EVOLUCIÓN DE LOS RESULTADOS DE LOS OPERATIVOS DE EVALUACIÓN DE LA CALIDAD EDUCATIVA

1995 – 1999

Puntuaciones equiparadas de las pruebas de lengua y de matemática

JUAN CRUZ PERUSIA

◆

Documento N°13

Estudios Especiales

MINISTERIO de EDUCACIÓN PRESIDENCIA de la NACIÓN

IDECE Instituto para el Desarrollo de la Calidad Educativa

EVOLUCIÓN DE LOS RESULTADOS DE LOS OPERATIVOS DE EVALUACIÓN DE LA CALIDAD EDUCATIVA. 1995 - 1999

El presente informe tiene como objetivo mostrar de manera sintética cuáles han sido las tendencias de los resultados para el total del país de los operativos de evaluación de la calidad educativa en el período que va desde 1995 a 1999.

Estas tendencias no pueden obtenerse a partir de una comparación directa de los resultados (porcentaje de aciertos) que se han ido informando luego de cada uno de los operativos de evaluación debido a que los niveles de dificultad de las pruebas han variado de año en año. En consecuencia, la variabilidad de resultados no se debe sólo a cambios en los niveles de conocimientos de los alumnos, sino también a que las pruebas han sido más fáciles o más difíciles.

Debido a lo expuesto en el párrafo anterior el Instituto para el Desarrollo de la Calidad Educativa ha hecho equiparar las puntuaciones con el objeto de analizar las tendencias en el tiempo de los resultados de las pruebas. De manera resumida, la equiparación es un proceso estadístico que se utiliza para ajustar las diferencias de dificultad que pudieran existir entre los exámenes tomados en distintos años. Una vez ajustado este factor, las diferencias en las puntuaciones sólo pueden originarse por cambios en los contenidos de las pruebas o por distintos niveles de conocimientos de los alumnos.

Para realizar la equiparación de puntuaciones se utilizó el modelo de Rasch de la Teoría de Respuesta al Ítem, lo que permitió obtener estimaciones del nivel de aptitud de cada uno de los alumnos evaluados. Luego, los resultados obtenidos para un mismo curso a lo largo de los distintos años fueron llevados, por medio de una transformación lineal, a una escala o distribución con media igual a 500 y desvío estándar de 100[1].

Este informe está organizado en 3 secciones más un anexo. En la primera se presentan estimaciones de los índices de dificultad de las pruebas. En la segunda sección se expone en detalle cuál ha sido la evolución de los resultados nacionales equiparados con el objeto de mostrar las diferencias que existen con las tendencias de resultados sin equiparar. En última sección se analiza la evolución de los resultados equiparados nacionales para los distintos cursos a través de las estimaciones de los promedios de rendimiento por materia para cada curso y mediante el cálculo de las

[1] Los resultados que se presentan en este informe fueron tomados deMonseur, C. 2000. *"Trends indicators of student achievement. Item response theory scaling of literacy and mathematics achievement data from 1995 to 1999 in five grades"*. ACER, Australia.

probabilidades de respuestas correctas. Finalmente, en el anexo se presenta un conjunto de cuadros donde se exponen los principales resultados de las estimaciones realizadas.

1. Dificultad de las pruebas

En esta sección se exponen estimaciones de los niveles de dificultad de las pruebas luego de realizada la equiparación de los distintos ítems que compusieron cada una de ellas.

El índice de dificultad de una pregunta o ítem hace referencia al nivel de conocimientos que un alumno debería tener para tener una probabilidad del 50 % de responder correctamente esa pregunta, por lo que los índices de dificultad y el nivel de rendimientos de los alumnos están expresados en la misma escala.

Los valores que asumen los índices de dificultad de cada prueba por año y curso son un promedio de los índices de dificultad de cada pregunta que compuso la evaluación de ese año. El promedio de todos los ítems para todos los años en un mismo curso ha sido centrado en el valor 500 con una desviación estándar de 100; esta escala es la que comúnmente se utiliza en los informes de resultados de las pruebas TIMSS organizada por la IEA (International Association for the Evaluation of Educational Achievement) y de la NAEP (National Assessment of Educational Progress) del Departamento de Educación de Estados Unidos.

Dada la naturaleza del proceso de estimación de los índices de dificultad, se considera oportuno destacar que los mismos son comparables sólo entre años de *un mismo* curso y *no entre* cursos.

En los cuadros 1 y 2 se presentan los índices de dificultad con la escala descripta anteriormente por área del conocimiento evaluada y por cursos[2]. Están resaltados los mayores índices de dificultad correspondientes a cada materia y curso.

Índices de dificultad de las pruebas

Cuadro 1: Lengua

Año	3º	6º	7º	9º	III
1995	-	-	513	520	456
1996	485	531	516	495	502
1997	513	499	507	504	503
1998	498	476	465	486	501
1999	503	500	490	494	542

Cuadro 2: Matemática

Año	3º	6º	7º	9º	III
1995	-	-	483	568	529
1996	508	553	534	497	504
1997	551	516	525	493	483
1998	486	466	477	466	475
1999	472	465	479	468	505

Puede verse que los niveles de dificultad de las pruebas no han sido uniformes para las distintas evaluaciones. En general, para cada curso (excepto para III), los índices de dificultad tienen sus valores máximos en el mismo año, independientemente del área evaluada.

2. Evolución de los resultados nacionales

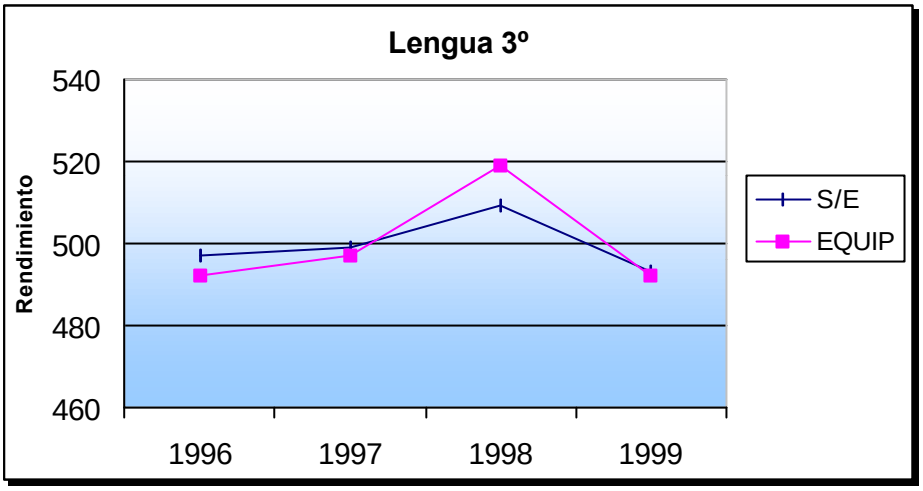
El análisis de la evolución de los resultados se realiza a través de un gráfico y un cuadro para cada curso y materia evaluada En el primero, se exponen la evolución de las puntuaciones equiparadas y sin equiparar. En el cuadro, a su vez, se presentan los resultados de las pruebas de hipótesis realizadas para determinar si las diferencias entre años pueden ser consideradas significativamente distintas de cero o no. Cuando esta diferencia es significativa, el valor que se consigna es 1; en cambio, si la diferencia no es estadísticamente significativa el valor consignado es 0.

Para evitar interpretaciones incorrectas de estos gráficos es importante remarcar una vez más que la comparación de resultados de las pruebas sin equiparar entre años no es conceptualmente correcta porque las dificultades de las mismas no son iguales; sin embargo, se ha decidido incluir en estos gráficos de tendencias estos resultados con el objeto de que se puedan advertir las diferencias que surgen cuando se analizan las tendencias en el tiempo una vez que las pruebas se ajustan por su nivel de dificultad.

[2] Los cursos son 3º, 6º, 7º y 9º Año de la EGB y el 3er. año de la Educación Polimodal / 5º-6º Nivel Medio (designado con III de aquí en adelante para que no se confunda con los niveles de EGB).

CURSO 3º

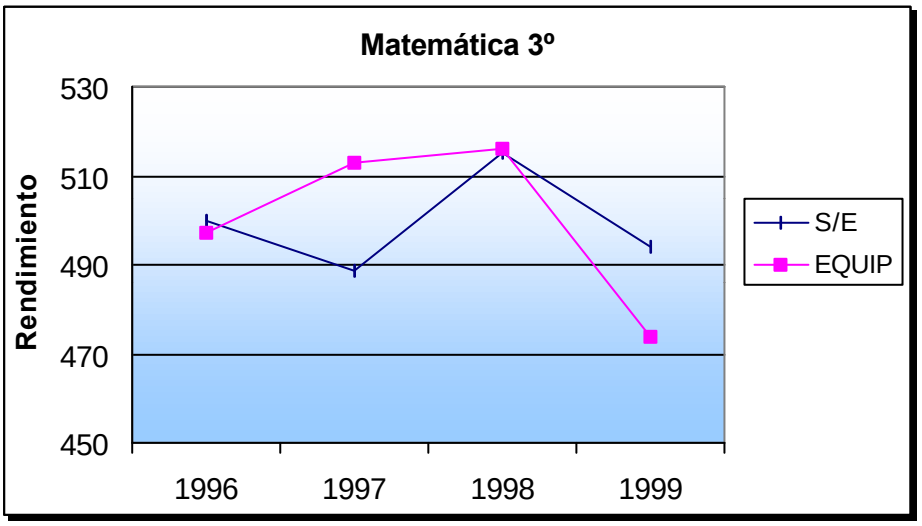
Lengua



Cuadro 3: Comparaciones entre años

	1997	1998	1999
1996	0	1	0
1997		1	0
1998			1

Matemática

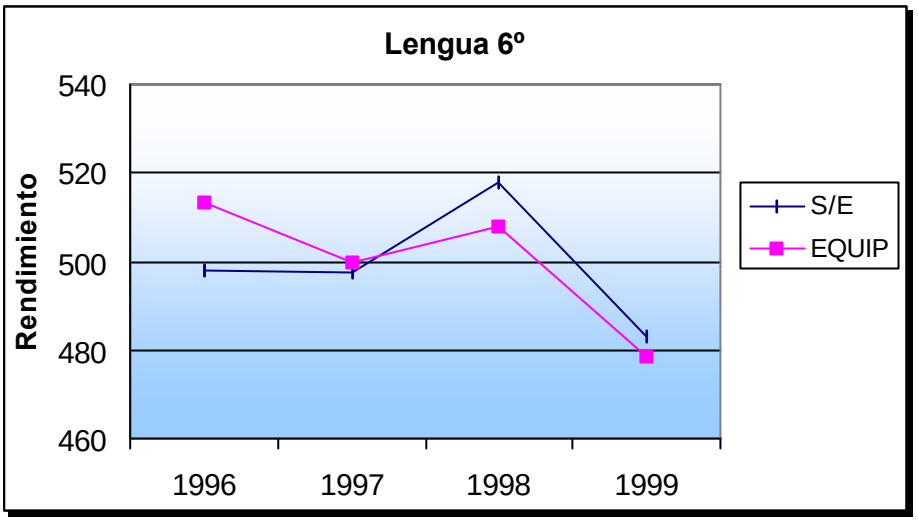


Cuadro 4: Comparaciones entre años

	1997	1998	1999
1996	1	1	1
1997		0	1
1998			1

Curso 6º

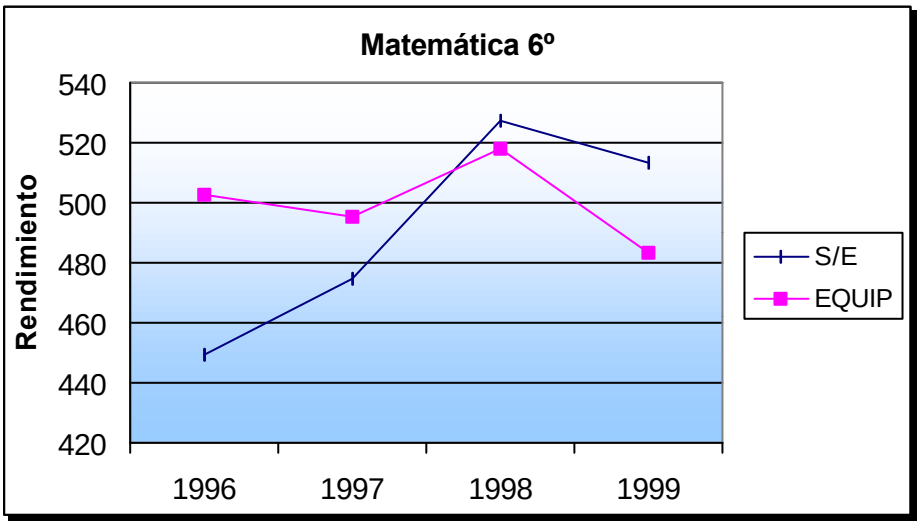
Lengua



Cuadro 5: Comparaciones entre años

	1997	1998	1999
1996	1	0	1
1997		1	1
1998			1

Matemática

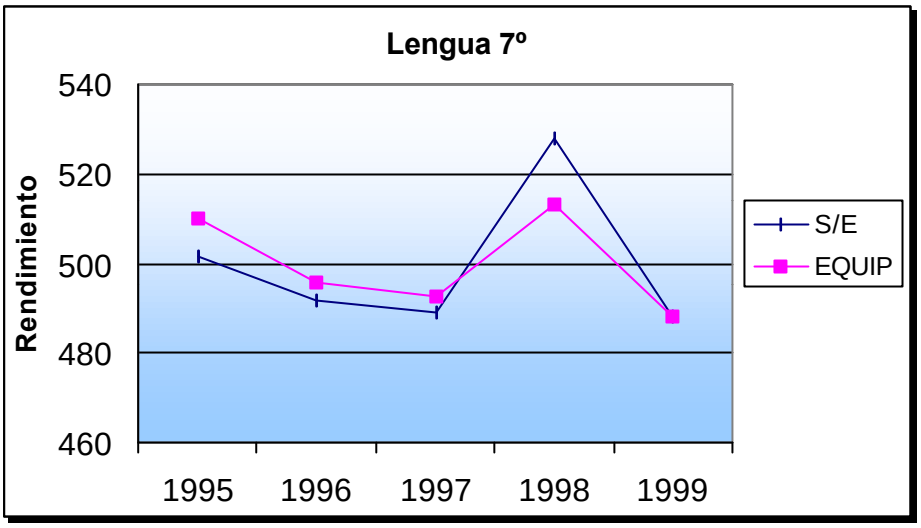


Cuadro 6: Comparaciones entre años

	1997	1998	1999
1996	1	1	1
1997		1	1
1998			1

CURSO 7º

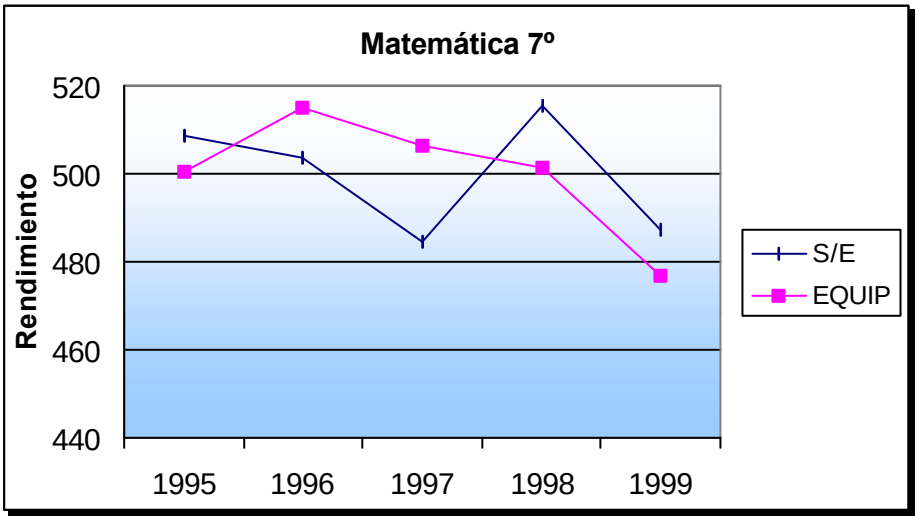
Lengua



Cuadro 7: Comparaciones entre años

	1996	1997	1998	1999
1995	1	1	0	1
1996		0	1	1
1997			1	0
1998				1

Matemática

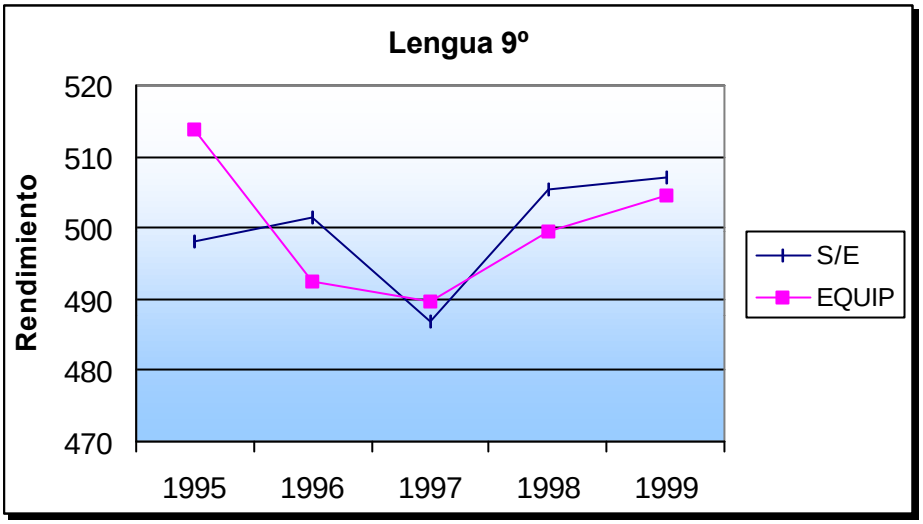


Cuadro 8: Comparaciones entre años

	1996	1997	1998	1999
1995	1	0	0	1
1996		1	1	1
1997			0	1
1998				1

CURSO 9º

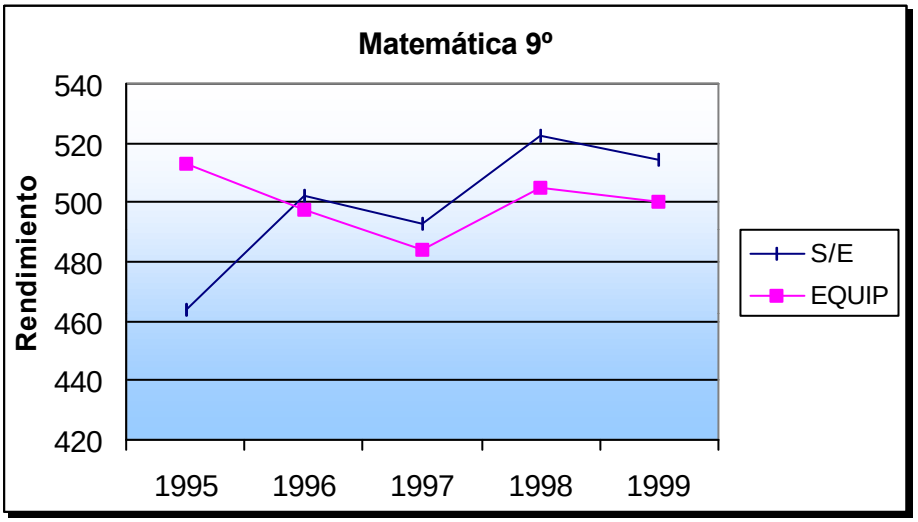
Lengua



Cuadro 9: Comparaciones entre años

	1996	1997	1998	1999
1995	1	1	1	0
1996		0	1	1
1997			1	1
1998				0

Matemática

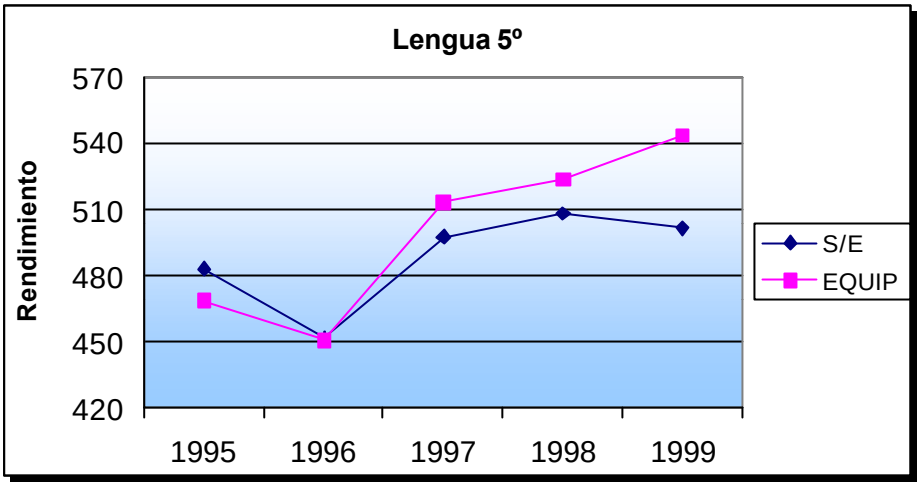


Cuadro 10: Comparaciones entre años

	1996	1997	1998	1999
1995	1	1	1	1
1996		1	1	0
1997			1	1
1998				0

CURSO III

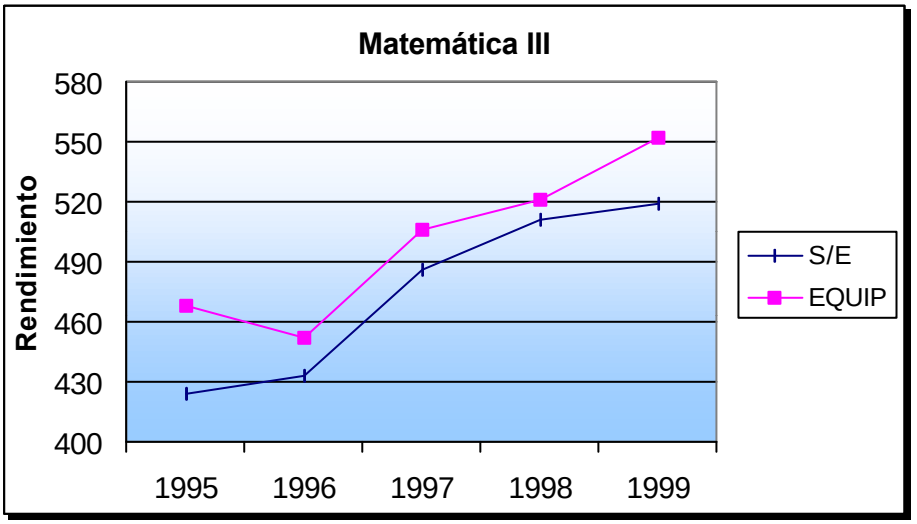
Lengua



Cuadro 11: Comparaciones entre años

	1996	1997	1998	1999
1995	1	1	1	1
1996		1	1	1
1997			1	1
1998				1

Matemática



Cuadro 12: Comparaciones entre años

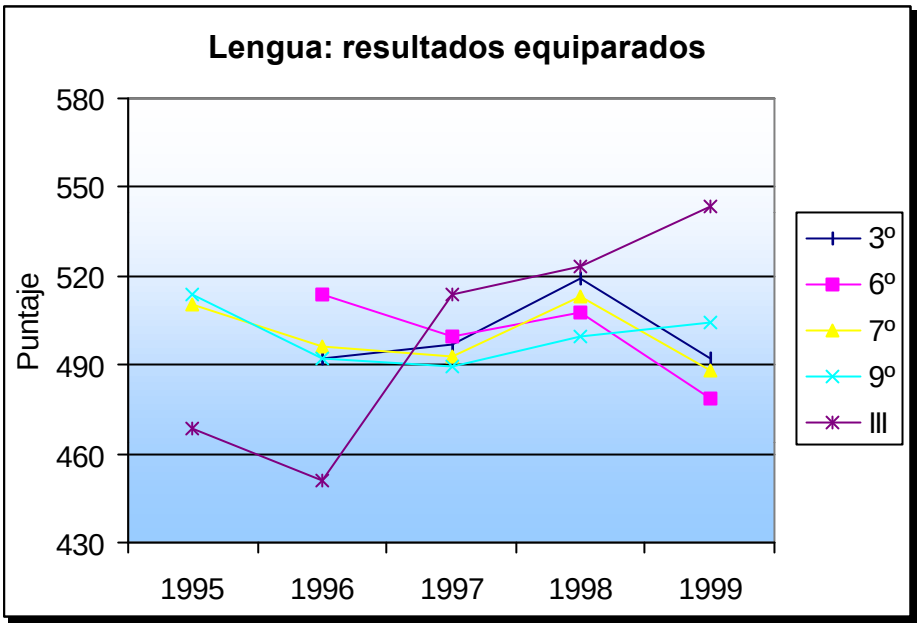
	1996	1997	1998	1999
1995	1	1	1	1
1996		1	1	1
1997			1	1
1998				1

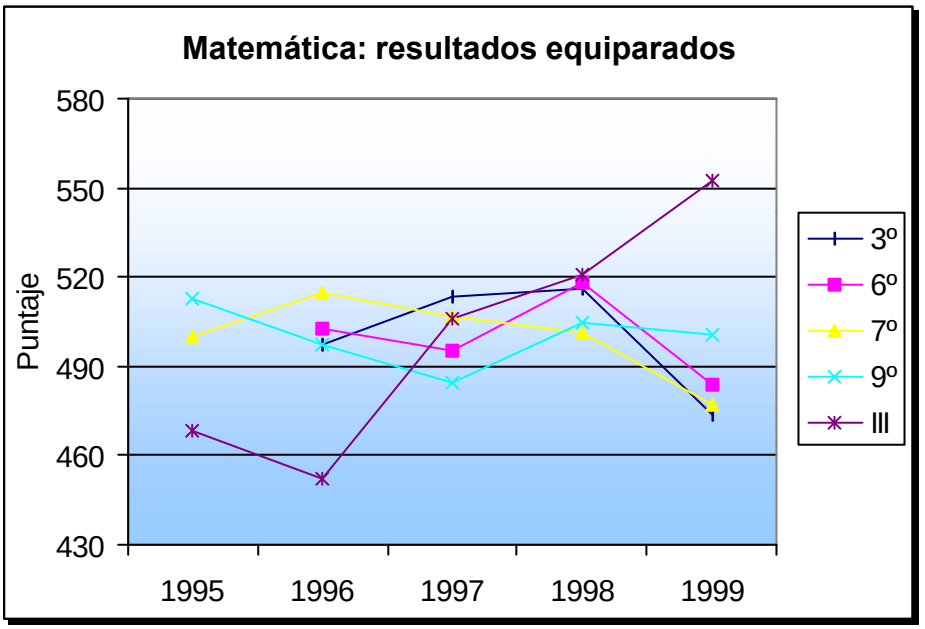
Una primera conclusión que podemos obtener al observar la evolución de los resultados equiparados es sobre el peligro de realizar comparaciones entre años sin tener en cuenta el nivel de dificultad de las pruebas. Existen casos en que la tendencia de los resultados cambia de manera importante, éstos son: Matemática de 3º entre 1996 y 1997, Lengua y Matemática de 6º en el mismo período, ambas materias de 9º entre 1995 y 1996 y finalmente, en Matemática de 7º, hay modificaciones entre 1995 y 1996 y entre 1997 y 1998.

Estos cambios de tendencias se dan, como es lógico, en aquellos períodos en que los niveles de dificultad de las pruebas han diferido bastante del promedio de todo el período que se analiza. Tomando por ejemplo Matemática de 3º entre 1996 y 1997, al observar los resultados sin equiparar, parecería ser que el nivel de aptitud o conocimientos de los alumnos ha caído, sin embargo, lo que sucedió en realidad es que la prueba de 1997 tuvo un índice de dificultad muy elevado (551, ver cuadro 2 de la sección 2) y por este motivo los resultados obtenidos en el operativo de evaluación fueron más bajos que los de 1996.

3. Análisis de los resultados nacionales equiparados

En los gráficos que siguen se expone la evolución de los promedios nacionales de rendimiento por materia evaluada para cada uno de los cursos examinados.





Debido a que los tamaños de las muestras son en todos los casos muy grandes, la mayoría de los tests de hipótesis encuentran que las diferencias observadas son estadísticamente significativas. En consecuencia, para realizar una mejor apreciación de la evolución de los resultados de año en año, se recomienda analizar la probabilidad promedio esperada de respuesta correcta que proporciona el modelo de Rasch.

La probabilidad estimada para cada materia y año debe interpretarse de la siguiente manera: si todos los alumnos de un determinado curso en un año determinado hubiesen respondido todas las preguntas de los exámenes de los distintos años, uno

esperaría observar una probabilidad promedio de éxito (que todos los alumnos respondan correctamente todas las preguntas) igual a esa probabilidad proporcionada por el modelo de Rasch. Este número *no* es el porcentaje de respuestas correctas a esperar de todos los alumnos de ese año para la prueba de ese mismo año solamente.

En los cuadros que siguen se exponen estas probabilidades para los distintos cursos evaluados y por áreas.

Probabilidades de respuestas correctas

Cuadro 13: Matemática

Año	3º	6º	7º	9º	III
1995	-	-	0.5367	0.5312	0.5402
1996	0.6197	0.5695	0.5729	0.4952	0.4980
1997	0.6559	0.5529	0.5518	0.4648	0.6367
1998	0.6622	0.6053	0.5400	0.5125	0.6718
1999	0.5645	0.5248	0.4787	0.5022	0.7409

Cuadro 14: Lengua

Año	3º	6º	7º	9º	III
1995	-	-	0.6396	0.5895	0.5852
1996	0.6389	0.6102	0.6066	0.5440	0.5427
1997	0.6506	0.5792	0.5986	0.5383	0.6842
1998	0.7026	0.5980	0.6459	0.5591	0.7045
1999	0.6389	0.5308	0.5880	0.5698	0.7428

Del análisis de los cuadros de probabilidades y de los gráficos de esta sección se pueden inferir las siguientes conclusiones:

- Para 3º, 6º y 7º Año;
  - hay una importante reducción de los resultados entre 1998 y 1999 que es independiente de las áreas evaluadas;
  - 1998 es el año en que se obtuvieron, en general, los mejores resultados (excepto lengua 6º y matemática 7º);
  - existe una cierta estabilidad de resultados que se ve alterada por el incremento producido entre 1997 y 1998.
- Para 9º Año;
  - hay una caída en los resultados en los tres primeros períodos, seguida luego por una recuperación que permite llegar de manera aproximada a los niveles de resultados similares a los obtenidos en 1995.

- Para 3er. Año Polimodal / 5º-6º Medio;
    - hay un incremento muy importante de los resultados a partir de 1996.

Anexo

En los cuadros 1 a 10 se presentan diferentes resultados por cada año y materia evaluada. Para cada año escolar, cada cuadro provee:

- N: número de alumnos utilizado para la estimación de los parámetros.
- Media: promedio del rendimiento o habilidad estimado para cada alumno.
- SD: desvío estándar de la distribución.
- SE (Media): error estándar de la media.
- Prob.: probabilidad promedio esperada de éxito de los alumnos para responder todas las preguntas que se evaluaron en los distintos años. Se derivan del modelo de Rasch.

Las estimaciones de media, desvío estándar y error estándar que se presentan aquí están en una escala basada en una distribución logística, donde la mayoría de los valores se encuentran comprendidos entre –3 y 3. Estos valores son los que luego fueron reescalados a la métrica de media 500 y desvío estándar 100 y que se exponen en los gráficos de las distintas secciones del presente documento.

3º AÑO EGB

Lengua

Cuadro 1: Estimaciones de la media, desvío estándar y error estándar.

Año	*Ni*	*Media*	*SD*	*SE(Media)*	*Prob.*
1996	37705	0.5707	1.0477	0.0158	0.6389
1997	36897	0.6217	0.9904	0.0161	0.6506
1998	44611	0.8596	1.1120	0.0158	0.7026
1999	39204	0.5706	1.1384	0.0177	0.6389

Matemática

Cuadro 2: Estimaciones de la media, desvío estándar y error estándar.

Año	*Ni*	*Media*	*SD*	*SE(Media)*	*Prob.*
1996	41386	0.4881	0.9218	0.0127	0.6197
1997	36634	0.6451	0.9416	0.0152	0.6559
1998	43114	0.6732	1.0070	0.0147	0.6622
1999	38130	0.2593	1.0089	0.0159	0.5645

6º AÑO EGB

Lengua

Cuadro 3: Estimaciones de la media, desvío estándar y error estándar.

Año	*Ni*	*Media*	*SD*	*SE(Media)*	*Prob.*
1996	11343	0.4482	0.9052	0.0207	0.6102
1997	40810	0.3196	0.9388	0.0125	0.5792
1998	41642	0.3972	0.9486	0.0125	0.5980
1999	36489	0.1235	0.9133	0.0127	0.5308

Matemática

Cuadro 4: Estimaciones de la media, desvío estándar y error estándar.

Año	*Ni*	*Media*	*SD*	*SE(Media)*	*Prob.*
1996	11883	0.2797	0.8383	0.0212	0.5695
1997	39301	0.2124	0.9092	0.0134	0.5529
1998	40580	0.4278	0.9817	0.0136	0.6053
1999	36218	0.0992	1.0149	0.0140	0.5248

7º AÑO EGB

Lengua

Cuadro 5: Estimaciones de la media, desvío estándar y error estándar.

Año	*Ni*	*Media*	*SD*	*SE(Media)*	*Prob.*
1995	33750	0.5736	0.9925	0.0155	0.6396
1995	34004	0.4331	0.9570	0.0151	0.6066
1997	37168	0.3997	0.9348	0.0136	0.5986
1998	34706	0.6011	1.0257	0.0147	0.6459
1999	29594	0.3559	0.9793	0.0152	0.5880

Matemática

Cuadro 6: Estimaciones de la media, desvío estándar y error estándar.

Año	*Ni*	*Media*	*SD*	*SE(Media)*	*Prob.*
1995	32996	0.1472	0.9987	0.0174	0.5367
1995	34204	0.2936	1.0057	0.0176	0.5729
1997	36324	0.2078	0.9756	0.0153	0.5518
1998	34725	0.1602	0.9935	0.0163	0.5400
1999	30399	-0.0852	0.9305	0.0159	0.4787

9º AÑO EGB

Lengua

Cuadro 7: Estimaciones de la media, desvío estándar y error estándar.

Año	*Ni*	*Media*	*SD*	*SE(Media)*	*Prob.*
1995	36142	0.3621	0.8566	0.0138	0.5895
1995	48005	0.1765	0.8374	0.0109	0.5440
1997	38320	0.1534	0.8534	0.0125	0.5383
1998	39814	0.2375	0.8778	0.0131	0.5591
1999	40317	0.2809	0.9031	0.0135	0.5698

Matemática

Cuadro 8: Estimaciones de la media, desvío estándar y error estándar.

Año	*Ni*	*Media*	*SD*	*SE(Media)*	*Prob.*
1995	36348	0.1249	0.8724	0.0155	0.5312
1995	49337	-0.0194	0.9400	0.0135	0.4952
1997	38201	-0.1412	0.9036	0.0146	0.4648
1998	39075	0.0502	0.9267	0.0148	0.5125
1999	40584	0.0090	0.9377	0.0148	0.5022

3ER. AÑO POLIMODAL / 5º-6º NIVEL MEDIO

Lengua

Cuadro 9: Estimaciones de la media, desvío estándar y error estándar.

Año	*Ni*	*Media*	*SD*	*SE(Media)*	*Prob.*
1995	29250	0.3441	0.9070	0.0138	0.5852
1996	31514	0.1711	0.8407	0.0126	0.5427
1997	28629	0.7733	0.8803	0.0052	0.6842
1998	31309	0.8689	0.9346	0.0053	0.7045
1999	31600	1.0606	0.9273	0.0052	0.7428

Matemática

Cuadro 10: Estimaciones de la media, desvío estándar y error estándar.

Año	*Ni*	*Media*	*SD*	*SE(Media)*	*Prob.*
1995	28094	0.161	0.9969	0.0188	0.5402
1996	32971	-0.0081	0.9170	0.0161	0.4980
1997	29949	0.5609	0.9682	0.0056	0.6367
1998	29803	0.7164	1.0681	0.0062	0.6718
1999	30047	1.0504	0.9966	0.0057	0.7409