

# Aprender 2019. Análisis comparativo de los niveles de desempeño en Ciencias Naturales.

 **la educación**  
nuestra bandera

Secretaría de Evaluación  
e Información Educativa



Ministerio de Educación  
**Argentina**

# Material producido por el Ministerio de Educación de la Nación

## Autoridades

### Presidente

Dr. Alberto Fernández

### Vicepresidenta

Dra. Cristina Fernández de Kirchner

### Jefe de Gabinete de Ministros

Dr. Juan Luis Manzur

### Ministro de Educación

Lic. Jaime Perczyk

### Gabinete de Asesores

Prof. Daniel José Pico

### Secretario de Evaluación e Información Educativa

Dr. Germán Lodola

### Subsecretaria de Planeamiento, Prospectiva e Innovación

Mg. Gladys Kochen

Secretaría de Evaluación e Información Educativa

Aprender 2019 : análisis comparativo de los niveles de desempeño en Ciencias Naturales / 1a ed. - Ciudad Autónoma de Buenos Aires : Ministerio de Educación de la Nación, 2022.

Libro digital, PDF

Archivo Digital: descarga y online

ISBN 978-950-00-1639-1

1. Educación. 2. Evaluaciones. 3. Educación Secundaria. I. Título.  
CDD 373.27

Se permite la reproducción total y/o parcial con mención de la fuente.

Esta licencia abarca a toda la obra excepto en los casos que se indique otro tipo de licencia.

Material de distribución gratuita, prohibida su venta.

## ÍNDICE

1. Introducción	4
2. Alcances y objetivos	7
3. Notas sobre el insumo y la calidad de la información	8
4. Plan de análisis	9
5. Estadísticos descriptivos de niveles de desempeño por variables de corte	9
5.1. Los resultados por sector de gestión	9
5.2. Nivel de desempeño según ámbito (Urbano – Rural)	11
5.3. Nivel de desempeño según jurisdicción	13
6. Evaluación estadística de ítems	17
6.1. Notas sobre el diseño de Modelos - Instrumento de evaluación	17
6.2. Metodología de comparación	18
6.3. Cantidad de respuestas correctas a ítems: estadísticos y medidas de tendencia central	19
6.4. Comparación de distribuciones cantidad de aciertos por niveles de desempeño según tipo de ítem.	20
7. Hipótesis orientativas para la interpretación de la distribución de los niveles de desempeño en Ciencias Naturales – Aprender 2019.	24
7.1. El marco muestral	26
8. Anexo técnico	27

## Resumen ejecutivo

En el presente informe se realiza un análisis estadístico con el fin de esbozar algunas hipótesis que permitan comprender la distribución de los niveles de desempeño en Ciencias Naturales, en el operativo de evaluación – Aprender 2019 de nivel secundario.

El análisis se realizó con las bases de datos correspondientes a las evaluaciones de Ciencias Naturales de los operativos Aprender 2016 y 2019, provistas por la Secretaría de Evaluación e Información Educativa del Ministerio de Educación de la Nación.

El plan de análisis de los datos intentará determinar si existen factores exógenos al instrumento y/o datos que puedan estar introduciendo algún nivel de distorsión en la distribución de los resultados. Para ello, se aislarán algunas variables de corte que permitan identificar si los problemas tienen su origen en algún subuniverso relevante y si éstos afectan la distribución general.

En segundo lugar, se evaluará si existen problemas de diseño del instrumento que estén sesgando la distribución. El objetivo es descartar la existencia de problemas en los diseños de las preguntas (ítems) de las evaluaciones que hayan redundado en un nivel de dificultad menor de la evaluación de 2019, de modo que explique el mejor nivel general de desempeño en relación al operativo inmediato anterior (2016).

Finalmente, se formularán algunas hipótesis que tiendan a explicar el origen de las diferencias en base a un análisis estadístico de las bases de datos mencionadas.

## 1. Introducción

El análisis de los datos recabados de la prueba Aprender 2019 ha arrojado diferencias llamativas en los resultados de las evaluaciones de Ciencias Naturales con respecto a evaluaciones anteriores aplicadas en el nivel secundario. Esto es, un mejor desempeño general.

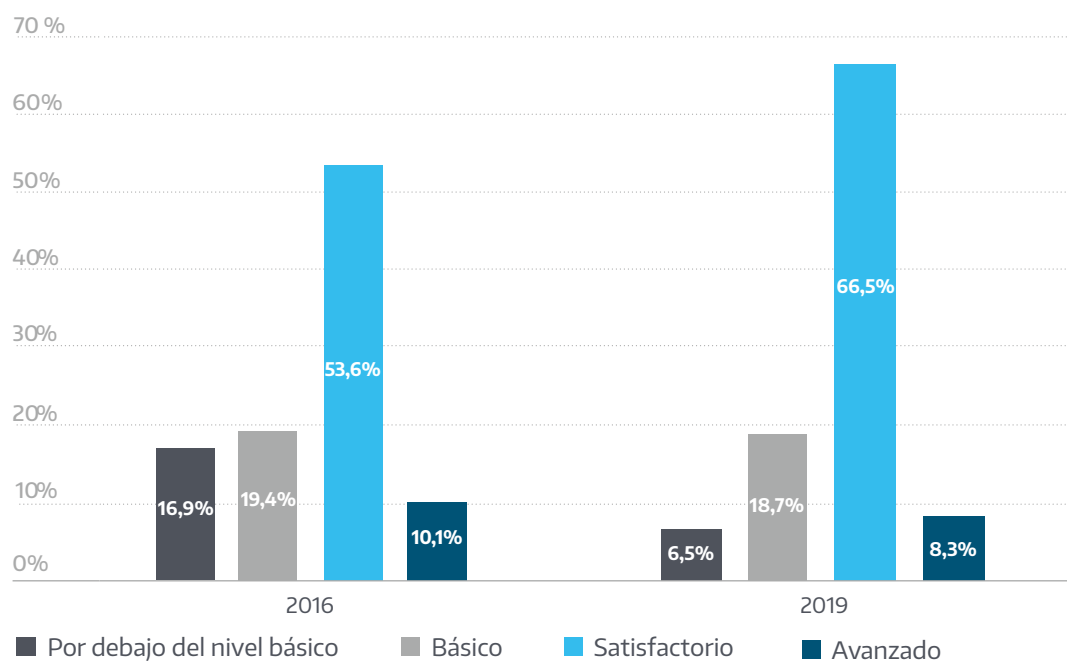
La magnitud de las diferencias entre los relevamientos de 2016 y 2019 en las evaluaciones de Ciencias Naturales, no son esperables en períodos de tiempo tan breves.

---

<sup>1</sup> Para un detalle metodológico sobre el tratamiento de los datos y la ficha técnica de cobertura ver el Anexo Técnico

**Gráfico 1.1.**

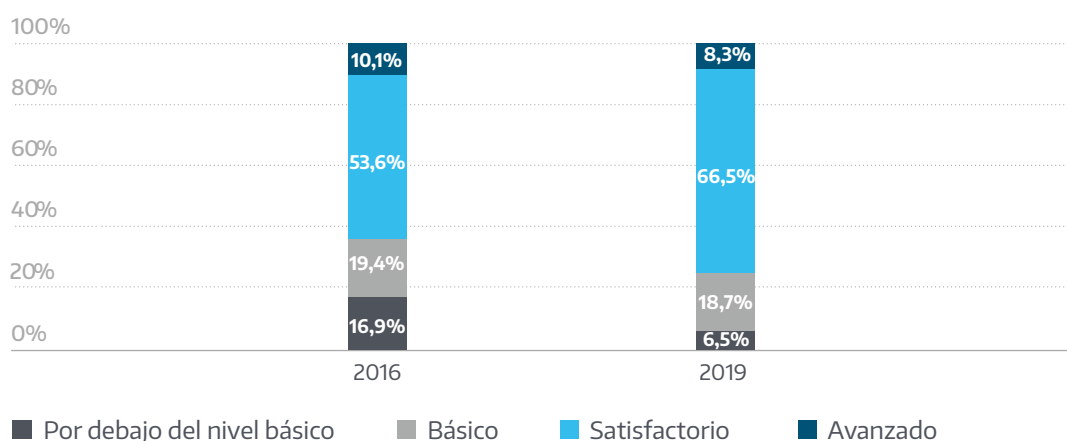
Nivel de desempeño en Ciencias Naturales, por operativo de evaluación.  
Aprender 2016-2019



**Fuente:** Evaluaciones de Ciencias Naturales. Aprender 2016 y Aprender 2019. DNIE. Ministerio de Educación de la Nación.

**Gráfico 1.2.**

Nivel de desempeño en Ciencias Naturales, por operativo de evaluación.  
Aprender 2016-2019



**Fuente:** Evaluaciones de Ciencias Naturales. Aprender 2016 y Aprender 2019. DNIE. Ministerio de Educación de la Nación.

Mientras que en 2016 la media del nivel de desempeño en Ciencias Naturales era de 2,57, en 2019 se elevó a 2,83.

En términos estadísticos, la distribución relativa de casos por categoría indica que existiría un traslado de las categorías 1 y 2 (*Por debajo del nivel nivel básico y Básico*) hacia la categoría 3 (*Satisfactorio*).

**Tabla 1.1.**

Porcentaje de estudiantes según su nivel de desempeño en Ciencias Naturales según operativo de evaluación. Aprender 2016-2019

Nivel de desempeño	2016	2019	Diferencia
1 Por debajo del nivel básico	16,9%	6,5%	-10,4
2 Básico	19,4%	18,7%	-0,7
3 Satisfactorio	53,6%	66,5%	12,9
4 Avanzado	10,1%	8,3%	-1,8
<b>Total</b>	<b>100%</b>	<b>100%</b>	

**Fuente:** Evaluaciones de Ciencias Naturales. Aprender 2016 y Aprender 2019. DNIE. Ministerio de Educación de la Nación.

Este fenómeno implica que se produce una menor dispersión de los resultados de las evaluaciones, y que tienden a concentrar a dos tercios de los casos en una sola categoría. De este modo, en la tabla siguiente observamos la reducción de todas las medidas de dispersión.

**Tabla 1.2.**

Medidas de dispersión

	2016	2019
Desviación estándar	0,887	0,67948
Varianza	0,786	0,462
Curtosis	-0,591	1,291

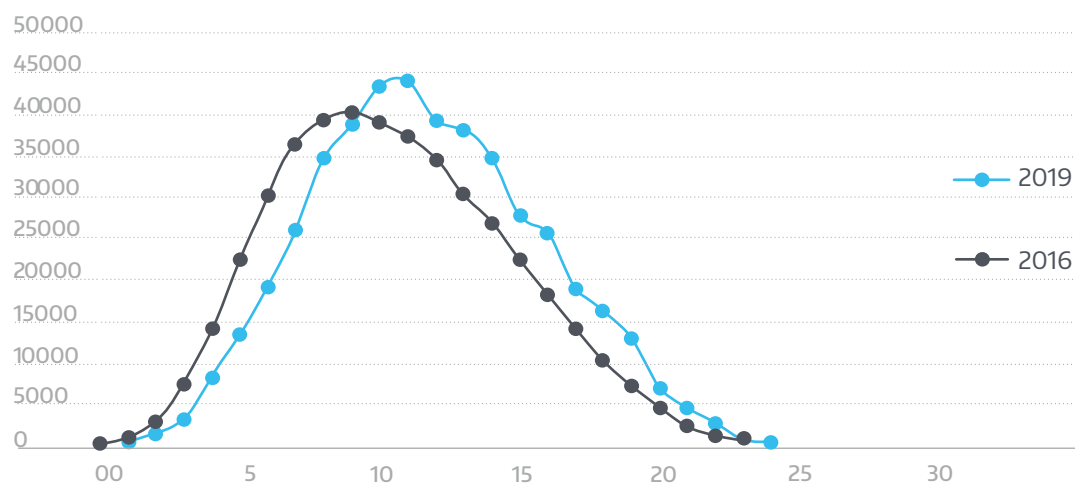
**Fuente:** Evaluaciones de Ciencias Naturales. Aprender 2016 y Aprender 2019. DNIE. Ministerio de Educación de la Nación.

Puntualmente, es importante señalar el cambio de signo de la curtosis, que nos indica un cambio en la distribución. En tanto se pasa de un tipo de distribución platicúrtica (tendiendo a mesocúrtica) con una leve asimetría hacia la izquierda en 2016, a una distribución leptocúrtica en 2019 .

<sup>1</sup> La media de esta variable categorial no tiene correlato empírico para este constructo, pero se decidió incorporarla como indicador matemático por sus capacidades abstractas de resumen de la información.

### Gráfico 1.3.

Cantidad de respuestas correctas en Ciencias Naturales, según operativo de evaluación. Aprender 2016-2019



**Fuente:** Evaluaciones de Ciencias Naturales. Aprender 2016 y Aprender 2019. DNIE. Ministerio de Educación de la Nación.

En el gráfico 1.3, en la distribución de frecuencias de cantidad de respuestas correctas podemos constatar que la asimetría se atenuó en 2019.

Una vez definido el problema se explorarán los datos con el fin de arribar a algunas hipótesis sobre el origen de estas diferencias.

## 2. Alcances y objetivos

En este trabajo se observarán diferentes factores que puedan estar incidiendo de manera no deseada en la distribución de los niveles de desempeño en Ciencias Naturales en el operativo de evaluación de 2019.

En primer lugar, se constatará que el comportamiento es homogéneo en el conjunto del universo.

Una vez constatado esto, se analizarán estadísticamente los datos de manera de tratar de formular hipótesis significativas para explicar el fenómeno.

### 3. Notas sobre el insumo y la calidad de la información

En este trabajo, nos proponemos analizar los datos de los conjuntos de datos provistos para esbozar algunas hipótesis que permitan interpretar el origen de las diferencias detalladas en la Introducción.

Se trabajará con las dos bases de datos de los operativos de evaluación de Ciencias Naturales correspondientes a 2016 y a 2019.

Dichas bases de datos contienen información desagregada, cuya unidad de registro es el alumno.

Las bases tienen tres bloques de datos:

- a) **Datos de contexto:** Jurisdicción, localidad/paraje. Datos de los establecimientos y secciones y características del establecimiento (ámbito, sector de gestión, etc.)
- b) **Datos de ítems:** Datos codificados de cada una de las respuestas para los 24 ítems del modelo, los campos de codificación resumidos de cada uno de los ítems.
- c) **Datos estadísticos calculados:** Medidas de resumen y Campos calculados tales como nivel de desempeño y campos de ponderación del registro.

Si bien todas las bases de datos son perfectibles, se entiende que el set de datos provisto ha superado las distintas instancias de evaluación de calidad dentro del ministerio y por lo tanto, escapa al alcance de este informe el análisis de consistencia.

Los modelos de construcción y asignación de niveles de desempeño se basan en marcos teóricos sustentados en modelos estadísticos <sup>3</sup> bajo estándares internacionales y consensos federales que desbordan ampliamente el alcance del presente informe y por lo tanto tampoco serán objeto de análisis o evaluación.

---

<sup>3</sup> Para un desarrollo de la implementación de la Teoría de Respuesta al ítem, ver "2019. Informe metodológico Aprender. Aprender 2019. Equipo de Metodología. DINIEE, Ministerio de Educación".



## 4. Plan de análisis

El presente plan de análisis tiene como primer objetivo corroborar que el fenómeno sea homogéneo a los distintos recortes significativos que puedan hacerse al interior del universo. Es decir, que no existen factores externos al instrumento que estén introduciendo algún nivel de distorsión que afecte la distribución general de resultados.

En segundo término, se evaluará si existen algunos factores que, desde el diseño del instrumento, están afectando la distribución de resultados. En términos concretos, se analizarán las distribuciones de respuestas correctas acumuladas de cada tipo de ítems (anclados y no anclados).

Para realizar este análisis, la metodología consistirá en usar los ítems anclados como grupo de control para poder medir si los no anclados inciden de manera significativa en el fenómeno observado.

De esta manera quedan configurados los dos capítulos de este informe.

- 1) Estadísticos descriptivos de niveles de desempeño por variables de corte.
- 2) Evaluación estadística de ítems.

## 5. Estadísticos descriptivos de niveles de desempeño por variables de corte

En este capítulo se analizarán las distribuciones de los niveles de desempeño segmentados por diferentes variables de contexto, de manera de poder confeccionar subuniversos.

Las variables disponibles para el análisis son:

- Sector de gestión: estatal y privada
- Ámbito: urbano y rural
- Jurisdicción: Provincial

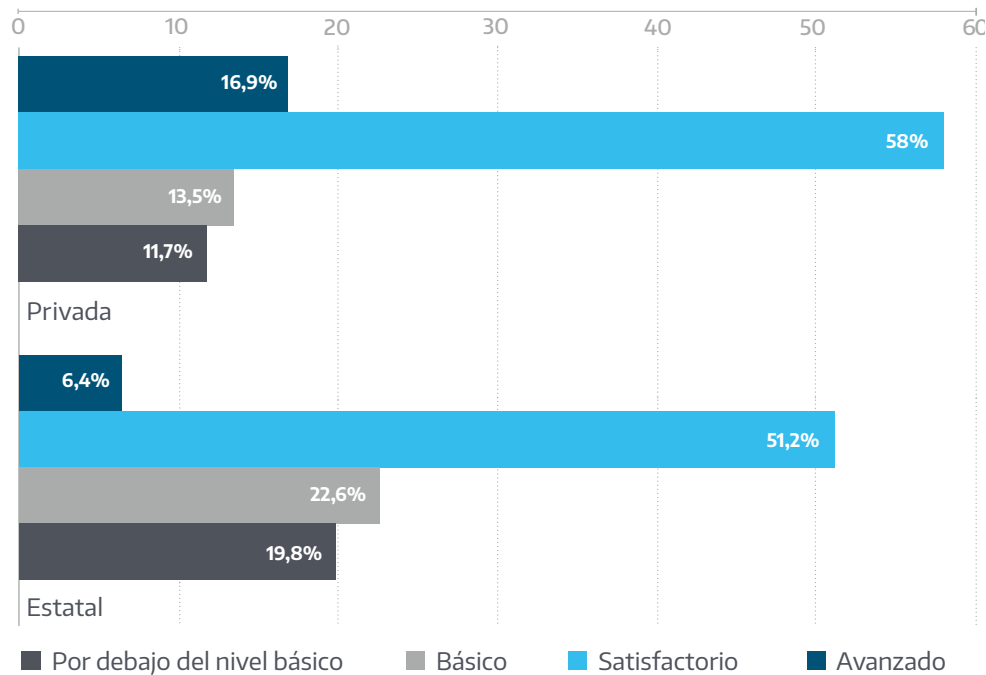
### 5.1. Los resultados por sector de gestión

Si analizamos los datos por sector de gestión observamos que la tendencia está marcada por el sector de gestión mayoritario en el sistema educativo argentino, es decir, el estatal.

Más allá de las magnitudes, debemos señalar que existen movimientos coincidentes en relación con el proceso general en ambos sectores de gestión.

### Gráfico 5.1.1.

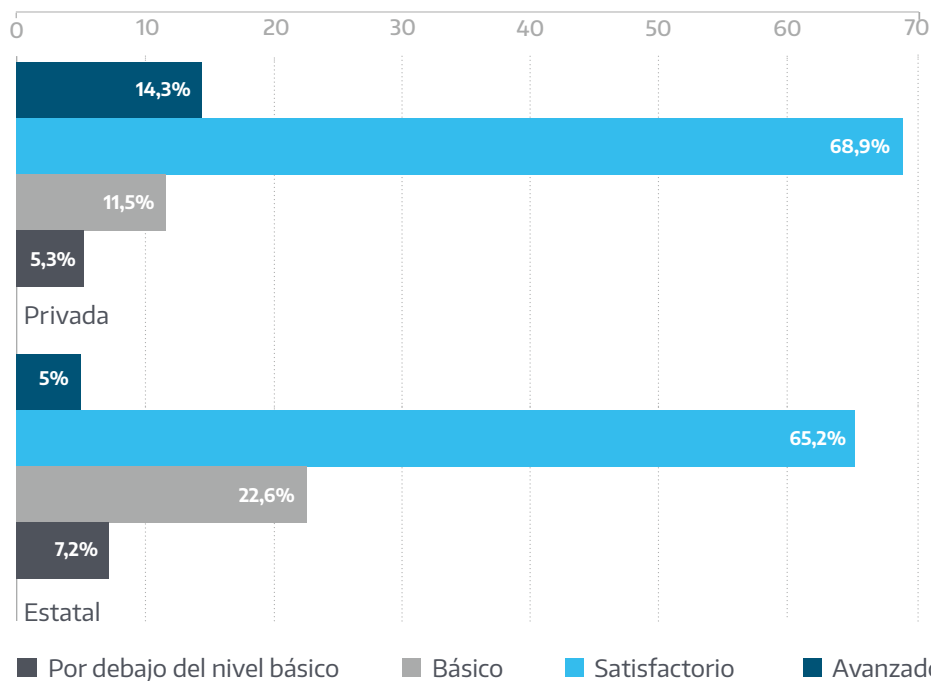
Nivel de desempeño por sector de gestión. Aprender 2016



**Fuente:** Evaluaciones de Ciencias Naturales. Aprender 2016 y Aprender 2019. DNIE. Ministerio de Educación de la Nación.

### Gráfico 5.1.2.

Nivel de desempeño por sector de gestión. Aprender 2019



**Fuente:** Evaluaciones de Ciencias Naturales. Aprender 2016 y Aprender 2019. DNIE. Ministerio de Educación de la Nación.

Aun cuando se señaló que ambos sectores de gestión tienen movimientos coincidentes hacia la pérdida de relevancia de los niveles “bajos” de desempeño (*Por debajo del nivel básico y Básico*), en el sector de gestión estatal esto es mucho más marcado. Mientras que el aumento diferencial de casos en los niveles más altos en las secundarias estatales llega a 14 puntos, en el caso de la educación privada no alcanza los 11 puntos (10.94 pp).

**Tabla 5.1.1.**

Distribución de frecuencias relativas de los niveles de desempeño por sector de gestión según operativo de evaluación. Aprender 2016-2019


Nivel de desempeño	2016		2019		Diferencia 2016-2019	
	Estatal	Privada	Estatal	Privada	Estatal	Privada
1 Por debajo del nivel básico	19,81	11,71	7,2	5,28	-12,61	-6,43
2 Básico	22,64	13,46	22,6	11,53	-0,04	-1,93
3 Satisfactorio	51,16	57,98	65,2	68,91	14,04	10,94
4 Avanzado	6,39	16,86	5	14,28	-1,39	-2,58

**Fuente:** Evaluaciones de Ciencias Naturales. Aprender 2016 y Aprender 2019. DNIE. Ministerio de Educación de la Nación.

**Tabla 5.1.2.**

Diferencias porcentuales de los niveles de desempeño entre operativos de evaluación según sector de gestión. Aprender 2016-2019

Diferencias 2016 - 2019			
	Diferencia total	Estatal	Privada
1 Por debajo del nivel básico	-10,4	-12,61	-6,43
2 Básico	-0,7	-0,04	-1,93
3 Satisfactorio	12,9	14,04	10,94
4 Avanzado	-1,8	-1,39	-2,58



**Fuente:** Evaluaciones de Ciencias Naturales. Aprender 2016 y Aprender 2019. DNIE. Ministerio de Educación de la Nación.

Para ambos sectores se observa una pérdida de casos en las categorías *Por debajo del nivel básico*, *Básico* y *Avanzado*, así como su concentración en la categoría *Satisfactorio*.

## 5.2. Nivel de desempeño según ámbito (Urbano – Rural)

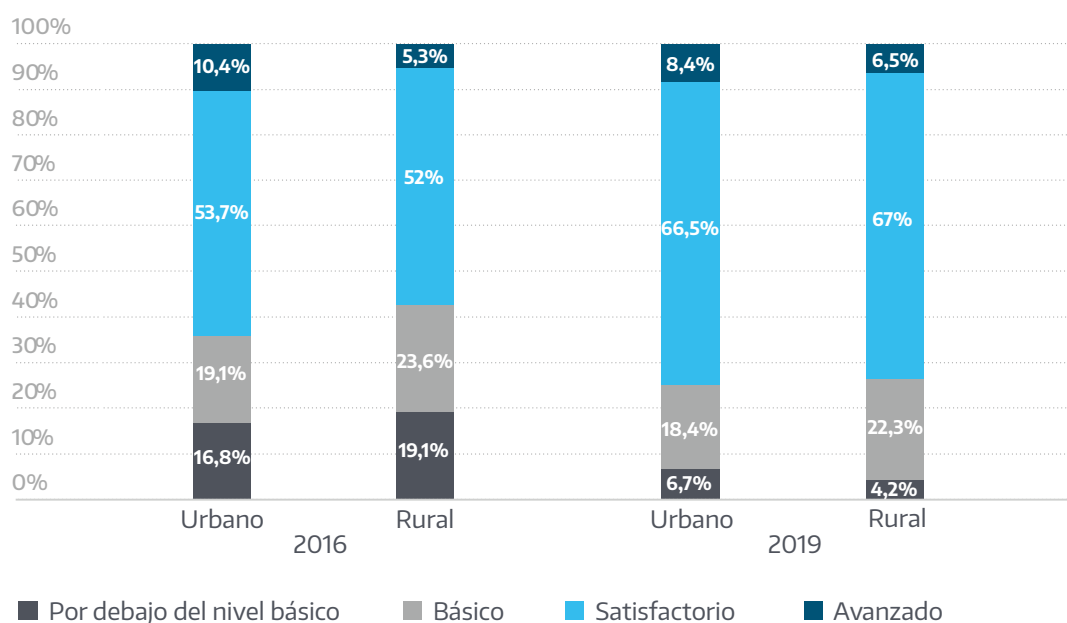
Si se observan los niveles de desempeño por ámbito, encontramos que la tendencia a la

concentración de casos en relación a la media se mantiene, mientras que se reducen de manera muy significativa en la categoría *Por debajo del nivel básico*.

Esta reducción en las categorías *Por debajo del nivel básico*, se produce de manera mucho más marcada en el ámbito rural, con 5 pp. de diferencia (10.1 -14.83).

### Gráfico 5.2.1.

Niveles de desempeño según ámbito por operativo de evaluación.



**Fuente:** Evaluaciones de Ciencias Naturales. Aprender 2016 y Aprender 2019. DNIE. Ministerio de Educación de la Nación.

### Tabla 5.2.1.

Distribución de frecuencias relativas de los niveles de desempeño por ámbito según operativo de evaluación. Aprender 2016-2019

Nivel de desempeño	2016		2019		Diferencias 2016-2019	
	Estatal	Privada	Estatal	Privada	Estatal	Privada
1 Por debajo del nivel básico	16,78	19,05	6,68	4,23	-10,10	-14,83
2 Básico	19,08	23,61	18,45	22,31	-0,63	-1,30
3 Satisfactorio	53,69	52,00	66,48	66,98	12,78	14,98
4 Avanzado	10,45	5,34	8,40	6,49	-2,05	1,15

**Fuente:** Evaluaciones de Ciencias Naturales. Aprender 2016 y Aprender 2019. DNIE. Ministerio de Educación de la Nación.

### 5.3. Nivel de desempeño según jurisdicción

Para analizar las diferencias por jurisdicción, se trabajará con la razón de los promedios.

De este modo, se obtiene una medida en la cual el 1 indica paridad entre un operativo y otro. Los valores mayores a 1 indican un aumento en relación a la base, y los inferiores señalan descensos.

**Tabla 5.3.1.**

Diferencias de niveles de desempeño entre operativos de evaluación por jurisdicción. Aprender 2016-2019

Jurisdicción	1 Por debajo del nivel básico	2 Básico	3 Satisfactorio	4 Avanzado
2 Ciudad de Buenos Aires	0,619	0,975	1,136	0,958
6 Provincia de Buenos Aires	0,412	1,017	1,244	0,721
10 Provincia de Catamarca	0,086	0,623	1,727	0,560
14 Provincia de Córdoba	0,390	0,787	1,182	1,003
18 Provincia de Corrientes	0,269	0,580	1,469	1,295
22 Provincia del Chaco	0,406	0,902	1,524	0,298
30 Provincia de Entre Ríos	0,297	0,847	1,263	0,810
34 Provincia de Formosa	0,207	1,004	1,462	0,420
38 Provincia de Jujuy	0,181	0,625	1,418	1,052
42 Provincia de La Pampa	0,955	0,934	1,037	0,933
46 Provincia de La Rioja	0,324	0,744	1,368	1,555
50 Provincia de Mendoza	0,436	0,970	1,188	0,683
54 Provincia de Misiones	0,351	0,602	1,296	2,129
58 Provincia del Neuquén	0,485	1,489	1,218	0,146
62 Provincia de Río Negro	0,191	1,565	1,122	0,659
66 Provincia de Salta	0,222	1,284	1,151	0,650
70 Provincia de San Juan	0,373	0,991	1,235	0,639
74 Provincia de San Luis	0,520	0,961	1,172	0,763
78 Provincia de Santa Cruz	0,104	0,853	1,253	0,825
82 Provincia de Santa Fe	0,431	1,148	1,114	0,880
86 Provincia de Santiago del Estero	0,255	0,867	1,385	1,718
90 Provincia de Tucumán	0,393	1,010	1,301	0,694
94 Provincia de Tierra del Fuego, Antártida e Islas del Atlántico Sur	0,223	1,196	1,282	0,529

\*Provincia de Chubut. Sin Datos disponibles para Aprender 2019.

**Fuente:** Evaluaciones de Ciencias Naturales. Aprender 2016 y Aprender 2019. DNIE. Ministerio de Educación de la Nación.

**Tabla 5.3.2.**

Medidas de diferencia de niveles de desempeño jurisdiccionales entre operativos de evaluación. Media, valores mínimos y máximos. Aprender 2016-2019

	1 Por debajo del nivel básico	2 Básico	3 Satisfactorio	4 Avanzado
Media	0,354	0,955	1,285	0,866
Mínimos	0,086	0,580	1,037	0,146
Máximos	0,955	1,565	1,727	2,129

**Fuente:** Procesamientos propios en base a Evaluaciones de Ciencias Naturales. Aprender 2016 y Aprender 2019. DNIE. Ministerio de Educación de la Nación.

De estas tablas, se puede extraer la siguiente información para analizar:

### Promedios de diferencias entre relevamientos

En términos generales, se observa entre las jurisdicciones una tendencia al descenso del peso de la categoría 1 (Por debajo del nivel básico) mientras que la categoría 2 (Básico) se mantiene relativamente constante, con diferencias estadísticamente insignificantes. Por el contrario, se observa un incremento en las diferencias promedio jurisdiccionales entre relevamientos en la categoría 3 (Satisfactorio), con diferencias cercanas a un tercio (1.285). En cuanto al promedio de la categoría más alta de nivel de desempeño (Avanzado), también pierde algo de relevancia.

### Valores mínimos y máximos de diferencias entre relevamientos

La amplitud entre las jurisdicciones de menor variación y las de mayor variación es muy marcada. En ambas categorías rondan valores cercanos a 1.

También se pueden observar diferencias extremas en la categoría Avanzado que rondan casi los 2 puntos.

Por el contrario, las diferencias entre valores mínimos y máximos en la categoría Satisfactorio son relativamente reducidas. Esto induce a pensar que las diferencias en el resto de las categorías pueden deberse a una menor concentración de casos. Este elemento se retomará al momento de formular hipótesis orientativas para la interpretación.

### Jurisdicciones con diferencias negativas extremas: las categorías *Por debajo del nivel básico* y *Básico*

En las seis (6) jurisdicciones en las que se produjo un descenso llamativo en la categoría 1 Por debajo del nivel básico, se observa un fenómeno similar que refiere a que las diferencias se corresponden con la tendencia observada a la concentración de casos en las categorías 2 *Básico*, y 3 *Satisfactorio*.

Advertencia:

Al ser un relevamiento muestral, varias jurisdicciones implementaron el operativo de evaluación 2019 en una cantidad muy limitada de establecimientos (entre 4 y 9). Tales son los casos de las provincias de La Pampa, Las Rioja, Catamarca, Provincia del Neuquén, Provincia de Río Negro, Santa Cruz, Salta, San Juan, San Luis, Santiago del Estero. Son jurisdicciones que tienen baja incidencia en el universo de matriculados en la educación secundaria por el tamaño de su población.

En este sentido, la información que provee este corte de los datos debe ser analizado con especial atención teniendo en cuenta este aspecto.

**Tabla 5.3.3.**

Jurisdicciones con mayor descenso de casos en el nivel de desempeño *Por debajo del nivel básico*. Aprender 2016-2019

	1 Por debajo del nivel básico	2 Básico	3 Satisfactorio	4 Avanzado
10 Provincia de Catamarca	0,086	0,623	1,727	0,560
78 Provincia de Santa Cruz	0,104	0,853	1,253	0,825
38 Provincia de Jujuy	0,181	0,625	1,418	1,052
62 Provincia de Río Negro	0,191	1,565	1,122	0,659
34 Provincia de Formosa	0,207	1,004	1,462	0,420
66 Provincia de Salta	0,222	1,284	1,151	0,650

**Fuente:** Evaluaciones de Ciencias Naturales. Aprender 2016 y Aprender 2019. DNIE. Ministerio de Educación de la Nación.

Entre las jurisdicciones que mayor descenso tuvieron en la categoría *Por debajo del nivel básico*, se observa que en general los casos se distribuyeron en las categorías de mejores niveles de desempeño, a excepción de la provincia de Catamarca, en la cual se observa un fuerte proceso de concentración de niveles de desempeño alrededor de la categoría 3 (*Satisfactorio*).

**Tabla 5.3.4.**

Jurisdicciones con mayor descenso de diferencias en el nivel de desempeño *Básico*. Aprender 2016-2019

	1 Por debajo del nivel básico	2 Básico	3 Satisfactorio	4 Avanzado
18 Provincia de Corrientes	0,269	0,580	1,469	1,295
54 Provincia de Misiones	0,351	0,602	1,296	2,129
10 Provincia de Catamarca	0,086	0,623	1,727	0,560
38 Provincia de Jujuy	0,181	0,625	1,418	1,052
46 Provincia de La Rioja	0,324	0,744	1,368	1,555
14 Provincia de Córdoba	0,390	0,787	1,182	1,003

**Fuente:** Evaluaciones de Ciencias Naturales. Aprender 2016 y Aprender 2019. DNIE. Ministerio de Educación de la Nación.

### Jurisdicciones con diferencias positivas extremas en niveles de desempeño *Satisfactorio* y *Avanzado*.

En las jurisdicciones con aumentos relativamente más importantes en las diferencias entre relevamientos en el nivel de desempeño *Satisfactorio*, se observa de manera amplificada la tendencia general de una distribución. Tal como fue mencionado anteriormente, las medidas de tendencia central de la distribución de los niveles de desempeño en Ciencias Naturales en Aprender 2019 muestran una concentración de casos precisamente en esta categoría (*Satisfactorio*).

De este modo, el caso de Catamarca vuelve a aparecer en el análisis inverso de diferencias extremas, es decir aquellas que más aumentan en la categoría 3.

**Tabla 5.3.5.**

Jurisdicciones con mayor aumento de diferencias en el nivel de desempeño *Satisfactorio*. Aprender 2016-2019

	1 Por debajo del nivel básico	2 Básico	3 Satisfactorio	4 Avanzado
10 Provincia de Catamarca	0,086	0,623	1,727	0,560
22 Provincia del Chaco	0,406	0,902	1,524	0,298
18 Provincia de Corrientes	0,269	0,580	1,469	1,295
34 Provincia de Formosa	0,207	1,004	1,462	0,420
38 Provincia de Jujuy	0,181	0,625	1,418	1,052
86 Provincia de Santiago del Estero	0,255	0,867	1,385	1,718

**Fuente:** Evaluaciones de Ciencias Naturales. Aprender 2016 y Aprender 2019. DNIE. Ministerio de Educación de la Nación.



En cuanto al nivel de desempeño Avanzado, en las jurisdicciones que tuvieron mayores diferencias entre operativos, se observa en todos los casos descensos relevantes en las categorías 1 y 2, y aumentos significativos también en la categoría 2.

**Tabla 5.3.6.**

Jurisdicciones con mayor aumento de diferencias en el nivel de desempeño *Avanzado*. Aprender 2016-2019

	1 Por debajo del nivel básico	2 Básico	3 Satisfactorio	4 Avanzado
54 Provincia de Misiones	0,351	0,602	1,296	2,129
86 Provincia de Santiago del Estero	0,255	0,867	1,385	1,718
46 Provincia de La Rioja	0,324	0,744	1,368	1,555
18 Provincia de Corrientes	0,269	0,580	1,469	1,295
38 Provincia de Jujuy	0,181	0,625	1,418	1,052
14 Provincia de Córdoba	0,390	0,787	1,182	1,003

**Fuente:** Evaluaciones de Ciencias Naturales. Aprender 2016 y Aprender 2019. DNIE. Ministerio de Educación de la Nación.

## 6. Evaluación estadística de ítems

### 6.1. Notas sobre el diseño de Modelos – Instrumentos de evaluación. <sup>4</sup>

En el proceso de diseño, los instrumentos de los operativos de evaluación atraviesan diferentes etapas. Se trata de seis modelos de evaluaciones diferentes que constan de veinticuatro (24) preguntas.

La evaluación consta de 6 bloques de 12 ítems cada uno y que cada modelo combina 2 de estos bloques.

Esta combinación de bloques y preguntas /ítems configura un total de 74 ítems, una parte de las cuales se repite de un operativo de evaluación a otro; éstos son los ítems llamados "Anclajes" y tienen por objetivo dar comparabilidad entre operativos de evaluación. La otra parte de los ítems varía de un operativo de evaluación a otro, de manera de que una parte de las preguntas no se repitan, estos ítems se denominan "No Anclados".

Existen cuatro modelos diferentes de evaluaciones con distribuciones distintas de ítems anclados y no anclados.

<sup>4</sup> Función de información para los ítems de Ciencias Naturales. Curva estandarizada con promedio 0 y desvío estándar igual a 1. 2019. Informe metodológico aprender. Aprender 2019. Equipo de Metodología. DINIEE, Ministerio de Educación.

## 6.2. Metodología de comparación

A partir de esta estructura de modelos / instrumentos de evaluación se analizará el comportamiento de los ítems para observar en qué medida pueden afectar la distribución de los niveles de desempeño.

Tal como fue señalado en la introducción, escapa al alcance de este informe:

- a) La evaluación del modelo teórico de asignación estadística de niveles de desempeño.
- b) El análisis de procedimientos teóricos y metodológicos de asignación de niveles de dificultad de ítems.
- c) Los procesos de evaluación de calidad y de consistencia de los datos.

Esto es así en tanto se entiende que los modelos están sujetos a estándares internacionales de comparabilidad, que han sido implementados en varias ocasiones y han permitido construir series de datos consistentes.

En relación a los ítems, aquellos que son anclados resultan una constante importante ya que han sido probados y demostraron operar como indicadores eficaces para medir el nivel de desempeño; además, brindan parámetros de comparabilidad con aquellos ítems "nuevos".

En este sentido, se han implementado procesos de evaluación pedagógica que han permitido descartar ítems que evalúan contenidos que, por cuestiones curriculares o de gestión educativa, no han podido ser dictados<sup>5</sup>.

Y para finalizar, se han implementado procedimientos de evaluación de la calidad y consistencia de los datos que han permitido identificar aquellos ítems que por razones pedagógicas o de diseño han presentado problemas; dichos ítems fueron identificados y eventualmente excluidos en el análisis de resultados del operativo<sup>6</sup>.

En este apartado, se observará si la cantidad de respuestas correctas de los estudiantes mantienen una distribución uniforme por el tipo de ítem. En concreto, se trata de aislar los ítems anclados de los no anclados para observar la existencia de distribuciones no esperables en los niveles de desempeño.

La metodología utilizada es la selección de los modelos que contienen igual cantidad de ítems anclados y no anclados. El objetivo es tener un marco de comparabilidad entre

---

<sup>5</sup> 2019. Informe de evaluación de Ítems. Ciencias Naturales. Aprender 2019. Equipo de Metodología. DINIEE, Ministerio de Educación.

<sup>6</sup> Informe de evaluación de Ítems. Ciencias Naturales. Aprender 2019. Equipo de Metodología. DINIEE, Ministerio de Educación.

relevamientos, provista por los ítems anclados, y así observar:

- a) Si el comportamiento entre ítems anclados es análogo en ambos operativos
- b) Si existen diferencias de distribución entre ítems anclados y no anclados.

Dado que todos los modelos tienen 24 preguntas, se seleccionaron los modelos 2 y 4 de ambos operativos de evaluación que tienen 12 preguntas ancladas y 12 no ancladas. En este sentido, el puntaje máximo es 24 y el puntaje mínimo (ninguna respuesta correcta) es cero (0).

El indicador utilizado es la sumatoria de respuestas correctas según tipo de ítem<sup>7</sup> (anclado o no anclado). Estas sumatorias, tienden a alinearse con la distribución de los niveles de desempeño. Aceptamos en este enfoque que existe una leve varianza producto del balanceo intencional del nivel de dificultad del ítem producto del balanceo intencional del nivel de dificultad de los ítems en la construcción del test. Lo central de esta verificación es la simetría en la distribución de los puntajes entre ítems anclados y no anclados.

### **6.3. Cantidad de respuestas correctas a ítems: estadísticos y medidas de tendencia central**

Al analizar los estadísticos se observa que las medias de los puntajes según tipo de ítem no presentan diferencias relevantes. De todos modos, en los ítems no anclados la media presenta mayores diferencias entre operativos de evaluación; en promedio 5.86 ítems correctos sobre 12 posibles en 2019, mientras que en 2016 el promedio era cercano a 5 (5.18).

La medida que llama la atención es que la varianza en los ítems anclados de 2016 es muy superior a la de los no Anclados del mismo año y a los dos tipos de ítems en 2019. Esta diferencia en la varianza refleja en parte el fenómeno que da origen a este informe.

En cuanto a los estadísticos de 2019 indicarían niveles de dificultad similares en ambos tipos de ítems ya que anclados y no anclados presentan medidas de tendencia central y dispersión bastante homogéneas.

---

<sup>7</sup> Esta aproximación recurre a un método de análisis más propio de la teoría clásica del test. Hemos optado por trabajar con las cantidades de respuestas correctas como método alternativo de verificación de la distribución. La distribución de la sumatoria de respuestas coincide con la distribución de la función de información del test antes de la ponderación. Además es la forma de observar el comportamiento del alumno con relativa independencia del nivel de dificultad del ítem.

**Tabla 6.3.1.**

Media y medidas de dispersión de cantidad de respuestas correctas por tipo de ítem según operativo de evaluación. Aprender 2016-2019

	2016		2019	
	Anclados	No Anclados	Anclados	No Anclados
Media	6,07	5,18	5,81	5,86
Desviación estándar	2,652	2,14	2,337	2,249
Varianza	7,034	4,581	5,459	5,056
Curtosis	-0,436	-0,392	-0,471	-0,182

**Fuente:** Evaluaciones de Ciencias Naturales. Aprender 2016 y Aprender 2019. DNIE. Ministerio de Educación de la Nación.

## 6.4. Comparación de distribuciones cantidad de aciertos por niveles de desempeño según tipo de ítem.

Al analizar las distribuciones por niveles de desempeño podemos observar los siguientes elementos:

### Al interior de cada operativo

Si se comparan los ítems anclados y no anclados al interior de cada relevamiento, se observa que aunque presentan algunas diferencias tienen parámetros estadísticos similares.

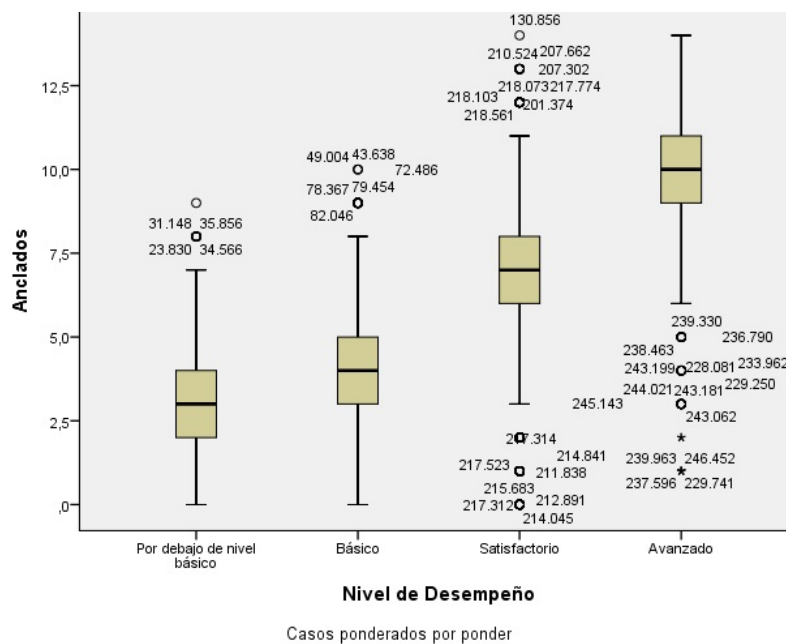
De las distribuciones de 2016 puede decirse que los ítems no anclados aparentan tener una menor media de respuestas correctas para el caso de las y los alumnos que están en la categoría *Por debajo del nivel básico*; en tanto, la media de respuestas correctas está cerca del límite inferior del primer desvío.

En cuanto al operativo de 2019 los ítems anclados aparecen con una menor cantidad promedio de respuestas correctas. La media de la categoría *Por debajo del nivel básico* tiende a ser baja, mientras que resultan llamativamente altas las medias – tanto de anclados como no anclados - de respuestas correctas en el nivel de desempeño *Avanzado*.

### Gráfico 6.4.1.

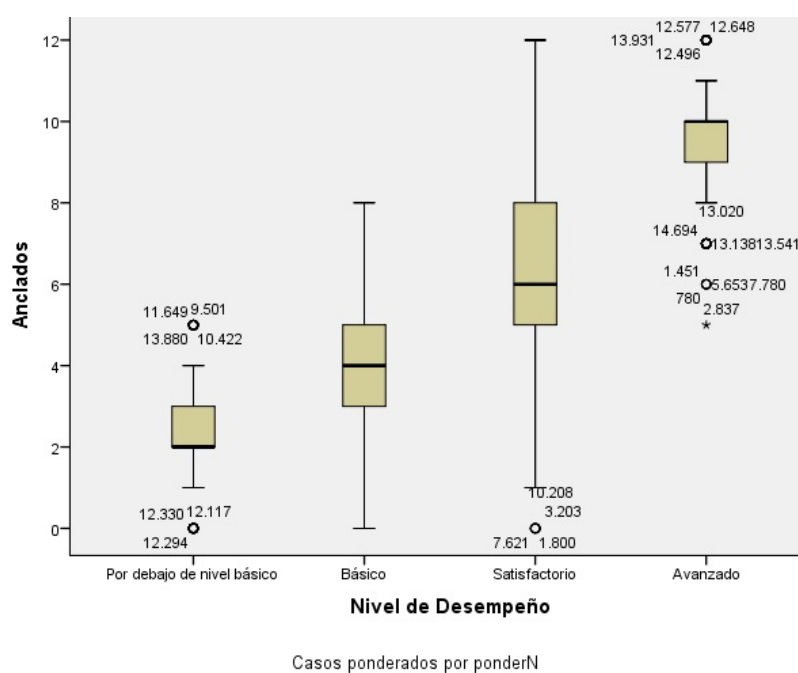
Boxplots de medidas de dispersión de la cantidad de aciertos por tipo de ítem y nivel de desempeño según operativo de evaluación.

#### Ítems anclados 2016



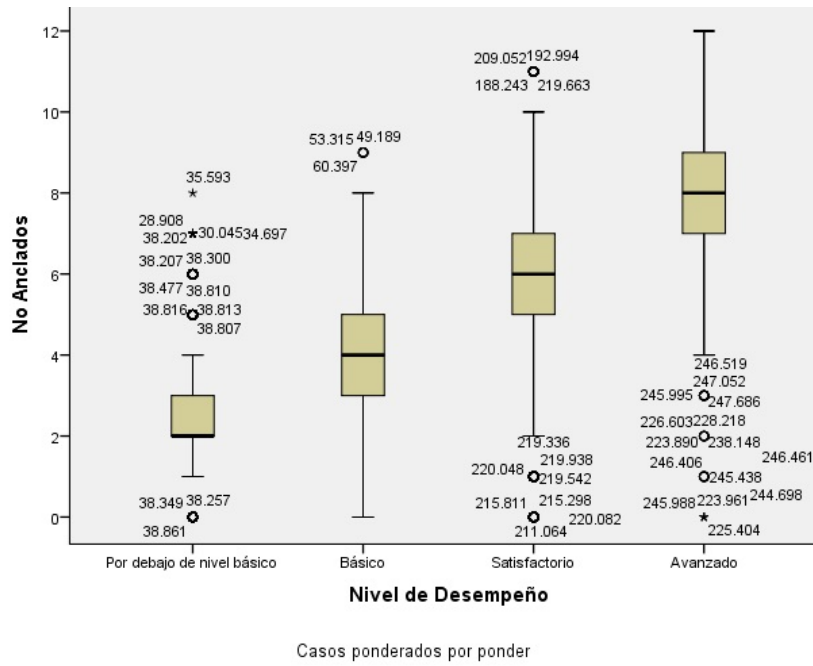
Fuente: Evaluaciones de Ciencias Naturales. Aprender 2016 y Aprender 2019. DNIE. Ministerio de Educación de la Nación.

#### Ítems anclados 2019



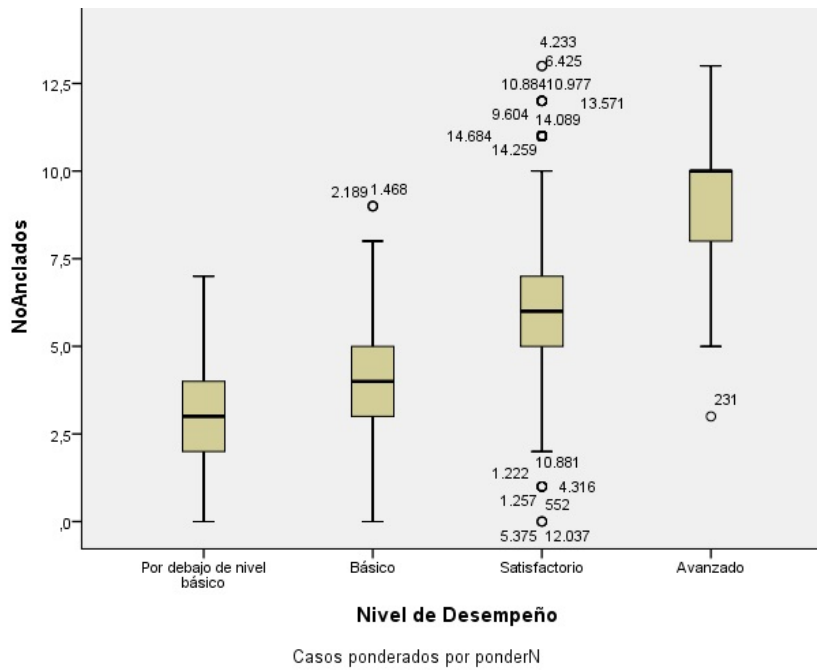
Fuente: Evaluaciones de Ciencias Naturales. Aprender 2016 y Aprender 2019. DNIE. Ministerio de Educación de la Nación.

### Ítems no anclados 2016



Fuente: Evaluaciones de Ciencias Naturales. Aprender 2016 y Aprender 2019. DNIE. Ministerio de Educación de la Nación.

### Ítems no anclados 2019



Fuente: Evaluaciones de Ciencias Naturales. Aprender 2016 y Aprender 2019. DNIE. Ministerio de Educación de la Nación.

## Ítems no anclados

En cuanto a los ítems no anclados se observa un fenómeno diferente con respecto a los anclados.

En primer lugar, una menor presencia de outliers. En cuanto a las medias, resultan más altas en las categorías extremas (Por debajo del nivel básico y Avanzado) y más bajas en las categorías Básico y Satisfactorio.

Es esperable que la distribución de respuestas correctas en los ítems no anclados tienda a la normal, ya que:

- a) Son ítems testeados en un período y una población coincidente con la población objetivo del operativo.
- b) Son ítems que, de no discriminar eficazmente en el nivel de dificultad previsto<sup>8</sup>, pueden ser descartados sin afectar ni la cantidad ni la calidad de la información provista por el relevamiento.

En este sentido, la distribución de los ítems no anclados no difiere de la distribución esperada ya que fueron seleccionados para producir este resultado.

## Ítems anclados

En relación a los ítems anclados, se observa una menor presencia de valores outliers o atípicos en Aprender 2019. Se trata de fenómenos esperables en un contexto de una distribución con menor varianza.

En los boxplots de gráfico 6<sup>9</sup>, se puede observar cómo la menor varianza de 2019 implica una distribución concentrada de registros dentro de los 2 desvíos en las categorías Básico y Satisfactorio, sin generar prácticamente outliers. Mientras que en 2016, presentaba una distribución más similar a una curva normal.

---

<sup>8</sup> Cada uno de los ítems está diseñado para medir un nivel de dificultad específico. Si cada ítem tiene 4 opciones de respuesta efectiva (más la respuesta múltiple – error- y la no respuesta) la probabilidad por azar simple es de  $\frac{1}{4}$  (o 25%). Esta probabilidad varía de acuerdo a mayor sea el nivel de desempeño del estudiante. Es decir, que a un nivel de desempeño alto, la probabilidad estadística de respuesta correcta tiende a 100%, mientras que en los de nivel de desempeño Por debajo del nivel básico tiende a cero. En tanto cada ítem tiene un nivel de dificultad específico, la probabilidad varía según el criterio de: a menor dificultad, mayor probabilidad de respuesta correcta, y viceversa.

<sup>9</sup> En el gráfico mencionado, los ejes Y indican la cantidad de respuestas correctas promedio para cada nivel de desempeño.

Las medias en 2019 tienden a ser ligeramente más bajas que en 2016 pero la varianza mucho menor.

En un sentido, resulta contra intuitivo que sean los ítems anclados los que tienden a mostrar una menor dispersión en cantidad de respuestas correctas, mientras que los no anclados son los que tienen una mayor varianza.

Los anclados han atravesado diferentes instancias de testeo:

- De diseño y testeo teórico por parte de los especialistas.
- Test piloto empírico en una muestra.
- Tests sucesivos en los diferentes operativos de evaluación en los que han sido incluidos.

En este sentido, la distribución observada pareciera no surgir de problemas de la capacidad de discriminación de niveles de desempeño de los ítems, sino ser emergentes de otros procesos.

Si el “instrumento de medición” está probado como fiable, deberíamos observar si existen en la población objetivo elementos que distorsionen de alguna manera los resultados.

## **7. Hipótesis orientativas para la interpretación de la distribución de los niveles de desempeño en Ciencias Naturales Aprender 2019.**

De acuerdo al plan de análisis planteado, se han realizado dos tipos de análisis estadísticos de los resultados de las evaluaciones de Ciencias Naturales – Aprender 2019.

El primero de ellos se centró en la observación de las distribuciones con variables de corte, es decir, observando los resultados al dividir el universo por algún atributo que permita la construcción de subconjuntos relevantes a ser comparados y, de ese modo, descartar factores ajenos al operativo que estén alterando significativamente los resultados.

En este primer análisis se pudo observar que las distribuciones por variables de corte muestran una tendencia a mantener las características de las distribuciones generales. No obstante, las distribuciones difieren en un cuanto grado en cada subuniverso.

En términos generales, en aquellos subuniversos relativamente más grandes y/u homogéneos, las diferencias tienden a ser menos marcadas. Por el contrario, en aquellos subuniversos (o sub-subuniversos) más pequeños o más heterogéneos en su interior, las diferencias entre las distribuciones son muy relevantes. Estas características generales se pueden observar claramente en la segmentación por ámbito y por sector de gestión, pero también en el análisis de jurisdicciones que son más homogéneas (caso CABA) versus aquellas más



heterogéneas (caso Santa Fe).

En cuanto a las diferencias de promedios, valores mínimos y valores máximos jurisdiccionales, tal como se ha señalado en el apartado específico, se observaron variaciones que parecieran estar originadas en sesgos de la muestra de cada una de las jurisdicciones<sup>10</sup>.

Cuando se comparan los casos por sector de gestión se observa una tendencia más pronunciada para el sector estatal. Cuando el corte se hace por ámbito, es el rural el que muestra una mayor concentración de casos en las categorías *Satisfactorio*.

Cuando se analizan las diferencias jurisdiccionales, resulta bastante notorio que las jurisdicciones que presentan variaciones (positivas y negativas) más extremas son aquellas en las que el marco muestral le asignó un número limitado de establecimientos.

Es notorio observar que las jurisdicciones que tuvieron resultados con distribuciones de los niveles de desempeño con menor variación entre relevamientos, son aquellas que:

- a) constituyen espacios sociales relativamente más homogéneos – como es el caso de la Ciudad de Buenos Aires - y/o
- b) cuentan con una mayor cantidad de establecimientos educativos seleccionados para la participación en el operativo de evaluación como es el caso de provincias grandes como Buenos Aires y Córdoba.

Esta situación hace desaconsejable la presentación de resultados con un nivel de desagregación jurisdiccional en tanto tenderá a amplificar los sesgos observados en la distribución general.

En la segunda parte de la evaluación estadística de los resultados de las Evaluaciones de Ciencias Naturales - Aprender 2019, el eje estuvo puesto en verificar que los resultados no se expliquen por problemas de diseño al nivel de los modelos o de los ítems. En este sentido se corroboró que aquellos ítems que estaban ajustados estadísticamente (los no anclados) mantenían una distribución similar a los de 2016, mientras que los anclados mostraban una distribución no esperada.

Si bien no es descartable, resulta poco probable que el origen de la distribución observada se origine en problemas de medición de un grupo de ítems.

Esto es así porque el período que separa los operativos de evaluación es lo suficientemente breve como para que se hayan operado cambios sustanciales en términos sociales o de los procesos pedagógicos que expliquen tal diferencia. De ser así, también sería esperable que estas “nuevas distribuciones” se hubieran manifestado de manera coincidente en aquellas evaluaciones que se realizaron de manera censal.

---

<sup>10</sup> Este tema se retomará de manera general en el apartado “El Marco Muestral”.

Tampoco se ha constatado el desarrollo de políticas nacionales tendientes a intensificar los conocimientos de las y los alumnos de la escuela secundaria en esta disciplina (Ciencias Naturales) que pueda explicar este proceso de mejora en los niveles de desempeño.

En principio, resulta conveniente descartar la hipótesis de caducidad de los instrumentos por cambios significativos en los niveles de desempeño y, por lo tanto, poner en debate el diseño de los ítems.

En cierto sentido, también se observó que la metodología de evaluación de los ítems permite identificar con exactitud aquellos ítems que no sean capaces de medir con precisión el nivel de dificultad para el que fueron diseñados; esta precisión se constata en que los ítems no anclados mantienen una distribución normal a partir de su ajuste a los niveles de desempeño "históricos" de los operativos de evaluación anteriores.

## 7.1. El marco muestral

Según este análisis, una parte del fenómeno de mejora en los niveles de desempeño observado como reducción de la varianza, concentración de casos alrededor de la mediana y un aumento significativo del peso de la moda -nivel de desempeño *Satisfactorio*- podría estar influido por el modo en que quedó conformado efectivamente el diseño muestral.

La desventaja de toda muestra, es que siempre presenta errores debidos al proceso de muestreo que, en otras palabras, son los que surgen como consecuencia de observar parcialmente a la población. Excepto que la muestra coincida con la población, este error no se anula, aunque sí puede medirse.

El modo en que se lleva adelante un operativo de evaluación nacional es un proceso complejo en el cual intervienen actores institucionales de diferentes niveles, que desarrollan sus actividades en contextos heterogéneos tanto en lo económico, lo social y hasta en lo geográfico, y esto en cada una de las etapas de planificación e implementación. Puntualmente, cada una de las etapas en el proceso de implementación de operativos de evaluación implica presuponer que en su desarrollo intervienen aspectos políticos, económico-presupuestarios, logísticos y administrativos que podrían introducir sesgos en el proceso general del operativo.

En este sentido, la realización de relevamientos muestrales implica diversas instancias de toma de decisiones en las que influyen todas estas dimensiones sin que los actores tomen conciencia plena de los efectos. En ese contexto, las decisiones metodológicas y de técnicas de diseño de la muestra, pueden verse constreñidas de manera significativa y hasta orientadas por las circunstancias institucionales, las limitaciones temporales para resolver problemas o los límites materiales para lograr una solución óptima.

Asimismo, una vez diseñada la selección de establecimientos, pueden intervenir otros factores que afecten su representatividad. Es posible que hayan intervenido en algunas instancias del proceso de diseño, implementación y/o ajuste de la muestra, alguno o varios sesgos que hayan terminado por afectar los resultados.

Este análisis desconoce los criterios para los procesos de selección de unidades de observación así como la metodología de sustitución. Parece probable que los límites en la selección de unidades de observación (por temas presupuestarios o logísticos, entre otros) haya obligado a la implementación del esquema de ponderación con rangos tan amplios y, por lo tanto, pesos relativos tan disímiles.

Es importante remarcar que aunque el problema puede manifestarse como un problema de la muestra, no necesariamente implica que ese sesgo esté originado en el diseño. Tal como se ha señalado, es posible que en las distintas etapas del proceso del operativo de evaluación se hayan tomado decisiones que hayan incidido en la naturaleza de la muestra. Eso, por supuesto escapa al análisis meramente estadístico.

## 8. Anexo técnico

El anexo técnico es una selección de capítulos del Informe Metodológico elaborado por la DINIEE. Lo incorporamos con el objetivo de documentar técnicamente algunas de las decisiones estadísticas definidas para el operativo Aprender 2019.

### **Informe metodológico Aprender 2019. Equipo de Metodología DINIEE.**

#### **Cap. 2. Criterios para la depuración de la base de datos**

Con el objetivo de mejorar la calidad de las estimaciones, se determinaron diferentes criterios de depuración para la base de datos. Se eliminaron del proceso de generación de puntajes TRI los siguientes casos:

*a) Cuestionarios de estudiantes que contestaron menos del 50% de las preguntas, debido a que no aportan información suficiente para estimar su habilidad con un nivel aceptable de precisión. Hay que recordar que uno de los objetivos de Aprender es producir información a nivel de escuela, por lo que se deben tomar recaudos para que esta información sea confiable.*

*b) Cuestionarios de estudiantes que siguen patrones de respuesta atípicos o aberrantes, es decir, que contestaron las preguntas en forma aleatoria o que intencionalmente contestaron mal varios ítems a pesar de su habilidad.*

## Ficha técnica de la evaluación de Ciencias Naturales.

Quinto/Sexto año Secundaria - Cs Naturales:

Cantidad de alumnos presentes en materia	15161
Cantidad de alumnos eliminados por menos del 50% de respuesta	165
Porcentaje de alumnos eliminados por menos del 50% de respuesta	1.09
Cantidad de alumnos eliminados por patrones aleatorios de respuesta	202
Porcentaje de alumnos eliminados por patrones aleatorios de respuesta	1.35
Porcentaje de alumnos eliminados en total	2.42
Cantidad de alumnos puntuados en materia	14794

### Cap. 4 tratamiento de valores faltantes en las pruebas

A lo largo del cuestionario de un determinado estudiante pueden encontrarse respuestas en blanco (missing) que pueden deberse a distintos motivos. Hay varias alternativas para tratar las respuestas faltantes a los ítems. Algunas son:

- a) *Un ítem en blanco está más cerca de una respuesta incorrecta que de una correcta y recomiendan considerarlo como respuesta incorrecta.*
- b) *Dejarlos como missing para el cálculo de los parámetros de los ítems, pero considerarlos como respuesta incorrecta para el cálculo del puntaje del estudiante.*
- c) *Dejarlos como missing tanto para el cálculo de los parámetros como para el cálculo de los puntajes.*
- d) *Imputarlos y estimar los parámetros y puntajes con el valor imputado.*

En Aprender 2019 se optó por esta última alternativa. La imputación se hizo con el módulo mi del Stata, que realiza en una de sus opciones una imputación multivariada, apoyándose en los ítems respondidos, mediante una regresión logística. Otra referencia para este método de imputación es el software IVEware. Recordar que en esta etapa ya se trabaja con un cuestionario que tiene al menos el 50% de respuestas a los ítems.

Al aplicar la imputación de respuestas, hay que diferenciar los casos en los cuales un ítem no fue respondido con la intención de volver luego a él, de aquellos en los cuales los ítems son dejados en blanco al final del cuestionario ya que es posible suponer que la o el estudiante se cansó, siendo éste el motivo de la no respuesta de los ítems finales. Por lo tanto, no se toman en cuenta en el proceso de imputación, excepto que haya menos de 5 ítems en blanco, en cada bloque al final del cuestionario.

## Fuentes de información

2016. Base de Datos de Ciencias Naturales. Aprender 2016. DINIEE. Ministerio de Educación.

2019. Base de Datos de Ciencias Naturales. Aprender 2019. DINIEE. Ministerio de Educación.

2019. Tabla de Anclaje de Ítems 2016-2019.

2019. Informe de evaluación de Ítems. Ciencias Naturales. Aprender 2019. Equipo de Metodología. DINIEE, Ministerio de Educación.

2019. Informe metodológico aprender. Aprender 2019. Equipo de Metodología. DINIEE, Ministerio de Educación.

2019. Evaluación de la Educación Secundaria 2019. Secretaría de Información y Estadística Educativa. Ministerio de Educación.

**Director Nacional de Evaluación, Información y Estadística Educativa**

Tomas Ciocci Pardo

**Directora de Evaluación Educativa**

Paula Viotti

**Elaboración del informe**

Flavio Guberman

**Edición y corrección**

Mercedes Mac Donnell, Florencia Carballido, Laura Melchiorre, Quimey Lassus Frusso,  
Flavia Petrini

**Diseño editorial**

Cecilia Román Ulloa