

9^o Seminario
Internacional
de Investigación
sobre la calidad de la educación

DOCENTES

BOGOTÁ D.C.



GOBIERNO
DE COLOMBIA



MINEDUCACIÓN

icfes
mejor saber

¿Cómo evaluar a los docentes de manera ‘justa’?

Kristof De Witte

Universidad de Leuven, Bélgica

Universidad de Maastricht, Holanda

Noviembre 2, 2018

* La traducción de esta presentación fue realizada por el equipo de la Oficina de Investigación del Icfes.

Bélgica versus Colombia

→ ¡Compartimos el gusto por las buenas bebidas!



Bélgica versus Colombia

→ ¡Compartimos el haber perdido contra Inglaterra en la copa del mundo 2018!



Recordemos de ayer

→ Atraer y recompensar a los buenos profesores es importante

“Los incentivos monetarios son efectivos” (Paul Glewwe)

“pero no para todos (e.g. doble salario in Indonesia)” (Karhik Mualidharan)

→ Aún: ¿quiénes son los mejores docentes?

“No hay un examen perfecto para evaluar a los docentes” (Maria Paulina)

Recordemos de ayer

→ Principios de la evaluación docente en Colombia (Laura Barragan)

Multidimensional

Reflexión

Autonomía

Transparencia

Depende del tamaño
de la clase

Recordemos ade ayer

→ Principios de la evaluación docente en Colombia (Laura Barragan)

Multidimensional

Reflexión

Autonomía

Transparencia

Depende del tamaño
de la clase

Esta presentación:
Desarrolla una técnica para evaluar
la calidad de los docentes
empleando evaluaciones de los
docents realizadas por los
estudiantes.

La técnica cumple con los principios
colombianos

Las evaluaciones que hacen los estudiantes sobre la enseñanza (SET por sus siglas en inglés) se utilizan cada vez más para medir el desempeño de los docentes

→ e.g. Portugal, Flandes, E.E.U.U., etc.

Sin embargo, estas siguen siendo controversiales

i.e., son 'injustas' en la medida en que no controlan por el impacto de factores que se encuentran fuera del control de los maestros.

- ↗ - Investigación académica muestra que las características de contexto tienen un efecto
- ↘ - La experiencia práctica de los docentes indica que algunos entornos son más constructivos para enseñanza de alta calidad.

“Cualquier sistema de evaluación de facultades necesita preocuparse por la justicia, lo que frecuentemente se traduce en preocupación sobre la comparabilidad. Usando el mismo sistema de evaluación para todos casi que garantiza que la evaluación va a ser injusta para todos”

(Emery et al., 2003, p. 44)

¿Cómo construir evaluaciones de enseñanza de los estudiantes (SET) de una manera justa?

Construcción típica de puntajes de las SET:

- Paso 1: Calcular las puntuaciones SET mediante la media aritmética de los ítems del cuestionario (de esta manera, sin tener en cuenta el entorno exógeno).
- Paso 2: Determinar el impacto de las características del entorno en los puntajes SET (usualmente mediante análisis de correlación, regresión, modelos multinivel).
- Paso 3: Ajustar los puntajes SET con las características del entorno.

Problemas con el método tradicional de medir SET:

1. Cálculo de los puntaje de la SET en el primer paso:

Implica frecuentemente que todos los aspectos del aprendizaje se valoran igual

↔ Los profesores valoran los aspectos de manera diferente

↔ No hay consenso de cómo se interrelacionan los aspectos del aprendizaje

↔ Usar pesos fijos es subjetivo

↔ Crea injusticia (y con esto docentes desilusionados)

2. Supuesto de separabilidad en los pasos 2 y 3

Se asume que no hay un vínculo directo entre los puntajes de la SET y el entorno de aprendizaje

¿Cómo valoramos las dimensiones subyacentes?

1. Cualquier conjunto común de pesos predeterminados favorecerá a ciertos docentes y perjudicará otros -> Injusticia
2. En la ausencia de un consenso sobre cómo se interrelacionan los aspectos del aprendizaje, cualquier decisión de pesos será, en alguna medida, **subjetiva**.
3. La selección de pesos puede afectar el puntaje que reciben los docentes y sus clasificaciones **socavando** su **credibilidad**.



- *“No hay un **modelo** para ser un docente efectivo”* (Fraser, 2000 p. 3).
- *“Sabemos cuáles son las características de un buen docente, pero **no sabemos como se relacionan entre ellas**”* Weimer (1990, p. 13)

Idea:

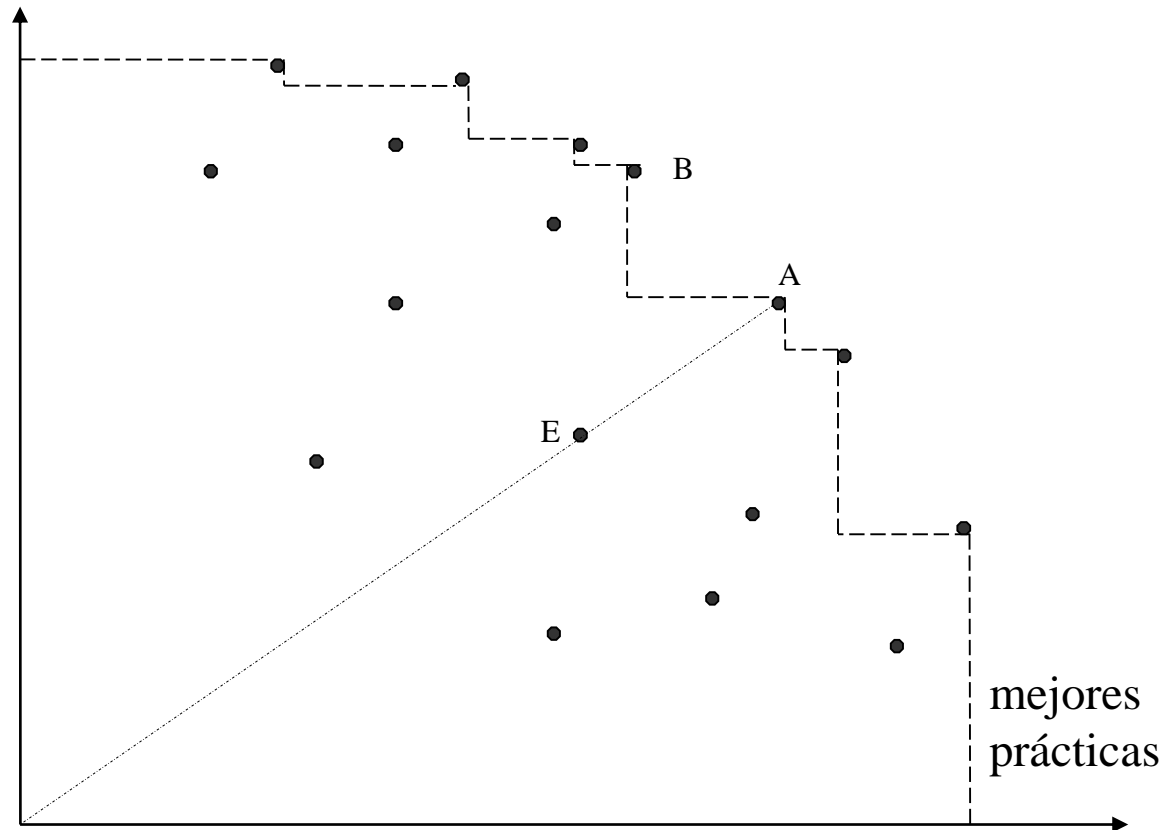
Empezar con los docentes de mejor rendimiento

y comparar el desempeño de los demás con estos mejores docentes

→ Modelo de Beneficio de la Duda (BoD)

e.g. El modelo analiza gráficamente dos dimensiones:

Resultado y_2 :
Clases están bien estructuradas



Resultado y_1 : Docente explica de manera clara

”Modelo de Beneficio de la Duda” (BoD)

Esta aproximación es conveniente porque la expresión algebraica detrás de esta gráfica determina los pesos endógena

i.e. La razón del desempeño de los docentes evaluados

al desempeño de los mejores docentes



$$SET_c = \max_{w_{c,i}} \frac{\sum_{i=1}^q w_{c,i} y_{c,i}}{\max_{y_{j,i} \in \{\text{evaluated lectures}\}} \sum_{i=1}^q w_{c,i} y_{j,i}}$$

→ ¿Dónde estamos ahora?

- Construcción de los puntajes SET basados en indicadores de desempeño de una única dimensión
- No tenemos una comprensión *a priori* de la importancia de estos indicadores



El modelo:

- Para cada ítem i del cuestionario, ajustar el desempeño del docente en su curso c (i.e., $y_{c,i}$) para que sea relativo con respecto al desempeño de los demás $y_{j,i}$

→ Un buen desempeño relativo: mayor valoración a este ítem

→ Un mal desempeño relativo: menor valoración a este ítem

Así: valoraciones/pesos óptimos que maximizan los puntajes SET

Desventajas del BoD:

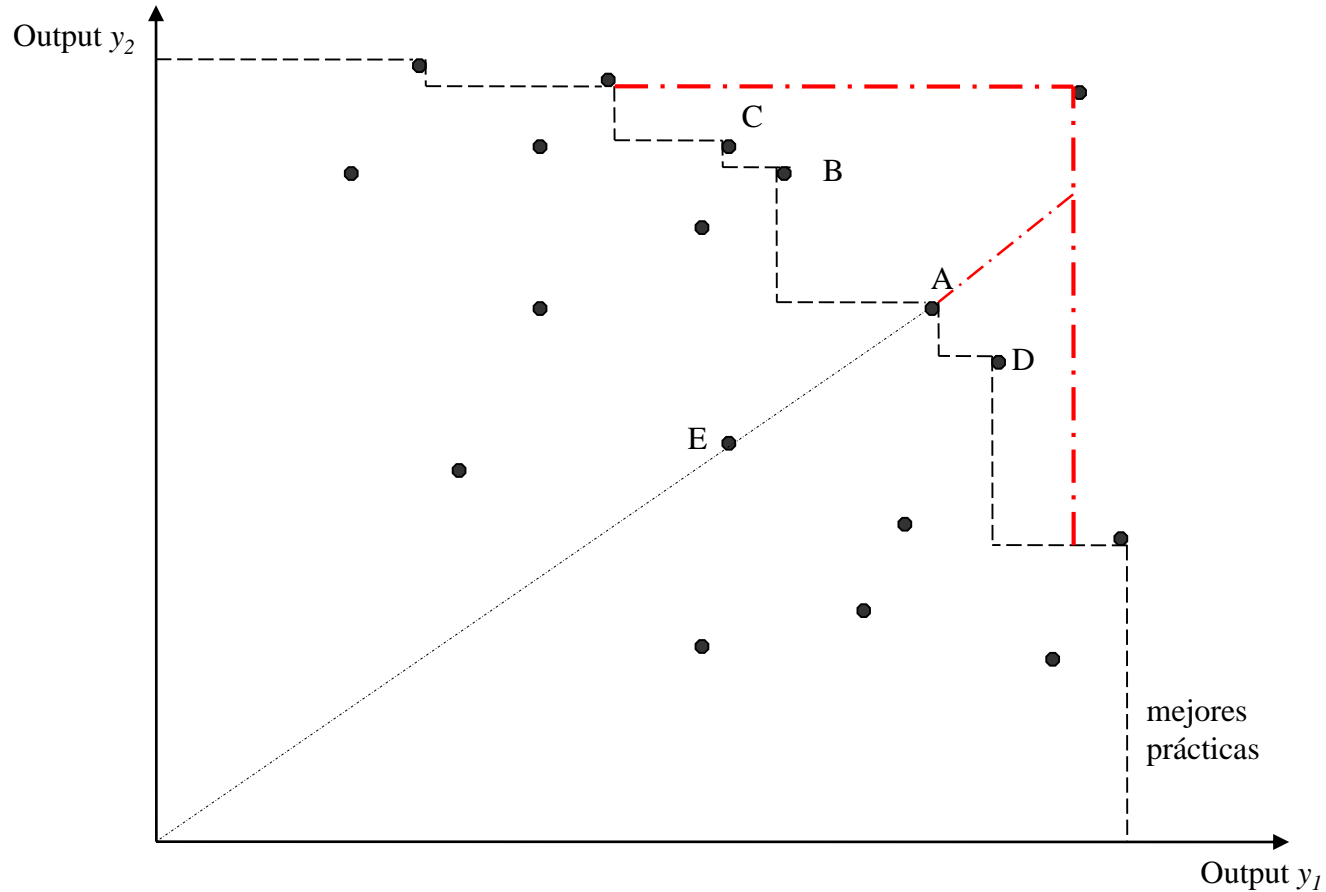
- ✓ Le puede permitir a los docentes aparentas ser excelentes profesores de una manera que es difícil de justificar (e.g. **pesos de cero o pesos que contradicen perspectivas previas.**

Solución:

Tener en cuenta la **opinión de los expertos/partes interesadas (e.g. estudiantes, maestros, etc.)**, reconociendo que el consenso en un esquema de valoración único y fijo es la excepción antes que la regla.

→ Modelo de Beneficio de la Duda (BoD)

e.g. el modelo analiza gráficamente dos dimensiones:



→ El modelo de Beneficio de la Duda robusto (BoD robusto)

→ Motivos:

1. Permite tener en cuenta observaciones extremas (e.g., error de medición)
2. Inferencia estadística

→ Implementación

Puntajes robustos de eficiencia de Cazals *et al.*, 2002

→ Idea:

- Tomar, de manera repetida y con reemplazo, m observaciones de la muestra original de n observaciones.
- Estimar relativo a este conjunto de referencia de tamaño m el modelo BoD
- Tomar la media aritmética del los B puntajes SET:

Teacher-related characteristics

	Significant correlation	Insignificant correlation
Instructor gender	<i>Higher SETs for females:</i> Kaschak (1981); <i>Higher SETs for males:</i> Feldman (1992); <i>Gender interaction:</i> Basow et al. (1987), and Basow (2000)	Basow et al. (1985), McKeachie (1979), Cashin (1995), Fernandez et al. (1997), Hancock et al. (1992), Marsh et al. (1997), Ellis et al. (2003), and Liaw et al. (2003)
Teacher age and experience	<i>Positive:</i> McPherson (2006), Smith et al. (1992), d'Appollonia et al. (1997), Wagenaar (1995); <i>Negative:</i> Baek et al. (2008), and Cochran et al. (2003); <i>Nonlinear relationship:</i> Langbein (1994)	Feldman (1983), Liaw et al. (2003), Ellis et al. (2003), and Koh et al. (1997)
Pedagogical training	<i>Positive:</i> Wagenaar (1995), Nasser et al. (2006),	
Teacher Rank (guest/part-time vs. full-time)	<i>Full-time teachers with lower SETs:</i> Aigner et al. (1986)	Cranton et al. (1986), Delaney (1976), Chang (2000), Steiner et al. (2006), and Willet (1980)
Doctoral degree	<i>Negative:</i> Cochran et al. (2003), Nasser et al. (2006)	Chang (2000)

Student-related characteristics

	Significant correlation	Insignificant correlation
Student grades	<i>Positive:</i> Greenwald et al. (1997), Langbein (1994), Baek et al. (2008), McPherson (2006), Isely et al. (2005), Marsh et al. (1997, 2000), Griffin (2001, 2004), Feldman (1997), Marsh (1980, 1983, 1984, 1987), etc.	Decanio (1986), Abrami et al. (1980)
Student heterogeneity	<i>Negative:</i> Dreeben et al. (1988), Ting (2000), and Perry (1997)	
Questionnaire response rates	<i>Positive:</i> Koh et al. (1997) <i>Negative:</i> McPherson (2006)	Isely et al. (2005)

→ El modelo de Beneficio de la Duda robusto y condicional

→ Motivos:

1. Incorporar las características del entorno en el modelo de BoD
2. Comparar 'manzanas con manzanas'
3. Relajar el supuesto de reparabilidad
4. Inferencia estadística del impacto de las características

→ Implementación

Estimadores de eficiencia condicional para variables exógenas mixtas (i.e., continuas y discretas) de De Witte and Kortelainen (2008)

→ Idea:

- Tomar, de manera repetida y con reemplazo, m observaciones de la muestra original de n observaciones empleando una probabilidad que $z_{c,r} \approx Z$

→ Configuración del cuestionario:

Se aplicó un cuestionario de 16 preguntas a 5,513 estudiantes

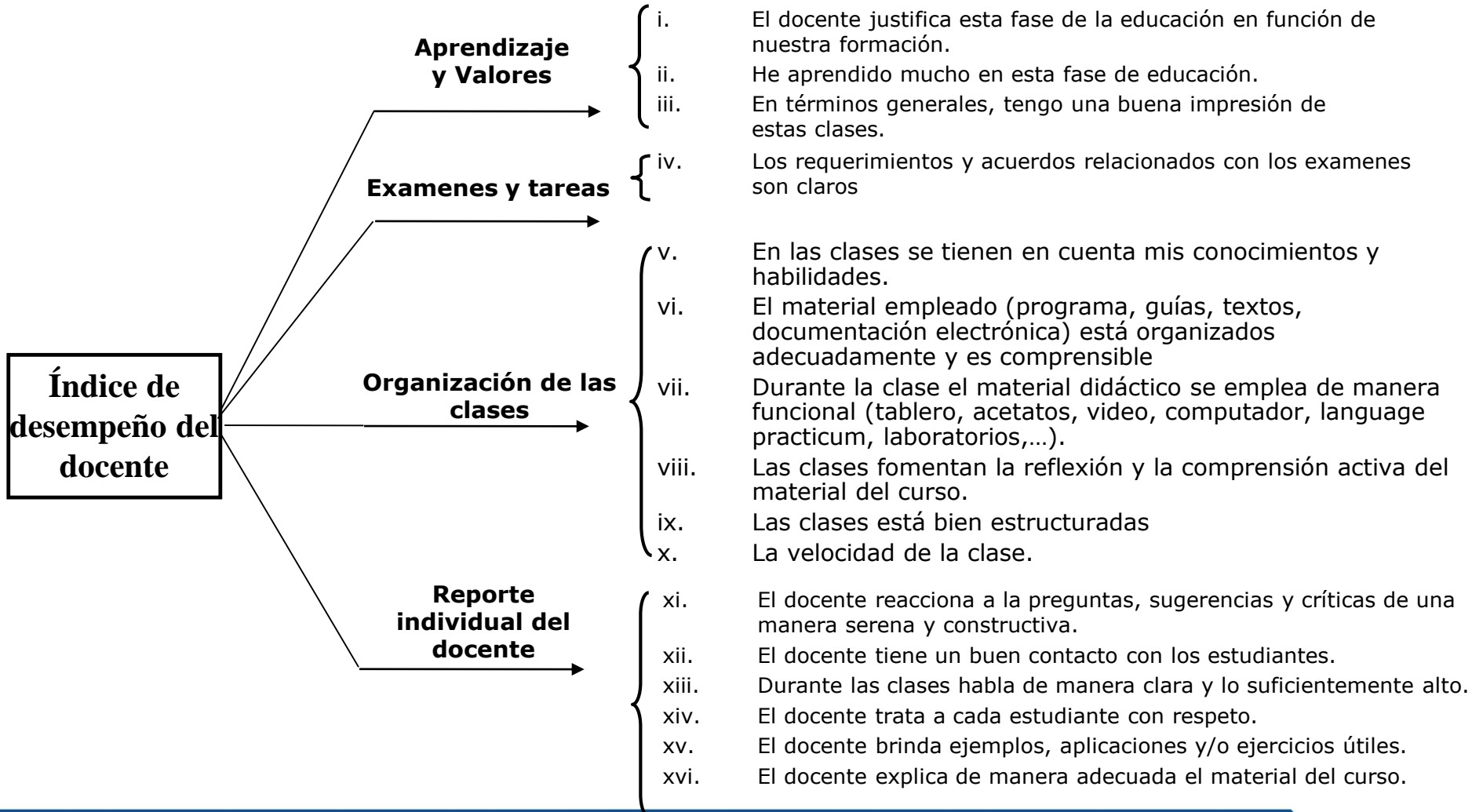
- 112 cursos universitarios de 69 docentes
- Ciencias comerciales en la Universidad de Bruselas (Bélgica)
- Año: 2006-2007



→ Dimensiones del cuestionario:

- Las preguntas eran calificadas con una Escala de Likert de 1 (en desacuerdo) a 5 (de acuerdo).
- Las preguntas fueron agrupadas por la coordinación de la universidad en 4 dimensiones:
 1. Aprendizaje y Valores
 2. Exámenes y tareas
 3. Organización de las clases
 4. Reporte individual del docente
- Se relacionan con las características del entorno

4 DIMENSIONES CLAVE



16 PREGUNTAS DEL CUESITONARIO

→ Resultados

Nr.	Teacher	Course	Class	Contact	EW	BoD	BoD_R	Order-m	BoD_R
...
8673	Professor B	Micro Economics A	1BW ¹	45	3.650	85.50%	79.25%		82.84%
8674	Professor B	Micro Economics B	1BW ²	30	3.697	86.19%	80.10%		83.92%
9487	Professor B	Micro Economics B	1DW ²	30	4.101	94.81%	88.14%		92.36%
66607	Professor C	Banks & Stock B	2JU¹	16	3.582	83.31%	83.05%		86.28%
1421	Professor C	Corporate finance	1EW²	30	3.981	94.31%	73.58%		76.81%
8522	Professor C	Banks & Stock A	1BE¹	30	3.677	85.09%	75.72%		78.26%
8636	Professor C	Banks & Stock A	1BW¹	30	3.750	89.77%	78.02%		81.08%
8911	Professor C	Corporate finance	1EW¹	30	3.801	91.79%	78.84%		82.29%
9029	Professor C	Banks & Stock B	1LC¹	16	3.250	77.16%	65.99%		68.95%
9157	Professor C	Banks & Stock B	1SB¹	16	2.944	76.61%	64.14%		66.96%
8927	Professor D	Quantitative Methods	1EW ¹	30	3.508	87.60%	75.51%		74.46%
9583	Professor D	Quantitative Methods	2LB ²	30	3.400	83.60%	75.22%		78.38%
...

¹: academic year 2005/2006, ²: academic year 2006/2007, EW = Equal Weighting, BoD = full flexibility Benefit of the Doubt weighting, BoD_R = Restricted Benefit of the Doubt weighting, and Order-m BoD_R = restricted and robust order-m Benefit of the Doubt weighting

→ Modelo de Beneficio de la Duda (BoD) condicional e incondicional

	Dimension 1 Learning and value	Dimension 2 Examinations and Assignments	Dimension 3 Lecture organization	Dimension 4 Individual Lecturer report	Aggregate BoD
Unconditional BoD model					
Average	0.79443	0.76371	0.82782	0.83868	0.83328
St. Dev.	0.11985	0.12301	0.09214	0.08122	0.09653
Min.	0.33605	0.35065	0.49471	0.54069	0.52400
Max.	1.00000	1.00000	1.00000	1.00000	1.00000
Conditional BoD model 1					
Average	0.80968	0.78222	0.85217	0.85474	0.86116
St. Dev.	0.12166	0.12507	0.09563	0.10437	0.09797
Min.	0.37430	0.35961	0.51006	0.49847	0.53853
Max.	1.01817	1.00904	1.02788	1.00949	1.01823

¿Qué se correlaciona con los puntajes SET?

Influencia favorable

- Entrenamiento en pedagogía
- Tamaño de clase

Influencia no favorable

- Profesor invitado
- Puntaje promedio de los estudiantes
- Cursos en la tarde

Sin influencia significativa

- Edad
- Dispersión de los puntajes de los estudiantes

→ Posibles aplicaciones en educación:

- Evaluar la enseñanza de los profesores de universidad
 - en la universidad HUB (Bélgica)
- Evaluar la investigación de los profesores de universidad

- Evaluación de los docentes de secundaria:

- cf. Portugal; ver OECD, 2009

”El modelo de evaluación de los docentes involucra el uso de una amplia gama de instrumentos, que incluyen autoevaluaciones, observaciones de clase, entrevistas, resultados de los estudiantes y formularios estandarizados de desempeño de los docentes – este es un modelo ambicioso, ya que busca tomar todas las áreas del funcionamiento de los docentes”

- Proyecto piloto en Flandes (Klasse, 2001)
- Amplia literatura en E.E.U.U.: la evaluación como una herramienta para el mejoramiento instruccional → surge del 'No Child Left Behind' Act.

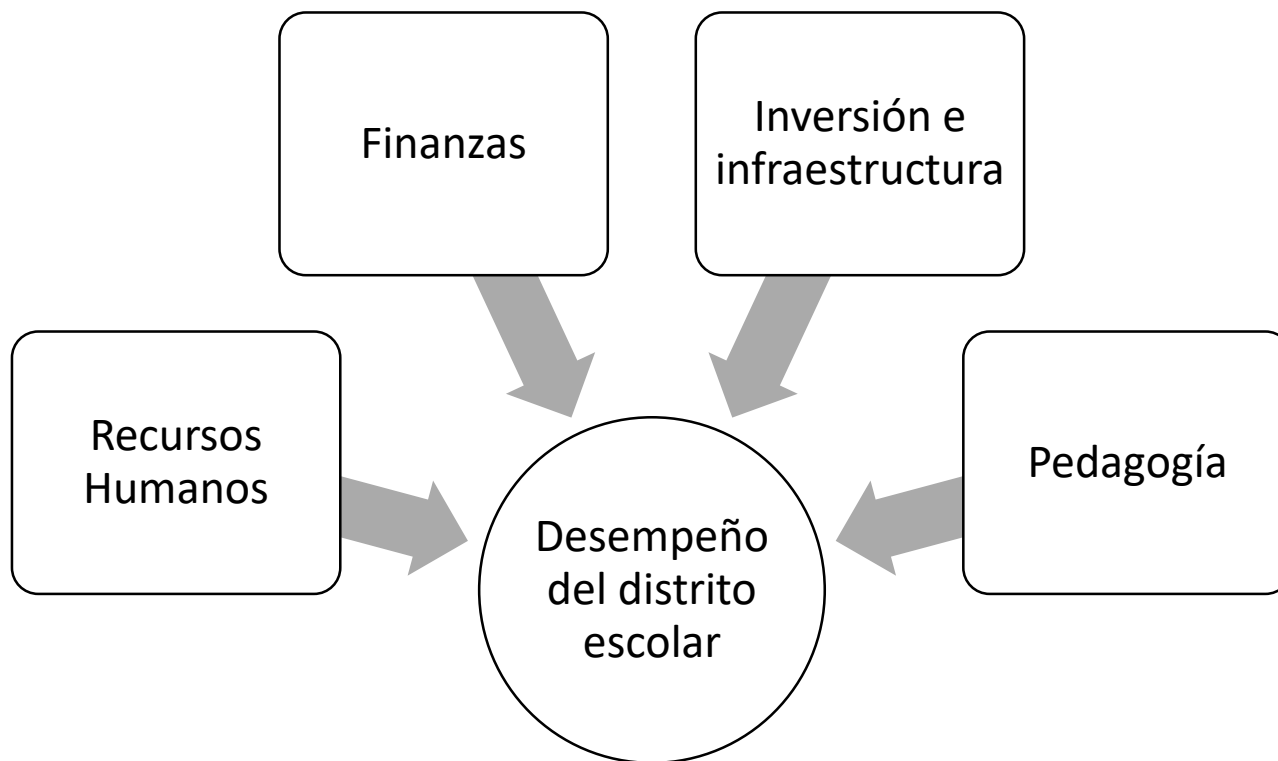
- Recompensar a los docentes de acuerdo a su evaluación
- Recompensar instituciones (e.g. colegios o universidades) de acuerdo a su desempeño.



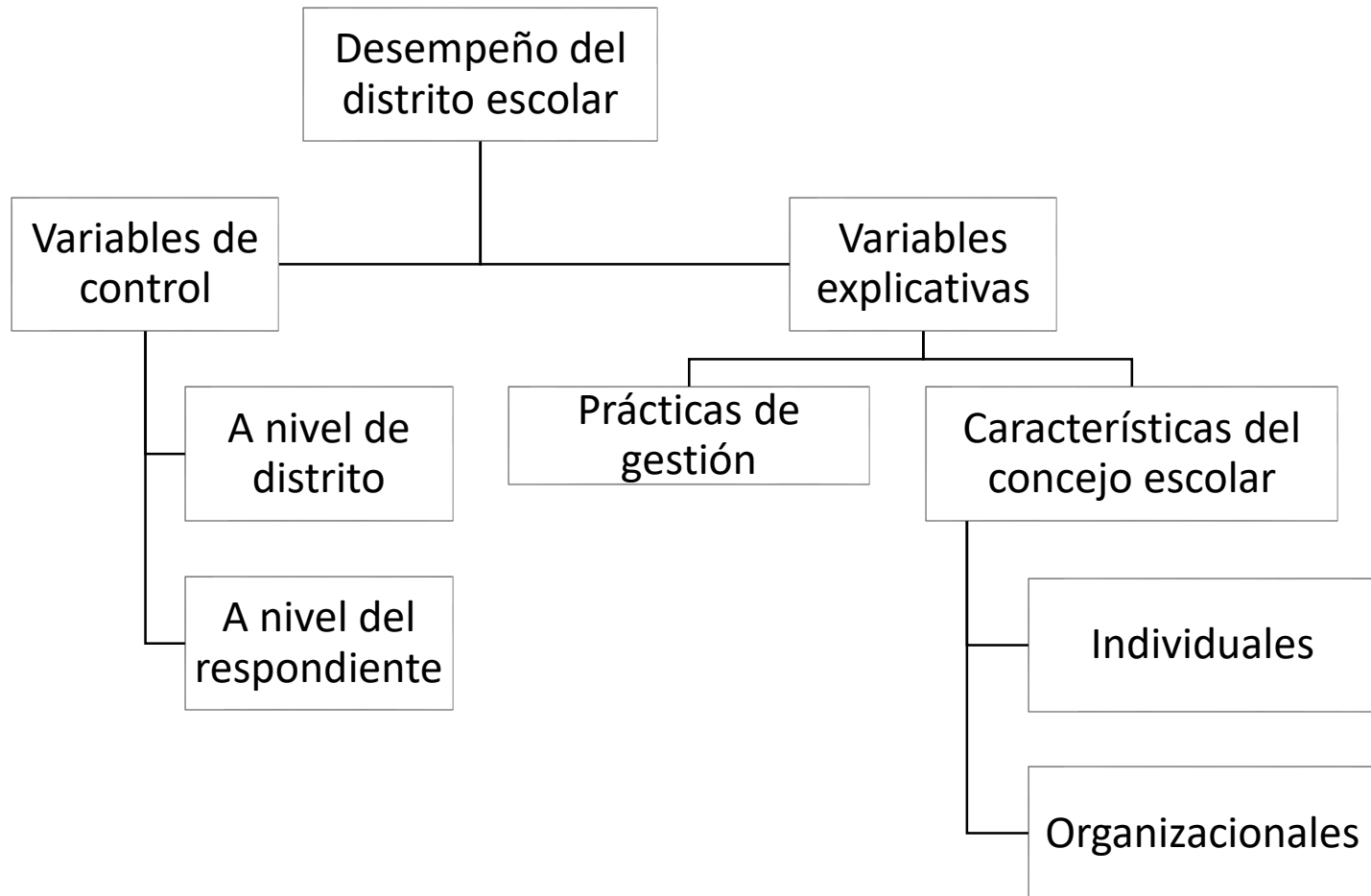
Únicamente es posible si la evaluación es considerada 'justa' por todas las partes:

- i.e.:
- Puntaje de rendimiento favorable
 - Se tienen en cuenta características del entorno

- Evaluación de los concejos escolares/distritos escolares



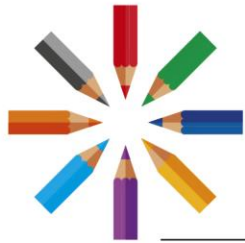
- Evaluación de los concejos escolares/distritos escolares



- Evaluación de los concejos escolares/distritos escolares

¿Qué se correlaciona con el desempeño de los distritos escolares?
(evidencia para Bélgica)

1. Desempeño más alto en los distritos no gubernamentales (concejos escolares privados).
2. Los estilos de gestión participativa son favorable para el desempeño.
3. La consolidación es mejor que la cooperación dentro de los concejos educativos.
4. Experticia de los miembros del concejo
5. El tamaño no importa <> se pueden realizar análisis de costo eficiencia (Schiltz & De Witte, 2016)



9^o Seminario
Internacional
de Investigación
sobre la calidad de la educación

DOCENTES

BOGOTÁ D.C.



GOBIERNO
DE COLOMBIA



MINEDUCACIÓN

icfes
mejor saber

¿Cómo evaluar a los docentes de manera ‘justa’?

Kristof De Witte

Universidad de Leuven, Bélgica

Universidad de Maastricht, Holanda

Noviembre 2, 2018

* La traducción de esta presentación fue realizada por el equipo de la Oficina de Investigación del Icfes.