http://dx.doi.org/10.1787/9789264167872-en.

[34] OECD (2013) *PISA 2012 Results: What Makes Schools Successful? Resources, Policies and Practices (Volume IV)*. PISA, OECD Publishing. http://dx.doi.org/10.1787/9789264201156-en.

[35] OECD (2014) *PISA 2012 Results: What Students Know and Can Do? Student Performance in Mathematics, Reading and Science"(Volume I, Revised edition, February 2014)* OECD Publishing. http://dx.doi.org/10.1787/9789264201118-en.

[36] Portela, M.C.A.S., and Thanassoulis, E. (2001) *Decomposing school and school type efficiency*. European Journal of Operational Research, 132(2): 114-130.

[37] Rasch, G. (1960) *Probabilistic Models for Some Intelligence and Attainment Tests*. Copenhagen, Denmark: Danmarks Paedogogiske Institut, (Chicago: University of Chicago Press).

[38] Santin, D., Sicilia, G. (2015) *Measuring the efficiency of public schools in Uruguay: main drivers and policy implications*. Latin America Economic Review 24(5): 1-28.

[39] Scheel, H. (2001) *Undesirable outputs in efficiency evaluation*. European Journal of Operational Research 132: 400-410.

[40] Sutherland, D., Price, R., Gonand, F. (2009) *Improving public spending efficiency in primary and secondary education*. OECD Journal of Economic Studies 1: 1-30.

[41] Thrall RM. (1996) *Duality, classification and slacks in DEA*. Annals of Operations Research 66: 109-138.

[42] Tone, K. (2001) *A slacks-based measure of efficiency in data envelopment analysis*. European Journal of Operational Research 130 (3): 498-509.

[43] Wright, B.D., Masters, G.N. (1982) *Rating scale analysis*. Chicago: MESA Press.

[44] Worthington, A.C. (2001) *An empirical survey of frontier efficiency measurement techniques education*. Education Economics 9(3): 245-268.

[45] Wu, M., Adams, R. (2007) *Applying the Rasch model to psycho-social measurement: A practical approach*. Educational Measurement Solutions, Melbourne.

Table 2.4: Main results. Benchmarks

	Reference												
DMU	Chile	Denmark	Estonia	Finland	Hungary	Ireland	Japan	Korea	Netherlands	Poland	Slovak Republic	Slovenia	Switzerland
Australia	-	0.244	0.273	-	-	-	-	0.363	-	-	-	-	0.119
Belgium	-	-	0.065	0.167	-	-	-	0.467	-	-	-	-	0.301
Canada	-	-	-	-	-	-	-	0.824	0.176	-	-	-	-
France	-	-	0.591	-	-	-	-	0.205	-	-	-	-	0.205
Germany	-	-	-	-	-	0.447		0.349	0.107	-	-	0.086	0.010
Iceland	-	-	-	-	-	-	0.800	0.200	-	-	-	-	-
Italy	-	-	-	-	-	0.212	-	0.310	-	0.380	-	-	0.098
Latvia	-	-	0.762	-	0.238	-	-	-	-	-	-	-	-
Luxembourg	-	-	-	-	-	-	-	0.889	0.111	-	-	-	-
New Zealand	0.026	-	-	-	-	0.387	-	0.487	-	-	0.101	-	-
Norway	-	-	-	-	-	-	-	0.278	0.222	-	-	-	0.500
Portugal	-	-	-	-	-	-	0.258	0.742	-	-	-	-	-
Spain	-	0.146	0.652	-	-	0.007	-	0.174	-	-	-	-	0.022
Sweden	-	-	-	0.065	-	-	-	0.337	-	-	-	0.326	0.272
United Kingdom	-	-	-	-	-	-	0.095	0.905	-	-	-	-	-
United States	-	-	-	-	-	0.247	-	0.042	0.458	-	-	-	0.253

CAPÍTULO 3

A multiplicative composite indicator to evaluate educational systems in OECD countries

In the evaluation of any political strategy, it is essential to carry out the collection and analysis of the available data. The presentation of a complex phenomenon by means of synthetic measures can improve both the political actors' understanding of the situation and the design of new measures. In this paper[1], we deal with the problem of the construction of a multidimensional composite indicator for the evaluation of educational systems in OECD countries. In our proposal, not only are those indicators included that measure the academic outcomes, but also a group of indicators that measures the social dimension of the educational system. A variation of a methodology based on Data Envelopment Analysis is developed to construct a composite multiplicative indicator that enables inter-temporal comparison and the detection of the sources of the variation.

[1]Domínguez, C., Segovia-González, M.M., Contreras, I. (2020) *A multiplicative composite indicator to evaluate educational systems in OECD countries*. Compare: A Journal of Comparative and International Education. En prensa.

3.1. Introduction

In recent decades, a strong commitment across the globe has emerged in the area of education. There is a general belief that education can transform the lives of people and plays a crucial role in the development of countries. The Incheon Declaration [34] sets out a new vision for education with a time horizon of 2030. This report focuses on access, equity, inclusion, quality, and lifelong-learning opportunities for all the inhabitants of every country.

With this work, we strive to provide results regarding the performance of the educational systems of certain OECD countries. Educational outcomes, equity, and students' well-being are considered. To this end, the problem of the construction of a multidimensional composite indicator of educational performance is addressed We propose a modified methodology based on the *Benefit of the Doubt* (BoD) principle ([7, 8]) to construct an indicator on the basis of a geometric aggregation and the free selection of weights. The proposed indices are built in two different periods, 2012 and 2015, so that an inter-temporal comparison can be established.

The proposed index is constructed on the basis of the data included in the Programme for International Student Assessment (PISA). The PISA report, initially launched in 2000, has since been carried out every three years and constitutes a major source of information for the execution of comparisons across economies. The PISA study has steadily increased in the number of participating schools and countries. In the first edition, 265,000 students from 32 countries were evaluated. The latest edition of this report, in 2018, covered more than 500,000 students from 80 countries. The main objective of the programme is to evaluate educational systems worldwide by testing the skills and knowledge of 15-year-old students in Mathematics, Science, and Reading skills, to which financial literacy was also included as an option in 2012. In addition to data on academic achievements (results from standardised test scores), the PISA database contains a vast amount of information on students, their households, and on the schools they attend. See [27] for a detailed discussion.

Numerous papers that criticise PISA appear in the literature, including criticism regarding the statistical methodology and the ideological basis and policy arguments upon which the report is based. It is clear that the ambitious aim expressed by the directors of PISA makes this study more vulnerable to criticism than other reports that either only attempt to measure and/or evaluate a limited part of the knowledge of students or that are focused on a specific country.

With respect to the inherent political ideology, several papers are strongly critical of the pragmatism and utilitarian approach through which this report is sustained. PISA reports strive to measure students' ability to apply the knowledge acquired in school to solve problems in everyday life [27]. This conception of what education is for has been refuted by a number of authors who consider that education policy should open up its perspective to a wider range of options, beyond classic targeted initiatives oriented towards the applicability of knowledge. In addition, the "race" started by several countries to improve their PISA -rankings as the main target of their educational policies, supposes a constraining boundary on political decisions in the education context. In [40] and [32], among others, provide a detailed critical discussion regarding PISA reports.

With respect to technical aspects, the basis of the survey and certain technical aspects used for the construction of a number of indices have also been problematized. Some of the relations between variables presented in the reports have been shown to be causal relations. On this point, it must be stated that several of these criticisms regarding methodological aspects have been considered and included in the successive reports by the PISA team. See [14] for a detailed discussion.

Nevertheless, despite all the aforementioned limitations, PISA has emerged as a valuable source of information for researchers and policymakers. The author in [18] states that PISA is one of the best databases for comparative research that justifies its use in research works. The authors see the orientation of the report as an advantage, since test are independent of the school's curricula. These features have leaded to a large number of research papers based on PISA information (according to [17], at least, 654 papers had been published until 2016).

The rest of the paper is organized as follows. In the following Section, a description of the selection of variables considered is presented and justified. Next section is dedicated to the description of the methodology for the construction of the composite indicator. The following section is devoted to the construction of the index and the discussion of the results. Last section includes the main conclusions.

3.2. A composite indicator for the evaluation of the educational systems

According to the OECD glossary of statistical terms, a composite indicator is formed when individual indicators are compiled into a single index on the basis of an underlying model of the multidimensional concept that is being measured. Composite indicators (CIs) are increasingly being accepted as a useful tool for performance comparisons, benchmarking, policy analysis, and public communication in various fields such as the economy and social sciences [25].

The inclusion of indicators related to educational aspects as a part of composite indicator is prevalent in analysis related to poverty, gender well-being or social inclusion among others. However, it is not so regular to find proposals of composite indicators for evaluating education aspects, beyond the well-known ranking of higher-education institutions ([12]; [22]). Others works, deals with the evaluation of institutions in the higher education context through the proposal of indicator systems ([2]; [21]).

Previous works have developed various proposals and methodologies for the construction of CIs for the evaluation of educational systems. Recently, [37] has proposed a multidimensional index to summarise relevant aspects of educational systems. Other authors are interested in the study of aspects regarding the outcomes of the system, among which, [30] and [20] include studies on the inequality and educational poverty of educational systems. To the best of our knowledge, only [33] apply the BoD approach for the construction of composite indicators in the evaluation of national educational systems.

This paper incorporates several innovations with respect to previous references. A cross-country and cross-period evaluation is proposed, certain innovations regarding the BoD methodology are given, and the panel of indicators incorporates additional and relevant aspects into the evaluation.

The relevant aspects are classified into three main areas: academic, equity, and students' self-perception of well-being.

With respect to the academic dimension, we strive to measure outcomes of the systems. Three different and connected aspects are considered: the students' performance, the excellence, and the inclusion of the systems. Regarding the academic performance, we strive to measure the overall achievement of the students, that is, the average student learning outcome for each country. This value is reflected through the average of the standardised test scores in Science included in the PISA report.

The second aspect to be considered is the excellence of the educational system. High-quality education is essential in modern society. In order to take this aspect into account the students' proficiency levels published in PISA for each subject have been examined. The PISA report considers six levels of educational proficiency, built from the test scores in each subject (a detailed explanation can be seen in [28]).

In the PISA reports, it is established that an educational system achieves the minimum objectives if the levels of proficiency of its students lie between levels 2 and 5. Students who reach levels equal to or greater than 5, are considered to have high performance in the acquisition competences. In order to take into account the excellence of an educational system, the high performance in a certain number of the subjects is considered as the indicator. This second indicator computes the percentage of students who obtained a level equal to or greater than 5 in any of the three main subjects measured in PISA (Mathematics, Science, and Reading).

With regard to inclusion, PISA 2015 includes, as one of the desirable objectives, that all students are guaranteed access to a quality education and reach a minimum of knowledge (defined as a baseline level of skills). Our proposal involves approaching this concept by considering a third indicator that is given by the percentage of individuals who reach at least level 2 in all subjects. It appears reasonable to consider that an educational system must reach at least a baseline level of skills in all subjects to consider that it works minimally. Note that the requirements in availability and comparability of information supposes a limitation in the aspects that can be taken into account for the measurement of academic outcomes. In this cross-country analysis, only the results of standardised tests are included.

The second main aspect to be studied in the multidimensional indicator is the equity of the educational system. Equity in education constitutes a specific target of the Sustainable Development Goals set by the United Nations in 2015. Equity does not mean that all students will have the same results in education, but that their results will not be conditioned by their social, economic, or any other circumstances ([10]; [29]). Two different approaches are considered: social equity and socio-economic incidence.

With regard to social equity, ideally individuals should be able to obtain excellent academic results depending only their individual abilities, no matter how adverse the conditions in their environment may be. In order to approximate social equality, an individual indicator represented by the percentage of resilient students is considered. PISA considers that a student is classified as resilient if he or she is placed in the bottom quarter of the PISA index of Economic, Social and Cultural Status (ESCS) and performs in the top quarter of students.

These measures, percentages of resilient students for each country, are computed differently in PISA 2012 and PISA 2015 reports because the students' performance results considered in each year are based on the main academic domain: Mathematics in PISA 2012, and Science in PISA 2015. In order to establish a comparison between these two periods, a new definition of a resilient student given by [1] is considered. The main concept of a resilient student remains, but certain nuances for an easier interpretation and more robust comparisons are included, which permits the comparison between successive periods.

The characteristic of educational equity in an educational system can be seen as granting all students the possibility of achieving full development of their talent, without being hindered by their economic and social circumstances. It is clear that the measure of this point is complicated. We approximate this concept through the percentage of variation in Science performance that cannot be explained by a student's socio-economic status. We seek to reflect the capacity of a country to minimise the influence of their economic, cultural, and social background on their academic achievements.

It is interesting to point out the differences with respect to the first indicator included in this dimension. With the concept of resilient students, only those students placed in the bottom quarter of the ESCS index are considered. With the former indicator, the incidence over the complete set of students is considered. The estimation of all the values has been carried out following the technical detail included in the PISA reports [26].

The third dimension included measures the students' well-being during their scholar stage. The well-being of the students is a valuable aspect for the evaluation of an educational system. In certain cases, an education system can be excellent at producing top student outcomes, but at a cost to their well-being. The inclusion of a measure regarding self-perceptions of their individual situation, the disciplinary climate, teacher-student relations, teachers' formative assessments, and teacher support, among others, is considered here. We have considered one of the indices proposed by OECD as a proxy of the well-being of the students at school: the *Sense of belonging* index. This value attempts to measure how accepted, respected, and supported students feel in their social context at school (Goodenow and Grady 1993). Previous studies have shown a positive association of this variable with other variables, such positive disciplinary climate, participating in extracurricular activities, family support, and positive teacher-student relations. This justifies the consideration of the *Sense of belonging* index as a proxy of well-being of students at school.

The second aspect included in this third dimension is related with the time spent learning after school. It is clear that is not easy to determine the boundary that separates a situation of reasonable working time after school from a situation in which the students have to sacrifice their social lives to obtain academic achievements. We strive to compute only the overload of working time for students with this indicator. We have considered the total learning time of the students after school as a new indicator. We propose computing only the excess of time spent learning after school, by considering that this excess is measured as the difference between the time dedicated to studying after regular school hours (in diverse occupations: homework, private tutoring, etc.) with respect to the average of OECD countries, but only in those situations in which this excess is positive. On the other hand, for those countries in which the time spent learning after school is below the average, the indicator is assigned a value equal to the unity in the construction of the composite indicator, in order to prevent its influence in the final value.

In short, seven indicators are considered, separated into three main dimensions:

- Academic features.
  - Academic performance (I_1).
  - Academic excellence (I_2).
  - Academic inclusion (I_3).
- Equity of the educational system.
  - Social equity (I_4).
  - Socio-economic incidence (I_5).
- Students' well-being.
  - *Sense of belonging* index (I_6).
  - Overload of working time (I_7).

3.3. Geometric Composite Indicator under the Benefit of the Doubt principle

3.3.1. Composite Indicators

From a technical point of view, a CI is a mathematical aggregation of a set of individual indicators (often referred as sub-indicators), for the measurement of multidimensional concepts that cannot be captured by only one single indicator [25]. The process of constructing composite indicators implies a number of successive decisions [25]: selection of initial indicators, the way that they are conceptually grouped, the decision as to whether to use a data normalization method, and the choice of the method to weight and aggregate initial indicators. A complete revision of the concepts related with the construction of CIs can be seen, among others, in [23] and [24].

The weighting and aggregation phase is one of the most crucial steps in the CI construction process ([31]; [13]). In this section, a methodology based on multiplicative aggregation is developed, in which the weighting factors are determined as a variable of the model, in an effort to reduce the subjective judgements. To this end, the so-called *Benefit of the Doubt* (BoD) approach is taken as reference ([7]; [8]; [22]) in the aggregation phase. This methodological group of studies uses Data Envelopment Analysis (DEA) ([6]; [3]) as a tool to construct composite indicators.

Let a set of n units or alternatives be evaluated by means of a CI. Suppose that a set of m individual indicators have been collected, where I_{ri} denotes the individual indicator r with respect to the alternative i, with $i = 1, 2 \ldots, n$ and $r = 1, 2, \ldots, m$. The composite indicator of alternative or unit i is denoted by CI_i. For the sake of simplicity, all the individual indicators are assumed to have been adequately treated and it is also assumed that the higher the value, the better.

In order to achieve CI_i the application of the multiplicative aggregation scheme is proposed such that:

$$CI_i = \prod_{r=1}^{m} I_{ri}^{\omega_r} \geq 0, \sum_{r=1}^{m} \omega_r = 1, \qquad (3.1)$$

where ω_r denotes the normalized weighting factor associated to the rth individual indicator.

Several authors have studied the advantages of the geometric aggregation. In [11] and [42], the authors point out that the multiplicative scheme present better theoretical properties than other aggregation schemes, such as the scale-invariance and a lower degree of compensation between individual indicators. The latter supposes that this aggregation function penalizes those alternatives with lower values in some individual indicators by assigning a lower composite indicator value [39]. In addition, modifications of values in an originally low-value indicator will cause a greater variation in the CI than in high-value indicators. Consequently, analysed units are encouraged to improve their weaknesses. Furthermore, [41] and [42] found that geometric aggregation leads to minimum information loss (in comparison with several other aggregation methods).

Once the aggregation scheme is selected, the next step involves the selection of the weighting profile. The proposed scheme for weight assignation is inspired from the BoD principle, which is based on DEA methodology. DEA-based procedures for the construction of CIs permit each alternative to select its own vector of weights in order to optimise the aggregate value, under a set of common constraints in order to guarantee that CI values are bounded (usually by the unity). The major benefit from this class of models is that the weight values are adapted to unit measures of the sub-indicators and that it concludes with an endogenously determined vector of weights. That is, the weighting factors are not determined by a subjective decision of the analyst. There are a number of recent procedures which combine both multiplicative aggregation and BoD weighting (e.g. [4]; [15]; [42]; [35]; [36]).

In this paper, a procedure based on the indirect CI-framework is developed as proposed in [35]. In the referred work, the authors propose the construction of a CI as a geometric average using the weights derived from a BoD approach. The procedure implies two successive steps. In the first stage, the importance weights are estimated using a BoD model. In the final step, the individual normalized indicators are aggregated using this weighting vector and a multiplicative scheme.

Consider the context of n entities and m indicators defined above. A max-min normalization process can be applied that enables the comparison between sub-indicators. For each individual value, a normalized value $I_{ri}^{N} \in [1,\ 2]$ is computed as follows:

$$I_{ri}^{N} = \frac{I_{ri} - I_{r}^{min}}{I_{r}^{max} - I_{r}^{min}} + 1 \tag{3.2}$$

where I_{r}^{min} and I_{r}^{max} denote the lowest and the highest value of indicator r across the n entities, respectively.

In order to determine the weights associated to sub-indicator r, our proposal computes a revisited optimistic/pessimistic BoD model developed in [9]. For each unit, the maximum and minimum aggregated values are evaluated where the traditional normalization constraint from DEA is substituted by the condition that the sum of the efficiency values across the set of units adds up to the unity. Note that with the consideration of the optimistic/pessimistic approach, each unit is evaluated not only in its most favourable context, but also its weaknesses are taken into account.

For each unit $o, (o = 1, \dots, n)$, the optimistic CI_o^+ and pessimistic CI_o^- evaluations are computed as follows

$$
\begin{array}{rll}
CI_o^+/CI_o^- = & max/min & \sum_{r=1}^{m} w_{ro}^+ \cdot I_{ro}^N \\
 & s.t. & \sum_{i=1}^{n} \sum_{r=1}^{m} w_{ro}^+ \cdot I_{ri}^N = 1 \\
 & & L_r \leq \frac{w_{or}^+ \cdot I_{ro}^N}{\sum_{r=1}^{m} w_{ro}^+ \cdot I_{ro}^N} \leq U_r, \qquad \forall\, r \\
 & & w_{ro}^+ \geq 0, \qquad \forall\, r, i,
\end{array}
\tag{3.3}
$$

where L_r and U_r are the lower and upper bounds respectively imposed for the determination of the optimal weights. Note that the additional conditions are imposed in terms of the relative weight of the virtual input $w_{ro} I_{ro}^N$. See, for instance, [8] and [38] for a detailed discussion on the inclusion of weight restrictions in BoD models. The outcome from the models (3.3) is a set of optimal weighting vectors for each unit. Following the procedure developed in [36], to mitigate the effect of outliers and/or the existence of errors, both models are robustified using the concepts proposed in [5]. To this end, 2,000 computation rounds of each model are obtained with a sub-sample of randomly selected units.

Note that at this point, an alternative proposal to determine the optimal weights ω_{ri}^+ and ω_{ri}^- has been included. In contrast with the proposal from [36], not only is one unit considered for the normalization condition (the maximum of each sub-indicator), but all the units are considered in the computation of optimal weights. The main benefit from the traditional evaluation of the BoD model (see [35]) is derived from the uniqueness of the optimal weighting profiles. In [19], it is proven that the optimal weighting vectors from models (3.3) are unique.

Following the ideas proposed by [35] and [36], once the importance weights have been determined, the optimistic and pessimistic sub-indicator exponents, denoted respectively by ω_{ri}^+ and ω_{ri}^-, are obtained as the ratio of the virtual output of sub-indicator r over the sum of the s sub-indicators:

$$\omega_{ri}^{+} = \frac{w_{ri}^{+} I_{ri}^{N}}{\sum_{r=1}^{m} w_{ri}^{+} I_{ri}^{N}}; \ \omega_{ri}^{-} = \frac{w_{ri}^{-} I_{ri}^{N}}{\sum_{r=1}^{m} w_{ri}^{-} I_{ri}^{N}}. \tag{3.4}$$

This last step involves the construction of the optimistic and pessimistic geometric indicators. It is important to bear in mind that the original values of the sub-indicators are considered in this stage. The normalized values are only used for the determination of the optimal weights from models (3.3). Here, a benchmark value or baseline sub-indicator value is considered for each r ($r = 1, \ldots, m$). These values are denoted as I_{rB}. In this study, the averages of the observed values have been considered as the baseline. Formally, CI_i^{+} and CI_i^{-} are defined as

$$CI_i^{+} = \prod_{r=1}^{m} \left(\frac{I_{ri}}{I_{rB}} \right)^{\omega_{ri}^{+}}, \ CI_i^{-} = \prod_{r=1}^{m} \left(\frac{I_{ri}}{I_{rB}} \right)^{\omega_{ri}^{-}}. \tag{3.5}$$

Note that in (3.5), the values ω_{ri}^{+} and ω_{ri}^{-} define how much the indicator r contributes to the aggregated values CI_i^{+} and CI_i^{-}. These values represent the percentage change in the *CI*-value as a result of a 1 % increase in $\frac{I_{ri}}{I_{rB}}$.

Once the values CI_i^{+} and CI_i^{-} are computed then the two indicators are added together to determine a compromise indicator such that

$$CI_i = \sqrt{CI_i^{+} \times CI_i^{-}} = \prod_{r=1}^{m} \left(\frac{I_{ri}}{I_{rB}} \right)^{w_{ri}^{*}} \tag{3.6}$$

where $\omega_{ri}^{*} = \frac{\omega_{ri}^{+} + \omega_{ri}^{-}}{2}$.

3.3.2. Inter-temporal decomposition

The geometric composite indicator proposed above permits the inclusion of a dynamic analysis of the entities. See [36] for a detailed analysis of the temporal decomposition of the geometric indicator. A set of observations is considered that corresponds to two periods, denoted respectively by t and t+1. This feature forces the notation of all the relevant variables to be extended accordingly in order to include the temporal reference. We denote by I_{ri}^{t} and I_{ri}^{t+1} , the values of the sub-indicators; by I_{rB}^{t} and, I_{rB}^{t+1} the baseline values; and by $\omega_{ri,\ t}^{*}$ and $\omega_{ri,\ t+1}^{*}$ the relative importance of the sub-indicators, in all the cases, for period t and $t + 1$ respectively.

A measure of the performance change for unit i, denoted by PC_i, can be measured as follows:

$$PC_i = \frac{\prod_{r=1}^{m} \left(\frac{I_{ri}^{t+1}}{I_{rB}^{t+1}} \right)^{\omega^*_{ri,t+1}}}{\prod_{r=1}^{m} \left(\frac{I_{ri}^{t}}{I_{rB}^{t}} \right)^{\omega^*_{ri,t}}}. \tag{3.7}$$

The interpretation of PC_i is straightforward. A value of PC_i greater (less) than the unity indicates that unit i has improved (deteriorated) its performance from t to t+1.

In [36], a tripartite decomposition of PC_i is proposed such that

$$PC_i = \Delta OWN_i \times \Delta BP_i \times \Delta W_i^*. \tag{3.8}$$

The component ΔOWN_i measures the changes derived from the variations in the sub-indicators of unit i. A value greater (less) than the unity represents an improvement (deterioration) in the performance of the individual indicators in period $t+1$ with respect to period t. That is, a value greater (less) than 1 indicates that the valuation of the indicators I_{ri} with the corresponding weighting vectors, is greater (less) in $t+1$ than in t.

With ΔBP_i, the changes derived from the variation of the base-line of t over $t+1$ are measured. Here, a value greater than the unity indicates that the composite value of the baseline on $t+1$ is lower than the corresponding value for period t. Note that, since the indicators I_r are included in relative terms with respect to the baseline $\left(\frac{I_{ri}}{I_{rB}}\right)$, a lower value of I_{rB} supposes an indirect improvement to the evaluation for this individual indicator. The combination of the first component (ΔOWN_i) and component together measures the progress or regress of the set of sub-indicators of unit i with respect to the evolution of the baseline performance indicator.

Finally, a value ΔW_i^* greater (less) than the unity indicates that the weighting scheme has changed in such way that it represents an advantage (disadvantage) for the sub-indicators of unit i. That is to say, a value greater (less) than the unity, indicates a positive (negative) impact derived from the selection of weights in the construction of the composite indicator in period $t+1$ with respect to period t.

3.4. Empirical application

In this section, we apply the methodology described in the previous section to the data of PISA reports for 2012 and 2015 for the OECD countries. Table 3.1 includes the acronyms of the countries that will be used henceforward. In Table 3.2, the main statistics that describe the dataset are summarised. It should to be pointed out that only 33 countries have been included in this study. The absence of some data for Austria forces it to be eliminated.

Table 3.1: Acronyms of countries included in the sample

Country	Acronym	Country	Acronym	Country	Acronym
Australia	AUS	Hungary	HUN	Norway	NOR
Belgium	BEL	Iceland	ISL	Poland	POL
Canada	CAN	Ireland	IRL	Portugal	PRT
Chile	CHL	Israel	ISR	Slovak Republic	SVK
Czech Republic	CZE	Italy	ITA	Slovenia	SVN
Denmark	DNK	Japan	JPN	Spain	ESP
Estonia	EST	Korea	KOR	Sweden	SWE
Finland	FIN	Luxembourg	LUX	Switzerland	CHE
France	FRA	Mexico	MEX	Turkey	TUR
Germany	DEU	Netherlands	NLD	United Kingdom	GBR
Greece	GRC	New Zealand	NZL	United States	USA

Table 3.2: Descriptive statistics of the dataset

2012	I_1	I_2	I_3	I_4	I_5	I_6	I_7
Min	414.92	0.92	36.26	3.00	73.58	0.64	4.26
Max	546.74	32.25	88.23	54.90	93.27	1.43	15.05
Mean	501.00	16.20	71.67	27.27	86.08	1.00	9.63
Median	498.97	15.76	72.05	26.20	85.93	0.96	8.77
Std. Deviation	29.12	6.75	11.37	12.04	5.03	0.22	2.63
2015	I_1	I_2	I_3	I_4	I_5	I_6	I_7
Min	415.71	0.59	35.91	3.50	78.60	0.56	11.02
Max	538.39	25.78	83.12	42.10	95.08	1.47	24.50
Mean	493.24	15.47	69.07	25.38	87.02	1.01	17.10
Median	501.10	15.83	71.65	25.80	87.48	1.00	17.24
Std. Deviation	28.57	6.16	11.16	9.74	4.00	0.19	2.82

In Figure 3.1, the values of the indicators over the corresponding baseline $\left(\frac{I_{ri}}{I_{rB}}\right)$ are represented. Note that in every case, the baseline is represented by the average of the observed values across the countries. Figures 3.1a and 3.1c

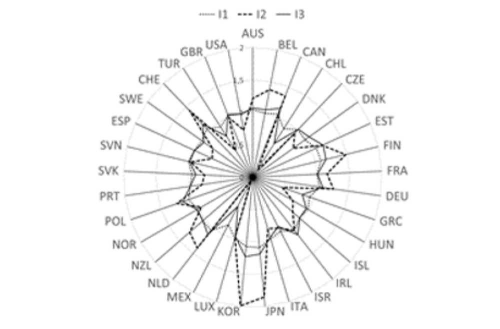

Figure 3.1a. Academic indicators 2012

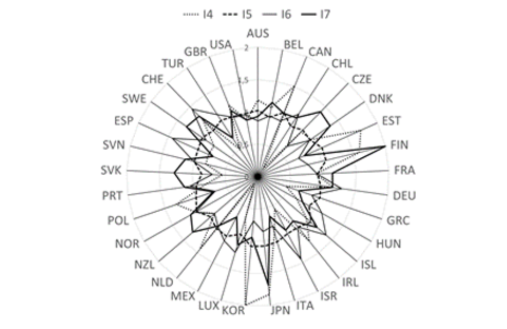

Figure 3.1b. Equity and Student 2012

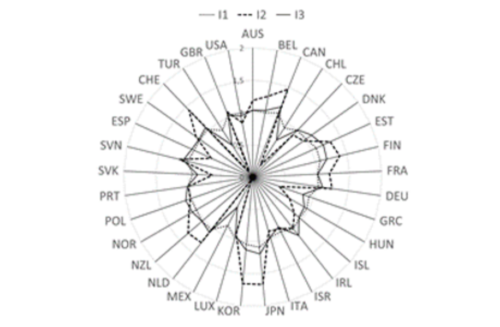

Figure 3.1c. Academic indicators 2015

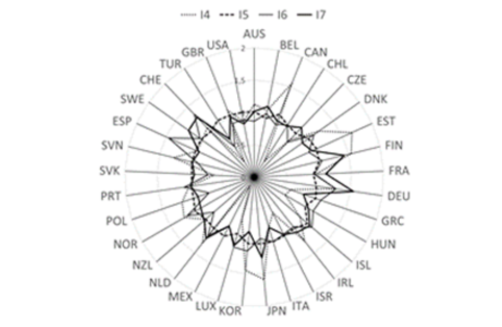

Figure 3.1d. Equity and Student 2015

Figure 3.1: Observations over benchmark values for 2012 and 2015

represent the indicators I_1, I_2, and I_3 (academic dimension) while Figures 3.1b and 3.1d represent the values of I_4, I_5, I_6 and I_7 (equity and student dimension).

The performance of the countries with respect to the indicators is uneven. Countries like Mexico, Chile, Greece, and Turkey are the worst performers in academic features (I_1, I_2, and I_3). At the opposite extreme, Korea, Japan, Finland, and Canada present the best results in all these items. As regards academic excellence (I_2), Switzerland is one of the top performers in both years, but fails to present good results in the remaining concepts of academic dimension. The correlation index between these three indicators is very high for both years (greater than 0.89 in 2012, and greater than 0.93 in 2015).

Regarding the equity dimension (I_4, and I_5), the worst results in social equity correspond to Mexico, Chile, Greece and Turkey; while the worst results in socio-economic incidence correspond to the Slovak Republic, Hungary and France. The best in the equity dimension are Korea, Japan and Estonia, which coincide with those countries with the best performance in academic features. The correlations between I_4 and I_5 is positive and only statistically significant in 2012 (0.52). With respect to students' well-being indicators

(I_6 and I_7), the worst results in the *Sense of belonging* index (I_6) correspond to the Czech Republic, Estonia, the Slovak Republic, and Poland; while the best results thereof correspond to Switzerland, Spain, Germany, and Iceland. The countries that spend the most time studying outside school hours are Italy, Greece, Mexico, and Turkey. The correlation between I_6 and I_7 is positive and statistically significant only in 2015 (0.409)

Table 3.3 includes the values of the multiplicative indicator for 2012 and 2015 and the ranking induced by these values. In this ranking, the first position is assigned to the top-performing country and 33^{rd} position to the worst-rated country. The multiplicative nature of the indicator enables the global value to be split into three separate components in accordance with the main dimensions considered in the study: academic (indicators I_1, I_2, and I_3), equity (indicators I_4 and I_5), and students' well-being (indicators I_6 and I_7). Note that the product of three partial indicators is equal to the global indicator.

Table 3.3: Multiplicative composite indicators and induced rankings

	Global				**Academic**				**Equity**				**Students**			
	2012		2015		2012		2015		2012		2015		2012		2015	
Country	CI_i	Rank	CI_i	Rank	CI_i	Rank	CI_i	Rank	CI_i	Rank	CI_i	Rank	CI_i	Rank	CI_i	Rank
AUS	1.020	14	1.016	17	1.042	10	1.034	11	1.022	11	1.015	14	0.957	24	0.968	24
BEL	1.082	9	1.050	12	1.094	3	1.065	5	0.994	18	0.985	22	0.995	15	1.000	12
CAN	1.139	5	1.170	5	1.053	8	1.073	4	1.100	4	1.129	2	0.983	18	0.965	26
CHL	0.808	31	0.729	32	0.836	32	0.837	32	0.928	31	0.891	33	1.042	6	0.977	21
CZE	0.896	27	0.922	26	0.996	17	0.992	21	0.995	17	0.962	25	0.904	29	0.966	25
DNK	0.949	22	1.040	13	0.956	26	1.013	18	0.996	16	1.029	10	0.996	14	0.998	14
EST	1.102	8	1.195	3	1.041	11	1.056	7	1.183	2	1.154	1	0.895	30	0.982	19
FIN	1.113	7	1.170	4	1.080	4	1.063	6	1.088	6	1.081	4	0.947	27	1.018	7
FRA	0.975	17	0.992	21	1.022	14	1.030	14	0.966	23	0.971	24	0.987	17	0.992	16
DEU	1.119	6	1.094	8	1.041	12	1.037	10	1.012	13	1.019	12	1.062	4	1.034	3
GRC	0.766	32	0.820	30	0.924	30	0.863	30	0.937	30	0.960	26	0.884	31	0.990	18
HUN	0.914	25	0.894	27	0.968	22	0.961	27	0.944	29	0.924	31	1.001	10	1.007	10
ISL	1.050	10	1.001	20	0.964	24	0.967	26	1.005	14	1.011	15	1.084	3	1.024	5
IRL	1.035	12	1.061	11	1.034	13	1.032	12	1.032	9	1.034	9	0.969	22	0.994	15
ISR	0.949	21	0.892	28	0.964	23	0.968	25	0.904	33	0.922	32	1.090	2	0.999	13
ITA	0.853	28	0.923	25	0.976	20	0.981	24	0.999	15	0.984	23	0.874	33	0.956	28
JPN	1.312	1	1.248	1	1.171	1	1.163	1	1.160	3	1.084	3	0.966	23	0.990	17
KOR	1.300	2	1.200	2	1.146	2	1.160	2	1.232	1	1.056	6	0.921	28	0.980	20
LUX	0.982	16	0.949	23	0.991	18	0.985	22	0.944	28	0.938	29	1.050	5	1.028	4
MEX	0.758	33	0.744	31	0.802	33	0.814	33	0.962	25	0.955	28	0.983	19	0.956	29
NLD	1.156	4	1.105	7	1.063	6	1.042	9	1.098	5	1.042	7	0.991	16	1.017	8
NZL	0.999	15	1.034	14	1.045	9	1.082	3	0.976	22	0.998	17	0.980	20	0.958	27
NOR	1.022	13	1.076	9	0.984	19	1.023	16	1.026	10	1.037	8	1.012	8	1.015	9
POL	0.975	18	0.971	22	1.057	7	1.023	15	1.043	7	1.025	11	0.884	32	0.926	31
PRT	0.943	23	1.021	16	0.973	21	1.005	20	0.963	24	0.996	20	1.007	9	1.020	6
SVK	0.834	29	0.820	29	0.953	27	0.941	29	0.916	32	0.958	27	0.955	25	0.910	32
SVN	0.961	20	1.066	10	1.003	16	1.031	13	0.957	26	1.062	5	1.001	11	0.974	23
ESP	0.968	19	1.025	15	0.960	25	0.941	28	0.982	21	0.997	19	1.027	7	1.092	1
SWE	0.919	24	1.011	18	0.936	29	1.011	19	0.985	20	0.998	18	0.997	13	1.003	11
CHE	1.188	3	1.129	6	1.068	5	1.052	8	1.019	12	0.999	16	1.092	1	1.074	2
TUR	0.831	30	0.692	33	0.913	31	0.858	31	0.956	27	0.936	30	0.952	26	0.862	33
GBR	1.045	11	1.008	19	1.005	15	1.017	17	1.042	8	1.017	13	0.998	12	0.975	22
USA	0.912	26	0.923	24	0.944	28	0.983	23	0.988	19	0.992	21	0.979	21	0.947	30
Average	0.996		1.000		1.000		1.003		1.011		1.005		0.984		0.988	

The eight top-performing countries (Japan, Korea, Switzerland, Netherlands, Canada, Germany, Finland, and Estonia) are the same in both evaluations but with variation in their relative positions. Japan and Korea were the top-ranked countries in 2012 and 2015, and Estonia and Finland occupied the third and fourth position in 2015, climbing up from the eighth and seventh position in 2012.

The five lowest-rated countries (Mexico, Greece, Chile, Turkey, and the Slovak Republic) also remain constant from 2012 to 2015, even though the relative position of each country varies from one evaluation to the other. In both years, all of these countries took one of the worst positions in all the composite sub-indicators (academic, equity, and students' indicators), except in Chile in 2012, which held a good position in the students' indicator.

The greatest variations from one period to the other were experienced by Iceland, Slovenia, Denmark, and the United Kingdom. Iceland and Slovenia suffered the greatest change in their ranking by 10 positions, although in a different respect to each other. Iceland worsened in all the sub-indicators and Slovenia improved the first two sub-indicators, with a more pronounced improvement in the equity sub-indicator. Denmark's ranking improved by 9 positions. All the sub-indicators improved, except in terms of the well-being of students, which remained constant. By contrast, the United Kingdom worsened 8 positions in the global composite indicator and 2, 5, and 10 positions in the academy, equity, and students' well-being composite sub-indicators, respectively.

Figure 3.2 summarises the former results and permits the comparison between countries. Each point represents the values of CI for 2012 (abscissa axis) and for 2015 (ordinate axis). That is, the distance from the origin to the right (upwards) indicates a better-valued unit. A continuous line to represent the bisector has been also incorporated. Those units located above this line are those in which the index in 2015 is greater than the value in 2012. The units located below the line present the inverse situation. Note that the larger the difference of angle with respect to 45^o, the greater the differences of the evaluation of both periods.

With respect to the values for Japan, Korea, and Switzerland, in all cases, the values in 2012 are larger than those in 2015 (all located under the bisector). Vertical lines provide a comparison with respect to the values achieved in 2012 (the order induced in this case is Japan, Korea, and Switzerland). Conversely, horizontal lines provide a comparison with respect to the values in 2015. It is easy to observe this induced ranking for Japan, Korea, Estonia, and Finland. On the other hand, Estonia, Finland, and Canada are

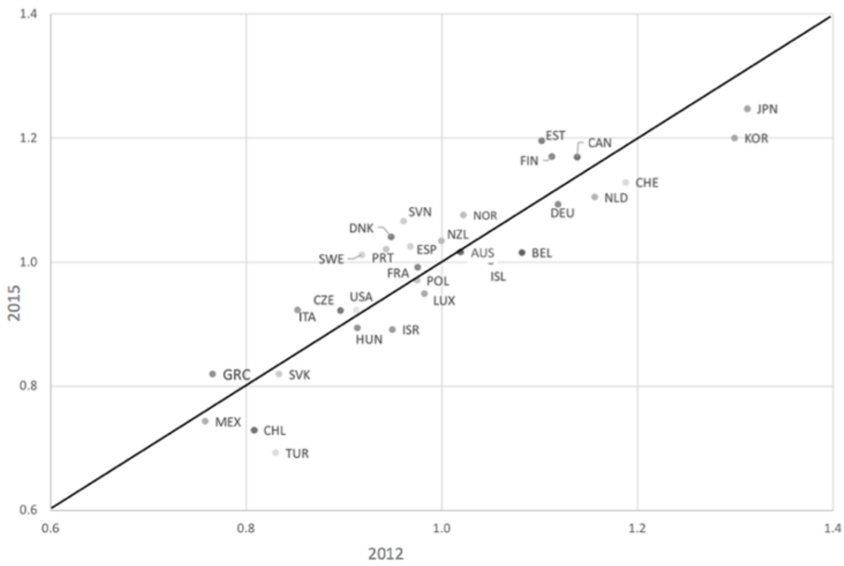

Figure 3.2: Composite indicators in 2012 vs. 2015

located above the bisector. At the opposite end, those points located closest to the origin represent the worst-valued countries. See, for instance, Turkey and Chile (below the bisector, their evaluation deteriorating over time), and Greece (above the bisector, its evaluation improving from 2012 to 2015).

Finally, Table 3.4 summarises the main results for an inter-temporal comparison. Similar to Table 3.3, the results have been included for the global results (indicators from I_1 to I_7) as well as for separated dimensions. In each case, the values of PC_i are splitted into three components as explained in previous sections. Therefore, the information included in Table 3.4 enables the location where this variation in the composite indicator is founded to be detected.

Table 3.4: Decomposition of the inter-period comparison

	Global				Academic				Equity				Students			
Country	PC_i	ΔBP_i	ΔOWN_i	$\Delta Wi*$	PC_i	ΔBP_i	ΔOWN_i	$\Delta Wi*$	PC_i	ΔBP_i	ΔOWN_i	$\Delta Wi*$	PC_i	ΔBP_i	ΔOWN_i	$\Delta Wi*$
AUS	0.996	0.945	1.035	1.019	0.992	1.013	0.980	1.000	0.992	1.007	0.987	0.999	1.011	0.926	1.070	1.021
BEL	0.970	0.955	1.031	0.985	0.974	1.016	0.965	0.993	0.991	1.005	0.989	0.997	1.006	0.936	1.080	0.995
CAN	1.027	0.953	1.071	1.006	1.019	1.010	1.004	1.005	1.026	1.016	0.989	1.021	0.982	0.929	1.078	0.980
CHL	0.902	0.922	1.149	0.852	1.002	1.011	1.033	0.959	0.960	1.003	1.043	0.918	0.938	0.909	1.066	0.967
CZE	1.029	0.894	1.169	0.984	0.996	1.011	0.983	1.003	0.967	1.005	0.965	0.997	1.069	0.880	1.233	0.985
DNK	1.097	0.897	1.338	0.914	1.060	1.014	1.031	1.014	1.033	1.007	1.025	1.001	1.001	0.879	1.266	0.900
EST	1.084	0.957	1.087	1.042	1.014	1.010	1.003	1.001	0.975	1.020	0.968	0.988	1.097	0.930	1.120	1.053
FIN	1.052	0.919	1.247	0.917	0.984	1.011	0.981	0.992	0.994	1.011	0.980	1.003	1.076	0.899	1.297	0.922
FRA	1.017	0.911	1.123	0.994	1.007	1.012	0.994	1.001	1.005	1.004	1.004	0.997	1.004	0.896	1.125	0.996
DEU	0.977	0.935	1.033	1.012	0.996	1.010	0.986	1.000	1.007	1.006	1.006	0.995	0.974	0.919	1.042	1.017
GRC	1.070	0.901	1.131	1.051	0.934	1.010	0.979	0.944	1.024	1.004	1.018	1.002	1.119	0.888	1.134	1.111
HUN	0.978	0.938	0.994	1.050	0.993	1.009	0.980	1.004	0.979	1.003	0.981	0.995	1.006	0.927	1.034	1.050
ISL	0.953	0.959	1.018	0.977	1.003	1.013	0.991	0.999	1.006	1.006	0.992	1.008	0.945	0.942	1.035	0.969
IRL	1.025	0.931	1.033	1.065	0.997	1.015	0.983	1.000	1.002	1.009	0.992	1.001	1.026	0.910	1.060	1.063
ISR	0.939	0.950	1.001	0.987	1.005	1.010	0.994	1.000	1.020	1.011	1.013	0.996	0.916	0.930	0.994	0.991
ITA	1.082	0.885	1.097	1.114	1.005	1.010	0.994	1.001	0.984	1.006	0.979	1.000	1.094	0.871	1.128	1.113
JPN	0.951	0.979	1.013	0.959	0.993	1.016	0.961	1.017	0.934	1.013	0.955	0.964	1.026	0.951	1.103	0.977
KOR	0.924	0.934	1.042	0.948	1.013	1.014	0.938	1.064	0.857	1.015	0.914	0.924	1.064	0.908	1.215	0.965
LUX	0.966	0.941	1.029	0.998	0.994	1.009	0.984	1.001	0.994	1.004	0.990	0.999	0.978	0.928	1.056	0.998
MEX	0.981	0.939	0.996	1.049	1.015	1.012	0.994	1.009	0.993	0.999	1.002	0.991	0.973	0.928	0.999	1.048
NLD	0.956	0.922	1.031	1.005	0.981	1.011	0.976	0.994	0.949	1.013	0.967	0.969	1.026	0.900	1.093	1.044
NZL	1.035	0.931	1.134	0.980	1.035	1.014	0.993	1.028	1.023	1.005	1.015	1.002	0.977	0.913	1.125	0.952
NOR	1.053	0.921	1.196	0.955	1.039	1.012	1.028	0.999	1.010	1.006	1.004	1.000	1.003	0.905	1.159	0.957
POL	0.996	0.949	0.989	1.061	0.968	1.014	0.955	1.000	0.983	1.010	0.975	0.998	1.048	0.927	1.063	1.063
PRT	1.083	0.918	1.199	0.984	1.033	1.010	1.023	1.000	1.034	1.005	1.029	1.001	1.013	0.905	1.139	0.983
SVK	0.983	0.886	1.238	0.896	0.987	1.010	0.970	1.008	1.045	1.003	1.034	1.008	0.953	0.875	1.234	0.882
SVN	1.110	0.937	1.221	0.970	1.028	1.011	1.019	0.998	1.110	1.016	1.092	1.001	0.973	0.913	1.096	0.972
ESP	1.059	0.913	1.083	1.071	0.980	1.012	0.997	0.972	1.015	1.006	1.010	0.999	1.064	0.897	1.076	1.103
SWE	1.101	0.905	1.241	0.981	1.080	1.013	1.057	1.008	1.013	1.010	1.016	0.987	1.007	0.884	1.155	0.985
CHE	0.950	0.958	1.002	0.991	0.985	1.011	0.984	0.990	0.981	1.006	0.975	1.000	0.983	0.942	1.044	1.000
TUR	0.833	0.878	0.868	1.094	0.940	1.009	0.885	1.052	0.979	1.003	0.964	1.013	0.906	0.867	1.018	1.027
GBR	0.965	0.947	1.052	0.969	1.012	1.012	1.002	0.998	0.977	1.011	0.981	0.985	0.977	0.926	1.070	0.986
USA	1.012	0.910	1.123	0.990	1.042	1.013	1.004	1.023	1.004	1.005	0.997	1.001	0.968	0.894	1.121	0.966
Average	1.005	0.928	1.091	0.996	1.003	1.012	0.989	1.002	0.996	1.008	0.996	0.993	1.006	0.910	1.107	1.001

For illustrative purposes, consider the case of Belgium. The global composite indicator of Belgium drops from 1.082 to 1.050.

The value PC_i assesses the value variation in relative terms of CI_i. The value 0.970 indicates a reduction of 3 % in 2015 with respect to the evaluation in 2012. This decrease is derived mainly from worse performances in the observations of the individual indicators: ΔOWN_i equal to 1.031 indicates that the composite evaluation of observed (non-normalised) sub-indicators in 2015 is 3.10 % higher. The weighting factors finally selected have a negative influence: $\Delta W_i^* = 0{,}985$. The value $\Delta BP_i = 0.955$ reflects the situation in which the baseline in 2015 is higher than that of 2012. Note that, since the indicators I_r are included in relative terms with respect to the baseline ($\frac{I_{ri}}{I_{rB}}$), a higher value of I_{rB} supposes an indirect deterioration in the evaluation of this individual indicator.

The most remarkable result is that the academic and equity dimensions experiment a reduction of 2.6 % and 0.9 %, respectively. Although in both the dimensions, ΔBP_i is greater than 1, which has a positive influence on the 2015 data, the negative influence of the other two components pushes the value into negative figures. The students' dimension experienced a raise of 0.6 %, even though the baseline and the weights exert a negative effect (0.936 and 0.995, respectively).

Note that, on average, the results for the 33 countries suppose a slight rise in the global indicator (by up to 0.5 %), in the evaluation of their own observations (9.1 %). This great difference is due to the low evaluation derived from the evolution of the baseline (0.928) and the negative impact of the weighting vectors (0.4 %). The three dimensions are considered separately: in the first two dimensions the evaluation of the indicators supposes a deterioration of the composite indicator (1.1 % and 0.4 % in the academic and equity dimensions, respectively), and the contribution of the third dimension is positive to the global evaluation (10.7 %); the diminution of the baseline in the first two dimensions exerts a positive influence on the final evaluation (1.2 % and 0.8 %, respectively) and a negative influence of 9 % on the students' well-being dimension. Finally, the selection of weights has mixed but limited consequences: in terms of a slight positive influence in the academic and students' well-being dimensions (0.2 % and 0.1 %, respectively) and a negative influence in the equity dimension (0.7 %).

3.5. Concluding remarks

In this paper, the problem of evaluating and analysing national educational systems has been addressed. The complex nature of the situation under study and the requirement for the construction of synthetic reports render composite indicators an appropriate tool for this purpose.

We propose a methodology based on the principles of Data Envelopment Analysis and a multiplicative aggregation. The selection of the weighting factors is carried out objectively, in that each entity can select its own vector of weights. A new procedure is proposed, with a new normalization constraint and considering both best and worst individual evaluations.

The consideration of a multiplicative aggregation scheme guarantees that compensation between the sub-indicators is avoided. Furthermore, a separate analysis of the results of each dimension and an inter-temporal analysis can be performed. Moreover, in the comparison between periods, the main sources of the variation are detected, especially those of the influence in the variation of the performance, the influence of the baseline, and the impact on the selection of the individual weighting vector in the composite indicator. In all cases, the procedure enables progresses and losses to be detected, together with their relative measure.

Three main dimensions have been considered for the construction of the composite indicator: academic, equity and students' self-perceptions. The idea here is to reflect that the objective that an educational system should pursue is not solely limited to the achievement of optimal academic results. A social objective must also be included as a priority aim.

Nevertheless, the procedure provides policy makers with a large amount of information regarding the evolution of the countries. Information is presented on the evolution of the indicator, considered globally and when the evaluation is carried out separately for the three aforementioned dimensions. Furthermore, the sources of the evolution from 2012 to 2015 are also summarised in such a way that the political actors can detect and design appropriate actions so that weaknesses can be corrected, and improvements can be promoted.

This work is of a pragmatic nature, in which information about each country is selected, treated, and synthetized, in an effort to facilitate a first approach to a complex system. The decision-making in a multidimensional context such as this must not be supported exclusively by a single measure or a group of single measures. A wide analysis is necessary, in which composite indicators must be considered as either a first approach to the status quo or the evolution of each alternative.

A methodology such as that proposed could enable the focus to be centred on the determination of which aspect constitutes a challenge for each national educational system, since it permits a first comparative analysis to be carried out between countries, and valuable information to be synthetized for a simpler and quicker analysis. However, the final decision and the policy measures finally applied should not be oriented exclusively towards obtaining a better position in the ranking or towards improving an aggregated value. In other words,, composite indicators such as that proposed here must be viewed as a tool for a first approach towards the analysis of a complex process, and in no way should its improvement be fixed as an objective of an educational policy. Once the status and evaluation of each country have been ascertained, a further analysis of the situation of each system is required in order to determine the measures to be taken.

The PISA reports conclude that expenditure is necessary to guarantee a suitable and equitable performance in education, but in itself it remains insufficient. We have investigated the relationships of the results obtained from the proposed index and two monetary indicators: cumulative expenditure on students up to age 15 and per capita income. None of the countries with the highest GDR per capita is among the top-performing countries when our ranking is considered. Similarly, the indices obtained, considering the global, academic, or equity results, are only weakly related to the cumulative spending on education.

Bibliografía

[1] Agasisti, T., Avvisati, F., Borgonovi, F., and Longobardi, S. (2018) *Academic resilience: What schools and countries do to help disadvantaged students succeed in PISA*. OECD Education Working Papers 167: 1-40. https://dx.doi.org/10.1787/e22490ac-en.

[2] Asif, M., and Searcy, C. (2014) *A composite index for measuring performance in higher education institutions*. International Journal of Quality and Reliability Management 31 (9): 983-1001.

[3] Banker, R.D., Charnes, A., and Cooper, W.W. (1984) *Some models for estimating technical and scale inefficiencies in data envelopment analysis.* Management Science 30 (9): 1078-1092.

[4] Blancas, F.J., Contreras, I., and Ramirez-Hurtado, J.M. (2012) *Constructing a composite indicator with multiplicative aggregation under the objective of ranking alternatives*. Journal of the Operational Research Society 64 (5): 668-678.

[5] Cazals, C., Florens, J.P., and Simar, L. (2002) *Nonparametric frontier estimation: A robust approach.* Journal of Econometrics 106 (1): 1-25.

[6] Charnes, A., Cooper, W.W., and Rhodes, E. (1978) *Measuring the efficiency of decision-making units.* European Journal of Operational Research 2 (6): 429-444.

[7] Cherchye, L., Moesen, W., and Puyenbroeck, T. (2003) *Legitimately Diverse, yet comparable: On synthesising social inclusion performance in the EU.* CES Discussion paper series 03.01, KU Leuven.

[8] Cherchye, L., Moesen, W., Rogge, N., and Puyenbroeck, T. (2007) *An Introduction to 'Benefit of the Doubt' composite indicators.* Social Indicators Research 82 (1): 111-145.

[9] Contreras, I., and Hinojosa, M.A. (2019) *A note on 'The cross-efficiency in the Optimistic-pessimistic framework'.* Operational Research International Journal. Advance online publication. doi: 10.1007/s12351-019-00484-2.

[10] Downey, D.B., and Condron, D.J. (2016) *Fifty years since the Coleman report: Rethinking the relationship between schools and inequality.* Sociology of Education 89 (3): 207-220. http://dx.doi.org/10.1177/0038040716651676.

[11] Ebert, U., and Welsch, H. (2004) *Meaningful environmental indices: a social choice approach.* Journal of Environmental Economic Management 47 (2): 270-283.

[12] Editor, L. H. (2008) *Rankings of Higher Education Institutions: A Critical Review.* Quality in Higher Education 14 (3): 187-207.

[13] Esty, D.C., Levy. M., Srebotnjak, T., and Sherbinin, A. (2005) *Environmental sustainability index: Benchmarking national environmental stewardship.* Yale Center of Environmental Law and Policy, New Haven.

[14] Fernandez-Cano, A. (2016) *A Methodological Critique of the PISA Evaluations.* e-Journal of Educational Research, Assessment and Evaluation 22 (1), art. M15. doi: http://dx.doi.org/10.7203/relieve.22.1.8806.

[15] Giambona, F., and Vassallo, E. (2014) *Composite indicator of social inclusion for European countries.* Social Indicators Research 116 (1): 269-293.

[16] Goodenow, C. and Grady, K. (1993) *The relationship of school belonging and friends' values to academic motivation among urban adolescent students.* The Journal of Experimental Education 62 (1): 60-71.

[17] Hopfenbeck, T.N., Lenkeit, J. El Masri, Y., Cantrell,K., Ryan, J., and Baird, J.A. (2018) *Lessons Learned from PISA: A Systematic Review of Peer-Reviewed Articles on the Programme for International Student*

Assessment, Scandinavian. Journal of Educational Research 62 (3): 333-353.

[18] Hopmann, S., Brinek, G., and Retzl, M. (2009) *PISA according to PISA- Does PISA keep what is promised.* University of Vienna Press. Vienna.

[19] Khodabakhshi, M., and Aryavash, K. (2012) *Ranking all units in Data Envelopment Analysis.* Applied Mathematics Letters 25 (12): 2066-2070.

[20] Lin CH.A. (2007) *Education expansion, educational inequality, and income inequality: Evidence from Taiwan, 1976-2003.* Social Indicators Research 80 (3): 601-615.

[21] Marginson, S. (2011) *The new world order in higher education.* In *Questioning Excellence in Higher Education: Policies, Experiences and Challenges in National and Comparative Perspective,* edited by Michele Rostan and Massimiliano Vaira, 320. doi: 10.1007/978-94-6091-642-7.

[22] Murias, P., de Miguel, C., and Rodriguez, D. (2007) *A Composite Indicator for University quality assessment: The case of Spanish higher education system.* Social Indicators Research 89: 129-146.

[23] Nardo, M., Saisana, M., Saltelli, A., Tarantola, S., Hoffman, A., and Giovannini, E. (2005a) *Handbook on constructing composite indicators: Methodology and user guide.* OECD Statistics Working Papers.

[24] Nardo, M., Saisana, M., Saltelli, A., and Tarantola, S. (2005b) *Tools for composite-indicator building.* Institute for the Protection and Security of the Citizen: European Commission.

[25] OECD (2008) *Handbook on constructing composite indicators. Methodology and user guide.* Paris: OECD Publishing.

[26] OECD (2009) *PISA Data. Analysis Manual SPSS Second edition.* Paris: OECD Publishing.

[27] OECD (2012) *PISA 2009 technical report.* PISA. OECD Publishing. http://dx.doi.org/10.1787/9789264167872-en.

[28] OECD (2017) *PISA 2015 Assessment and analytical framework: Science, reading, mathematics, financial literacy and collaborative problem solving.* Paris: OECD Publishing. http://dx.doi.org/10.1787/9789264281820-en

[29] Roemer J.E., and Trannoy A. (2015) *Equality of opportunity.* In *Handbook of income distribution* 2: 217-300. http://dx.doi.org/10.1016/B978-0-444-59428-0.00005-9.

[30] Saccome, D. (2008) *Educational inequality and educational poverty. The Chinese case in the period 1975-2004.* Working Paper 8/2008, Dipartimento di Economia, S. Cognetti de Martiis. Torino, Italy.

[31] Saisana, M., and Tarantola, S. (2002) *State-of-the-Art report on current methodologies and practices for composite-indicator development.* European Commission: Joint Research Centre.

[32] Sjøberg, S. (2015) *PISA and Global Educational Governance - A Critique of the Project, its Uses and Implications.* Eurasia Journal of Mathematics, Science and Technology Education 11 (1): 111-127.

[33] Stumbriene, D.E., Camanho, A.S., and Jakaitiene, A. (2020) *The performance of education systems in the light of Europe 2020 strategy.* Annals of Operations Research 288: 577-608. https://doi.org/10.1007/s10479-019-03329-5.

[34] Unesco. (2015) *Incheon Declaration. Education 2030: Towards inclusive and equitable quality education and lifelong learning for all.* http://unesdoc.unesco.org/images/0023/002338/233813M.pdf

[35] Van Puyenbroeck, T., and Rogge, N. (2017) *Geometric mean quantity index numbers with Benefit-of-the Doubt weights.* European Journal of the Operational Research 256 (3): 1004-1014.

[36] Verbunt, P., and Rogge, N. (2018) *Geometric composite indicators with compromise Benefit-of-the Doubt weights.* European Journal of the Operational Research 264 (1): 388-401.

[37] Villar, A. (2007) *Performance, inclusion and excellence: An index of educational achievements for PISA.* Advances in Social Sciences Research Journal 4 (3), 110-115.

[38] Wong, Y.H., and Beasley, J.E. (1990) *Restricting weights flexibility in data envelopment analysis.* Journal of the Operational Research Society 41 (9): 829-835.

[39] Yoon, K.P., and Hwang, C.L. (1995) *Multiple-attribute decision-making: An introduction.* Thousand Oaks, StateCA: Sage Publications.

[40] Zhao, Y. (2020) *Two decades of havoc: A synthesis of criticism against PISA.* Journal of Educational Change 21: 245-266.

[41] Zhou, P., Ang, B.W., and Poh, K.L. (2006) *Comparing aggregation methods for constructing the composite environmental index: An objective measure.* Ecological Economics 59 (3): 305-311.

[42] Zhou, P., Ang, B.W., and Zhou, D.Q. 2010. *Weighting and aggregation in composite indicator construction: A multiplicative optimization approach.* Social Indicators Research 96 (1): 169-181.

Conclusiones

En la presente memoria se han presentado tres trabajos de investigación en los que se proponen diferentes aproximaciones al estudio de la Economía de la Educación. La idea básica de los trabajos es, de manera muy sintética, estudiar la utilidad de los métodos cuantitativos a un campo complicado y difícil de abordar como es el estudio de los sistemas educativos. El objetivo de este estudio es, proveer a los decisores sobre de política educativa de algunos instrumentos adicional de carácter objetivo que puedan servir de ayuda en el proceso de toma de decisiones.

Dos elementos comunes relacionan los tres artículos incluidos. Por un lado, en todos ellos hemos optado por utilizar modelos no paramétricos. Esto es, procedimiento en los que no se requiere una determinación a priori de la forma funcional que represente el proceso productivo que se pretende estudiar. Las particularidades de la educación, entendida como proceso productivo, hacen que esta aproximación nos parezca más adecuada. En segundo lugar, en todos ellos hemos utilizado como base modelos basado en el Análisis Envolvente de Datos (DEA).

Se ha utilizado la metodología DEA en diferentes aproximaciones. En los dos primeros trabajos se utiliza en una visión más tradicional, como metodología con la que puede medirse la eficiencia en el desempeño de un conjunto de unidades o alternativas que, a partir de múltiples recursos, producen múltiples productos con una tecnología común. Si bien se utiliza esta aproximación más clásica, en ambos se proponen modificaciones o mejoras que adaptan el modelo al caso particular que se desea estudiar.

En el primer artículo, se propone analizar el sistema al nivel de las instituciones educativas. Proponemos una metodología en la que complementamos los modelos DEA con técnicas estadísticas multivariantes para extraer la información que puede generarse. Como se vió, concluimos que la ubicación geográfica de los centros educativos dentro de España no es significativa en términos de explicación de la varianza en la eficiencia relativa de las instituciones. Esto podría sugerir que las políticas tomadas desde el Gobierno central van a ser las verdaderas determinantes de los resultados y/o que el margen de maniobra de las Comunidades Autónomas es relativamente bajo, al menos en términos de resultados. También indica que existe una igualdad de oportunidades a lo largo del territorio español, es decir, que no hay un hándicap para los alumnos en función de su lugar de residencia. Si llegamos a la conclusión que tanto el tipo de gestión como el nivel socioeconómico son determinantes en el valor de eficiencia que alcanzan los colegios. El mejor desempeño de instituciones privadas y concertadas, frente a colegios públicos, puede indicarnos que algunas de las prácticas de gestión llevadas a cabo en estas instituciones pueden ser trasladadas a la gestión de centros públicos.

En el segundo artículo, se propone una modificación a un modelo existente para adaptarlo a la forma en de la información que tratamos. En este segundo trabajo, cambiamos a un enfoque estatal para evaluar los sistemas educativos nacionales en países OCDE. Para ello, analizamos los niveles de desempeño publicados en PISA, que son categorías en las que el estudio clasifica a los estudiantes en función de sus resultados. Cuanto mayor sea el porcentaje de alumnos en los niveles superiores, más cercano estará ese sistema de la excelencia. La forma en la que se dispone esta información, nos obliga a adaptar los modelos tradicionales para que tanto los valores observados como los valores de referencia respondan a sus características.

Como era de esperar, los resultados muestran que España no es eficiente. Es decir, los resultados en PISA no sólo no son buenos sino que, con el volumen de recursos empleados, debería alcanzar resultados superiores. Esto implica que incrementar el gasto en educación no puede ser la única solución al problema de España. Es necesario revisar la *tecnología* que subyace al proceso, entendida como la forma en que se combinan los recursos para la obtención del output.

Centrarnos para mejorar en aquellos países que obtienen resultados excelentes no es la mejor estrategia. Debemos mirar a aquellos sistemas educativos que el DEA nos ofrece como similares a nosotros. Tarea para futuros investigadores será determinar qué prácticas llevan a cabo Irlanda, Suiza, Corea, Dinamarca o Estonia que les permiten ser eficientes con una disposición similar de recursos.

En el tercer trabajo, los modelos DEA se utilizan como herramienta auxiliar, como elemento complementario al procedimiento de construcción de indicadores compuestos. En este tercer trabajo, se propone un nuevo panel de indicadores, que comprenda conceptos más allá de los meros resultados académicos para evaluar la bondad de un sistema educativo, y una nueva metodología de agregación en la que se intenta maximizar la objetividad en la determinación de las ponderaciones de cada indicador y evitar la compensación entre las dimensiones en las que se agrupan los indicadores.

En el tercero de los artículos se desarrolla una herramienta para evaluar los sistemas educativos, en el que se tienen en cuenta no sólo aspectos académicos sino también socio-económicos y de bienestar de los estudiantes. El objetivo de los sistemas educativos no debe ser sólo dotar a sus alumnos de unos conocimientos teóricos. Además, debe perseguir como objetivo convertirlos en ciudadanos funcionales, sin que ello suponga un sacrificio no deseable en tiempo y bienestar. Aspectos como la equidad, la resiliencia y el bienestar de los estudiantes son fundamentales para la construcción de una sociedad más justa y que permita un mayor desarrollo económico y social.

En el caso de España, si bien se consigue mejorar en el indicador desarrollado en el artículo entre los años 2012 y 2015, este resultado se alcanza gracias a mejoras en las valoraciones de las dimensiones de equidad y bienestar de los estudiantes. Los actores políticos deberían tener en cuenta estos datos para, manteniendo las políticas relativas a estos aspectos, buscar formas de mejorar el rendimiento académico. Esto es coherente con los que hemos visto en el resto de la tesis, España necesita reformar su sistema educativo con el objetivo de obtener un mejor nivel académico.

En las conclusiones de este tercer trabajo se resumen, de alguna forma, las conclusiones generales que pueden extraerse de esta memoria. Se pone de manifiesto la importancia y limitaciones de los análisis cuantitativos aplicados a cualquier campo de estudio y, por ende, al campo de la educación. Los resultados obtenidos del análisis cuantitativo deben verse como una herramienta de trabajo para los decisores, como una forma de presentar, tratar la información relativa a un fenómeno complejo. Esto es, una forma de analizar e interpretar problemas. Pero en ningún caso debe verse como un fin en sí mismo, como ha ocurrido con algunos indicadores que miden desempeños educativos de países o instituciones. Las mejoras en los resultados deben ser la consecuencia de mejoras en el desempeño y no un objetivo en sí mismo.

Los responsables de la toma de decisiones deben apoyarse, sin duda, en la información que proporcionan este tipo de análisis, que deben hacerse desde la neutralidad y sin carga ideológica. Las decisiones finales deben ser resultado de un análisis más profundo. Los resultados de los modelos deben ser una parte de un proceso complejo, en los que debe incluirse también el análisis del detalle de cada situación y análisis cualitativo.

En cuanto a las líneas futuras de investigación, es claro que el campo que queda por delante es amplísimo. La cantidad y calidad de la información con la que contamos cada año es más y mejor. A la información que proporciona el informa PISA se une cada año nuevos proyectos, tanto nacionales como internacionales, así como mejoras propuestas por algunos autores sobre la ya existente. Esta ingente cantidad de información permitirá ampliar las propuestas de análisis aquí presentadas, incluyendo análisis regionales o la incidencia de el género en los resultados. Y permitirá, asimismo, nuevas propuestas no factibles hasta el momento, debido a la no disponibilidad de la información necesaria.