



SERIE DE DOCUMENTOS TÉCNICOS/ 2

APRENDER 2016

NOTAS

METODOLÓGICAS



SERIE DE DOCUMENTOS TÉCNICOS/ 2

APRENDER 2016

NOTAS METODOLÓGICAS

AUTORIDADES

Presidente

Ing. Mauricio Macri

Ministro de Educación y Deportes

Lic. Esteban Bullrich

Jefe de Gabinete del Ministerio de Educación y Deportes

Dr. Diego Marias

Secretaria de Evaluación Educativa

Prof. Elena Duro

Secretario de Gestión Educativa

Lic. Maximiliano Gullmanelli

Secretario de Políticas Universitarias

Dr. Albor Cantard

Secretaria de Innovación y Calidad

Sra. María de las Mercedes Miguel

Secretario de Deportes, Educación Física y Recreación

Sr. Carlos Mac Allister

ÍNDICE

PRÓLOGO	9
INTRODUCCIÓN	11
1. CONSIDERACIONES INICIALES	13
2. ANTECEDENTES	14
3. CARACTERÍSTICAS	16
3.1 Pruebas referidas a criterio y Teoría de Respuesta al Ítem (TRI)	17
3.2 Marco de referencia y elaboración de ítems	18
3.3 Prueba piloto 2015	26
3.4 Ajustes 2016	26
4. CUESTIONARIO DE FACTORES ASOCIADOS	27
5. UNIVERSO Y MUESTRA DE ESTUDIANTES DE APRENDER 2016	33
5.1 Universo de Aprender 2016	33
5.2 Muestra de Aprender 2016	34
6. NIVELES DE DESEMPEÑO DE LOS ESTUDIANTES	36
6.1 Identificación de puntos de corte	36
6.2 Redefinición de los niveles de desempeño	37
7. ESCALAMIENTO Y COMPARABILIDAD ENTRE LOS PUNTAJES TRI DE ONE Y APRENDER 2016	38
8. CRITERIOS DE ORGANIZACIÓN DE LA APLICACIÓN TERRITORIAL: APRENDER 2016 Y OBJETIVO DEL CAMBIO	40
9. PROCESAMIENTO DE DATOS DE APRENDER 2016	42
9.1 Diseño de la base y carga de datos	42
9.2 Criterios para la depuración de la base de datos	42
9.2.1 Tratamiento de valores faltantes en las pruebas	43
9.2.2 Tratamiento de la no respuesta en las preguntas de contexto	46
10. ANÁLISIS PSICOMÉTRICO DE LAS PRUEBAS	47
11. REFERENCIAS	53

PRÓLOGO

La educación tiene un rol central en el desarrollo social y económico del país. Es nuestra obligación como funcionarios del Estado mejorar de forma continua los niveles de equidad y la calidad del sistema educativo. Para lograr estos objetivos, es fundamental contar con información confiable que guíe la toma de mejores decisiones en los distintos niveles de gestión.

Aprender 2016 generó un conjunto robusto de datos que permite tener una mirada más aguda sobre las fortalezas y desafíos que tenemos por delante. La evaluación por sí sola no corrige los problemas, pero nos brinda valiosa información para profundizar nuestra mirada y continuar encarando un cambio a través de acciones concretas para mejorar la calidad educativa en la Argentina.

Estamos llevando adelante una tarea que nos encomendó el Presidente: generar igualdad de oportunidades para aprender independientemente del lugar donde hayamos nacido.

La evaluación nacional de aprendizajes nos mostró grandes retos en distintas áreas del conocimiento. Al mismo tiempo, nos permitió visualizar un conjunto de escuelas con buenos niveles de desempeño pese a encontrarse en las áreas más vulnerables de nuestra sociedad. Es útil detectar qué cosas funcionan bien para aprovechar todas las experiencias educativas positivas que suceden en nuestro país.

Para concretar semejante tarea es importante destacar que fue invaluable el profundo compromiso demostrado por todos los ministros del país, los equipos de cada provincia, los directores que actuaron como veedores, los docentes aplicadores, los estudiantes y sus familias.

A partir de la información relevada en los resultados de Aprender, estamos trabajando con todas las provincias en políticas de apoyo a las escuelas con niveles de desempeño más bajos con la mirada puesta en la mejora. Es por eso que presentamos el Plan Maestr@, con metas claras para pensar la educación en la actualidad y en el largo plazo.

Con la transparencia que nos caracteriza, compartimos los resultados con la sociedad. Creemos que estos informes alentarán un debate permanente que alimente la revolución educativa que llevamos adelante. Estamos convencidos de que la educación nos une y es una herramienta que contribuye a construir el futuro que todos anhelamos para la Argentina.

La educación nos une.

Lic. Esteban Bullrich
Ministro de Educación y Deportes

INTRODUCCIÓN

La educación es un derecho y constituye una política central que habilita a ciudadanías plenas, abre puertas al ejercicio de otros derechos y posibilita el desarrollo y crecimiento de la sociedad. La generación de evidencia sólida y confiable sobre el estado de situación de la educación en nuestro país es un factor necesario -aunque no suficiente- para reorientar políticas y prácticas de cara a procesos de mejora educativa continua. El objetivo es aportar a garantizar una educación de calidad para todos.

La implementación de Aprender 2016 en más de 28 mil escuelas de todo el país, la participación de más de 950 mil estudiantes en esta evaluación, el alto compromiso que han manifestado las autoridades educativas jurisdiccionales, docentes, estudiantes y familias permite hoy devolver información valiosa, en tiempo y forma, a cada una de las escuelas participantes, a los decisores y a la sociedad. Esta información se constituye en un insumo fundamental para conocer y analizar ciertos aprendizajes prioritarios adquiridos por los estudiantes, notas sobre el contexto en el que los estudiantes aprenden y algunos factores que pueden incidir en sus procesos de aprendizaje.

La evaluación educativa es un recorte de una realidad mucho más compleja y por sí sola no mejora los aprendizajes, pero la información que genera, con apoyo y asesoramiento para su uso, se constituye en una herramienta potente a la hora de reflexionar en torno a las prácticas pedagógicas e institucionales, incita a elaborar nuevas preguntas, así como a redireccionar o elaborar prácticas y políticas con el objetivo de mejorar la educación.

La información entorno a Aprender 2016 pone en evidencia importantes desafíos que como sistema educativo hay que enfrentar y superar. Los datos nos muestran que aún es insuficiente la cantidad de estudiantes cuyos resultados se ubican en los niveles de desempeño satisfactorio y avanzado en todas las áreas. Por otro lado, prevalecen altos niveles de inequidad y desigualdad educativa a nivel nacional y jurisdiccional.

Al mismo tiempo, Aprender 2016 pone sobre la mesa la importancia de la educación como herramienta para alcanzar mayor justicia social. Los resultados muestran cómo todos los estudiantes pueden lograr mejores desempeños, incluso en los contextos más desaventajados, evidenciando el valor agregado que la escuela aporta a los niños y adolescentes de nuestro sistema educativo. Ese aporte es el que se pretende destacar y potenciar a partir de la devolución de los resultados de la evaluación.

El presente documento es parte de un conjunto de informes que incluye: reportes de resultados a nivel nacional y jurisdiccional, informes técnicos y temáticos. Por primera vez, las 24 jurisdicciones del país recibieron, luego de un operativo de esta magnitud, sus informes de resultado a cinco meses de la implementación del dispositivo de evaluación nacional.

Por otro lado, la totalidad de escuelas participantes de Aprender recibe un informe de resultados que hace foco en el potencial de mejora de cada una.

La finalidad de la información de la evaluación Aprender es poner a la evidencia al servicio de las escuelas, las autoridades, la academia, la comunidad educativa y aportar al debate educativo con miras a traccionar mejoras en la calidad y equidad de la educación argentina.

Prof. Elena Duro
Secretaria de Evaluación Educativa

En el siguiente documento se presenta información técnica sobre la prueba Aprender con la finalidad de contribuir a una mejor comprensión del proceso.

Se presentan conceptos que explican las características del proceso de construcción, la cobertura de la prueba Aprender y la conformación de la muestra de estudiantes evaluados. Se introduce también la metodología utilizada para establecer los puntajes y la definición de los niveles de desempeño. Además, se aborda la comparabilidad entre los resultados de la prueba ONE 2013 y Aprender 2016. La última sección se dedica a la cuestión del análisis y el procesamiento de los resultados.

1. CONSIDERACIONES INICIALES

Aprender es la evaluación nacional de carácter estandarizado que permite medir el nivel de desempeño de los estudiantes de nivel primario y secundario de todo el país en las áreas básicas de conocimiento, así como identificar distintos factores asociados que inciden en los aprendizajes.

Aprender produce información robusta y confiable sobre los aprendizajes de los estudiantes siguiendo las definiciones curriculares vigentes a nivel nacional, provincial y de la Ciudad Autónoma de Buenos Aires. El propósito de esta evaluación es constituirse en una potente herramienta para el proceso de mejora continua de la educación, fortaleciendo la toma de decisiones basada en evidencia, para contribuir a la disminución de las desigualdades en las oportunidades de aprendizaje.

2. ANTECEDENTES

En 1993, el Ministerio de Educación de la Nación (MEN) comenzó a implementar los Operativos Nacionales de Evaluación (ONE). Entre 1993 y 2002, los ONE fueron de carácter anual; de 2002 al 2010 fueron bienales; y a partir de 2010 comenzaron a ser trienales.

La Dirección Nacional de Información y Evaluación de la Calidad Educativa (DINIECE), en forma conjunta con las 23 provincias y la Ciudad Autónoma de Buenos Aires, tuvo a su cargo la implementación de los operativos hasta 2015. Los ONE evaluaron a los estudiantes de distintos grados de primaria y años de secundaria

Los operativos consistieron en la administración de pruebas de Lengua, Matemática, Ciencias Sociales y Ciencias Naturales, que se fundamentaron en un principio en los Contenidos Básicos Comunes (CBC) y luego en los Núcleos de Aprendizaje Prioritarios (NAP), los diseños curriculares de cada jurisdicción, los acuerdos con las jurisdicciones, los resultados de estudios piloto previos y la literatura específica relativa a los dominios/temas evaluados.

A su vez, en la mayoría de las instancias, se administraron cuestionarios complementarios cuyo objetivo fue contar con información sobre factores escolares (historia escolar, recursos) y extraescolares (nivel socioeconómico, nivel educativo de las familias) que permitiesen evaluar su asociación con el desempeño de los estudiantes. Estos cuestionarios fueron aplicados a estudiantes, docentes y directivos.

Tabla 1: Los ONE a través del tiempo¹.

LOS ONE A TRAVÉS DEL TIEMPO				
AÑO	NIVEL DE DATOS	NIVELES EDUCATIVOS	ÁREAS*	CUESTIONARIOS COMPLEMENTARIOS**
1993	Muestra	7º grado primaria 5º/6º año nivel medio	M y L M y L	A, DC y DR DC y DR
1994	Muestra	5º/6º año nivel medio 3º polimodal	M y L M y L	A, DC y DR A, DC y DR
	Muestra experimental	7º grado primaria / 7º año EGB 5º/6º año nivel medio 3º polimodal	CS y CN CS y CN CS y CN	
1995	Muestra	3º grado / 3º año EGB	M y L	A, DC y DR
		7º grado primaria / 7º año EGB	M, L, CS y CN	A, DC y DR
		2º año nivel medio	M y L	
		9º año EGB	M y L	
		5º/6º año nivel medio	M y L	
		3º polimodal	M y L	
1996	Muestra	3º grado / 3º año EGB	M y L	A, DC y DR
		6º grado / 6º año EGB	M, L, CS y CN	A, DC y DR
		7º grado primaria / 7º año EGB	M, L, CS y CN	A y DC
		2º año nivel medio	M y L	
		9º año EGB	M y L	DR
		5º/6º año nivel medio	M y L	

Referencias: * M = Matemática; L = Lengua; CS = Ciencias Sociales; CN = Ciencias Naturales; F = Física; Q = Química; H = Historia; G = Geografía; FEC = Formación ética y ciudadana.

** A = Alumno; DC = Docente; DR = Director; CC = Cuestionarios contextuales.

AÑO	NIVEL DE DATOS	NIVELES EDUCATIVOS	ÁREAS*	CUESTIONARIOS COMPLEMENTARIOS**
1997	Muestra	3° grado / 3° año EGB	M y L	A, DC y DR
		6° grado / 6° año EGB	M, L, CS y CN	A, DC y DR
		7° grado primaria / 7° año EGB	M, L, CS y CN	A, DC y DR
		2° año nivel medio	M y L	
		9° año EGB	M y L	
		5°/6° año nivel medio	M y L	
	Muestra experimental	5°/6° año nivel medio	CS y CN	DR
	Censo	5°/6° año nivel medio	M y L	DR
1998	Muestra	3° grado / 3° año EGB	M y L	A y DR
		6° grado / 6° año EGB	M, L, CS y CN	A y DC
		7° grado primaria / 7° año EGB	M, L, CS y CN	A, DC y DR
		2° año nivel medio	M y L	
		9° año EGB	M y L	
	Muestra experimental	5°/6° año nivel medio	F, Q, H y G	
	Censo	5°/6° año nivel medio	M y L	A y DR
1999	Muestra	3° grado / 3° año EGB	M y L	A, DC y DR
		6° grado / 6° año EGB	M, L, CS y CN	A, DC y DR
		7° grado primaria / 7° año EGB	M y L	A, DC y DR
		2° año nivel medio	M y L	
		9° año EGB	M y L	
	Censo	5°/6° año nivel medio	M y L	A y DR
	Muestra experimental	5°/6° año nivel medio	F, Q, H y G	
2000	Muestra	3° grado / 3° año EGB	M, L, CS y CN	
		2° año nivel medio	M y L	
		9° año EGB	M y L	
	Censo	6° grado / 6° año EGB	M, L, CS y CN	A, DC y DR
		5°/6° año nivel medio	M, L, F, Q, H y G	
2002	Muestra	3° grado / 3° año EGB	M y L	
		6° grado / 6° año EGB	M y L	A, DC y DR
		2° año nivel medio	M y L	
		9° año EGB	M y L	A y DR
		5°/6° año nivel medio	M y L	A, DC y DR
2005	Muestra	3° grado / 3° año EGB	M, L, CS y CN	
		6° grado / 6° año EGB	M, L, CS y CN	CC
		9° año EGB	M, L, CN y FEC	
		2° año nivel medio	M, L, CS y CN	
		5°/6° año nivel medio	M, L, CN y FEC	CC
2007	Muestra	3° grado / 3° año EGB	M, L, CS y CN	
		6° grado / 6° año EGB	M, L, CS y CN	
		2° año/ 3° nivel medio	M y L	
		5°/6° año nivel medio	M y L	DR
2010	Muestra	3° grado / 3° año EGB	M, L, CS y CN	CC
		6° grado / 6° año EGB	M, L, CS y CN	CC
		2° año/ 3° nivel medio	M, L, CS y CN	CC
	Censo	5°/6° año nivel medio	M, L, CS y CN	
2013	Muestra	3° grado / 3° año EGB	M, L, CS y CN	CC
		6° grado / 6° año EGB	M, L, CS y CN	CC
		2° año/ 3° nivel medio	M, L, CS y CN	CC
	Censo	5°/6° año nivel medio	M, L, CS y CN	

Fuente: DINIECE

3. CARACTERÍSTICAS DE LA EVALUACIÓN APRENDER

Aprender 2016 –al igual que los ONE– es una prueba estandarizada. Esto significa que se aplican los mismos instrumentos a todos los estudiantes del mismo grado/año en las mismas condiciones (día, horario, cantidad de tiempo para realizar la prueba) para luego ser evaluadas con los mismos criterios. Este tipo de evaluación se hace de acuerdo a normas que garantizan la homogeneidad del proceso evaluativo y de los resultados obtenidos.

En base a los resultados de la prueba piloto realizada por la DINIECE en marzo de 2015, la Secretaría de Evaluación Educativa (SEE) aplicó las evaluaciones a todos los estudiantes que asisten regularmente al último año de estudios de los niveles primario y secundario y a una muestra representativa de estudiantes que se encuentran en 3º año de primaria y en 2º o 3º año del nivel secundario, dependiendo de la estructura del sistema educativo de cada provincia y de la Ciudad Autónoma de Buenos Aires¹.

Las asignaturas elegidas para la evaluación de Aprender 2016 fueron Lengua y Matemática para todos los estudiantes a los que se aplicó el operativo (censal y muestral). Los estudiantes del último año de nivel secundario, fueron evaluados además, en Ciencias Naturales y Ciencias Sociales.

NIVEL PRIMARIO	3º grado (Muestral)	Lengua		Matemática	
	6º grado (Censal)				
NIVEL SECUNDARIO	2º/3º año (Muestral)	Lengua		Matemática	
	5º/6º año (Censal)	Lengua	Ciencias Naturales	Matemática	Ciencias Sociales

Tabla 2: Áreas de conocimiento evaluadas en Aprender 2016 y carácter censal/muestral de la población estudiantil evaluada, por curso.

¹ Las jurisdicciones que cuentan con una organización del sistema educativo de 7 años para la primaria y 5 para la secundaria en las cuales la muestra corresponde a estudiantes de 3º grado de primaria y 2º de secundaria son: Ciudad de Buenos Aires, Chaco, Jujuy, La Rioja, Mendoza, Misiones, Neuquén, Río Negro, Salta, San Juan, Santa Fe y Santiago del Estero. Las jurisdicciones restantes cuentan con un sistema de 6 años para la educación primaria y la secundaria por lo que se seleccionaron por muestra estudiantes del 3º grado primaria y 3º año de secundaria en Buenos Aires, Catamarca, Córdoba, Corrientes, Chubut, Entre Ríos, Formosa, La Pampa, San Juan, San Luis, Santa Cruz, Tierra del Fuego, Tucumán.

3.1 PRUEBAS REFERIDAS A CRITERIO Y TEORÍA DE RESPUESTA AL ÍTEM

Existen dos tipos de prueba: las pruebas o tests referidas a criterio (TRC) y las pruebas referidas a norma (TRN). Aprender es una prueba referida a criterio.

Las TRC se definen en relación a la relevancia y representatividad de los ítems respecto al dominio específico, por lo que la validez del contenido es fundamental. Privilegian la comparación de los logros de los estudiantes con respecto a las metas de aprendizaje o a las competencias que el sistema educativo persigue que los mismos alcancen. Las puntuaciones tienen carácter absoluto y están en relación al dominio medido en la prueba. En este sentido, sirven para retroalimentar y monitorear el progreso de los estudiantes o del sistema. El propósito es conocer y tomar decisiones sobre: a) si los individuos alcanzan o no el dominio o competencia evaluada; y b) determinar la eficacia de programas y sistemas educativos.

Las TRC se distinguen de las TRN. Estas últimas se definen como aquellas en las que la medida de desempeño es relativa al conjunto evaluado. Esto es, se enfocan en la comparación entre estudiantes o entre grupos de estudiantes. Por lo tanto, se omite la referencia a objetivos o logros esperados. La valoración es en comparación con el desempeño promedio del grupo.

A su vez, los ítems de las pruebas correspondientes a Aprender 2016 se analizaron con base en TRI, teoría en la que se enmarca el modelo general sobre el cual se basa la mayoría de las evaluaciones estandarizadas internacionales, y las propias ONE desde 2005. Sus supuestos básicos en el caso del modelo de dos parámetros son:

- a) Para cada disciplina evaluada en cierto año, el estudiante posee una habilidad, rasgo latente (no observado), competencia, etc. que puede asociarse con un número real θ .
- b) La probabilidad de responder correctamente a un cierto ítem i es una función creciente y continua del valor θ .

En el modelo logístico a dos parámetros la probabilidad de respuesta correcta se modeliza mediante:

$$\Pr(I_i = 1 | \theta) = \frac{\exp(D \cdot a_i \cdot (\theta - b_i))}{1 + \exp(D \cdot a_i \cdot (\theta - b_i))}$$

donde D es un escalar que normalmente se fija en 1,76;

a_i es la discriminación del ítem i (el ítem funcionará bien si este parámetro es mayor a 0);

y b_i es la dificultad del ítem.

Uno de los principales supuestos de TRI es la unidimensionalidad, esto es que una sola habilidad explica la probabilidad de respuesta a cada ítem en cierta disciplina y que esta

habilidad está asociada con un solo parámetro real, uno para cada estudiante. Otro supuesto, es la independencia local: la probabilidad de respuesta a un ítem (dada una cierta competencia) es independiente de la probabilidad de respuesta a otro ítem del mismo test.

En Aprender 2016, se optó por un modelo de dos parámetros, en vez del modelo logístico a un parámetro que fuera utilizado en ONE 2013. El modelo de dos parámetros aporta más flexibilidad al permitir discriminaciones distintas entre los ítems. Esto implica que, para cada ítem, se estimó no sólo su dificultad sino también su capacidad de discriminación (ver apartado “Escalamiento y comparabilidad entre los puntajes TRI de ONE y Aprender 2016”).

3.2 MARCO DE REFERENCIA Y ELABORACIÓN DE ÍTEMS

El marco de referencia de una evaluación es el conjunto delimitado de conocimientos y habilidades evaluados. En el caso de Aprender, ese conjunto consiste en los Núcleos de Aprendizaje Prioritarios (NAPs), los diseños curriculares jurisdiccionales y los consensos jurisdiccionales.

La SEE tomando como base el trabajo de elaboración de especificaciones (capacidades y contenidos para cada año evaluado) llevado adelante por la DINIECE durante el año 2015 seleccionó los criterios para la elaboración de los ítems de las pruebas.

CAPACIDADES Y CONTENIDOS

Los siguientes cuadros muestran las capacidades y contenidos, para cada área de conocimiento y grado o año, considerados para la elaboración de los ítems de la prueba piloto 2015. En todos los casos se evalúan los contenidos de todo un ciclo de manera que corresponde que los estudiantes contesten todas las preguntas.

CAPACIDADES Y CONTENIDOS DE LENGUA 2º/3º Y 5º/6º (NIVEL SECUNDARIO)	
BLOQUES DISCIPLINARES	CONTENIDOS
Extraer	<ul style="list-style-type: none"> • Información explícita en texto literario y no literario. • Secuencia en texto literario y no literario. • Resumen.
Interpretar	<ul style="list-style-type: none"> • Tema en texto literario y no literario. • Relaciones textuales. • Procedimientos de cohesión. • Elementos de enunciación en textos argumentativos: pequeños ensayos y notas especializadas. • Características de personajes. • Vocabulario. • Información inferencial. • Relación texto-paratexto.
Reflexionar y evaluar	<ul style="list-style-type: none"> • Estructura textual. • Recursos literarios y retóricos. • Tipos de narradores. • Tipologías y géneros discursivos.

CAPACIDADES Y CONTENIDOS DE LENGUA 3º Y 6º (NIVEL PRIMARIO)

CAPACIDADES COGNITIVAS	CONTENIDOS
Extraer	<ul style="list-style-type: none"> • Información textual explícita. • Estructura textual: información nuclear y periférica. • Secuencia en texto literario y no literario. • Características de hechos, objetos y personajes.
Interpretar	<ul style="list-style-type: none"> • Vocabulario. • Paratextos. • Cohesión y coherencia. • Resumen. • Tema. • Información inferencial.
Reflexionar y evaluar	<ul style="list-style-type: none"> • Géneros discursivos. • Estructura narrativa. • Identificación de hechos y opiniones. • Recursos literarios. • Tipos de narradores (solo en sexto). • Formatos textuales.

CAPACIDADES COGNITIVAS ESPECÍFICAS DE MATEMÁTICA (NIVELES PRIMARIO Y SECUNDARIO)

Reconocimiento de datos y conceptos	Capacidad cognitiva de identificar datos, hechos, conceptos, relaciones y propiedades matemáticas, expresadas de manera directa y explícita en el enunciado.
Resolución de operaciones	Resolver operaciones en los distintos conjuntos numéricos utilizando distintos procedimientos.
Resolución de situaciones en contextos intramatemáticos y/o de la vida cotidiana	<p>Capacidad cognitiva de solucionar situaciones problemáticas contextualizadas, presentadas en contextos que van desde los intramatemáticos hasta los de la realidad cotidiana.</p> <p>Interpretar información: comprender enunciados, cuadros, gráficos; diferenciar datos de incógnitas; interpretar símbolos, consignas, informaciones; manejar el vocabulario de la Matemática; traducir de una forma de representación a otra, de un tipo de lenguaje a otro.</p>
Comunicación en Matemática	Expresión de procedimientos y resultados: reconocer las distintas etapas de un cálculo, identificar una justificación, una argumentación.

CONTENIDOS DE MATEMÁTICA 5º/6º (NIVEL SECUNDARIO)

BLOQUES DISCIPLINARES	CONTENIDOS
Números reales	<ul style="list-style-type: none"> • Números racionales. Operaciones. Orden. • Números irracionales. Representación en la recta. Operaciones.
Funciones	<ul style="list-style-type: none"> • Funciones: lineal, cuadrática, polinómica, exponencial, logarítmica, trigonométrica, racional. • Dominio e imagen de las funciones numéricas más usuales. • Representación gráfica de funciones numéricas más usuales presentadas en forma algebraica o viceversa.
Ecuaciones e inecuaciones	<ul style="list-style-type: none"> • Ecuaciones de 1º y 2º grado. • Ecuaciones logarítmicas y exponenciales. • Sistemas de dos ecuaciones lineales. • Inecuaciones lineales.

BLOQUES DISCIPLINARES	CONTENIDOS
Geometría y medida	<ul style="list-style-type: none"> • Posiciones de rectas. • Figuras y cuerpos. • Teorema de Pitágoras. • Proporcionalidad geométrica. Semejanza de figuras. • Perímetro área y volumen. Equivalencia de medidas. • Distancia entre dos puntos en el plano cartesiano. • Relaciones trigonométricas y resolución de triángulos rectángulos.
Estadística y probabilidad	<ul style="list-style-type: none"> • Gráficos de barras, cartesiano, circular, pictogramas, cuadros, tablas. • Frecuencia. • Medidas de tendencia central. • Variaciones, permutaciones y combinaciones. • Probabilidad simple aplicada a situaciones numéricas.

CONTENIDOS DE MATEMÁTICA 6° (NIVEL PRIMARIO)

BLOQUES DISCIPLINARES	CONTENIDOS
Números y operaciones	<ul style="list-style-type: none"> • Números naturales, fraccionarios y expresiones decimales. Reconocimiento y uso. • Sistema decimal de numeración. Características • Representación y ubicación de naturales, fraccionarios y decimales en la recta numérica. • Operaciones: suma, resta, multiplicación y división entre naturales, decimales y fraccionarios (excluida la división entre decimales y fraccionarios). • Resolución de problemas que requieren diferentes significados de las cuatro operaciones, incluida la proporcionalidad con constante entera. • Relaciones entre números: divisibilidad. • Resolución fundamentada de cálculos y/o situaciones problemáticas.
Geometría y medida	<ul style="list-style-type: none"> • Relación entre sistemas de unidades: longitud, capacidad, peso, superficie y tiempo. • Cálculo de medida: estimación, aproximación y exactitud. • Perímetro: concepto. Perímetro de polígonos regulares. • Área: concepto, unidades, equivalencias. Área de polígonos comunes. • Sistemas de referencias para la ubicación de puntos en un plano. • Figuras geométricas: reconocimiento de elementos y propiedades de triángulos, cuadriláteros, circunferencia y círculo.
Geometría y medida	<ul style="list-style-type: none"> • Cuerpos geométricos: reconocimiento y propiedades de prismas, pirámides, cubo, cilindro, cono y esfera. • Resolución de problemas que requieren analizar, describir, comparar, clasificar y construir figuras en base a las propiedades conocidas.
Estadística y probabilidad	<ul style="list-style-type: none"> • Expresión e interpretación de datos a través de cuadros, diagramas y gráficos estadísticos. • Resolución de problemas que requieran interpretación de datos explícitos e implícitos en diferentes gráficos.

CONTENIDOS DE MATEMÁTICA 2°/3° (NIVEL SECUNDARIO)

BLOQUES DISCIPLINARES	CONTENIDOS
Números y operaciones	<ul style="list-style-type: none"> • Números naturales, enteros, racionales en su expresión fraccionaria o decimal. • Operaciones. Propiedades. • Orden. • Equivalencia. • La recta numérica. • Múltiples y divisores. • Notación científica. • Ecuaciones.

BLOQUES DISCIPLINARES	CONTENIDOS
Funciones	<ul style="list-style-type: none"> • Función lineal. • Función de proporcionalidad directa. • Razón y proporción. • Escala. • Porcentaje. • Función inversa.
Ecuaciones e inecuaciones	<ul style="list-style-type: none"> • Ecuaciones de 1° y 2° grado. • Ecuaciones logarítmicas y exponenciales. • Sistemas de dos ecuaciones lineales. • Inecuaciones lineales.
Geometría y medida	<ul style="list-style-type: none"> • Rectas paralelas y perpendiculares. • Sistema Métrico Legal Argentino (SIMELA) • Triángulos. Clasificación. Elementos. Propiedades. Perímetro. Área. • Circunferencia y círculo. Elementos. Propiedades. Longitud. Área. • Cuerpos. Cubos. Cilindro. Prismas. Elementos. Propiedades. Área. Volumen.
Estadística y probabilidad	<ul style="list-style-type: none"> • Gráfico de barras, circular, cartesiano, pictograma, histograma, cuadros y tablas. • Promedio. • Cálculo combinatorio. • Probabilidad simple.

CONTENIDOS DE MATEMÁTICA 3° (NIVEL PRIMARIO)

BLOQUES DISCIPLINARES	CONTENIDOS
Números y operaciones	<ul style="list-style-type: none"> • Números naturales. Reconocimiento, uso. Relaciones. Valor posicional. • Operaciones: suma, resta, multiplicación y división. • Cálculo exacto y aproximado. • Secuencias. • Análisis y elaboración de enunciados. • Registro y organización de datos en gráficos o tablas sencillas.
Geometría y medida	<ul style="list-style-type: none"> • Relaciones espaciales: sistemas de referencia para la ubicación de objetos o personas en un plano. • Figuras y cuerpos geométricos: Análisis, descripción y comparación, a partir de características conocidas. • Magnitudes: estimación o medición de cantidades con unidades convencionales y no convencionales: longitud, capacidad, peso. • Comparación y orden de diferentes unidades de uso frecuente, expresadas en enteros o medios y cuartos de ellas. • Uso del reloj y calendario para calcular duraciones en días, horas y minutos.

CAPACIDADES DE CIENCIAS NATURALES 5°/6° (NIVEL SECUNDARIO)

Reconocimiento de conceptos	Incluye la identificación e interpretación de conceptos propios de las Ciencias Naturales. Involucra reconocer y distinguir características, identificar relaciones causa efecto, identificar explicaciones de fenómenos naturales, clasificar y comparar.
Comunicación	La comunicación en Ciencias Naturales contempla tanto la identificación de datos como la organización, interpretación y traducción de información en distintos formatos (tablas, gráficos, diagramas de flujo, esquemas y símbolos).
Análisis de situación	Esta capacidad cognitiva contempla la identificación, interpretación y análisis de evidencias, conclusiones y procesos de investigación científica. Incluye analizar y relacionar datos, deducir a partir de datos, predecir, reconocer variables, identificar patrones, reconocer problemas científicos y relacionar conclusiones con evidencias.

CONTENIDOS DE CIENCIAS NATURALES 5º/6º (NIVEL SECUNDARIO)

BLOQUES DISCIPLINARES	CONTENIDOS
Los seres vivos	<ul style="list-style-type: none"> El organismo humano como sistema integrado y la salud: locomoción, nutrición, homeostasis, reproducción. La vida: continuidad y cambio: Teoría celular, bases moleculares y genéticas de la vida, teoría de la evolución, biotecnología.
Materia y energía	<ul style="list-style-type: none"> Estructura de la materia: modelo cinético-corpúscular, teoría atómica molecular, lenguaje de la química. Reacciones químicas: conservación de masa, procesos endotérmicos y exotérmicos. Soluciones: soluciones ácidas, básicas y neutras. Concentración. Fuerzas y movimientos: cinemática y dinámica. Ondas: características de ondas mecánicas y electromagnéticas, espectro electromagnético. Electricidad y magnetismo: circuitos eléctricos, consumo domiciliario, seguridad eléctrica, relación entre corriente eléctrica y campo magnético. Energía: conservación y transformaciones. Energía trabajo y potencia. Calorimetría, conducción, convección y radiación. Cambios de estado.
Medio ambiente	<ul style="list-style-type: none"> Ecosistemas: intercambios de materia y energía entre los seres vivos, ciclos de la materia y flujo de energía. Problemas ambientales: problemas ambientales locales y globales, consecuencias de la obtención, uso y reutilización de los recursos en el ambiente y la salud. Fuentes de energía renovables y no renovables, impacto de su aprovechamiento en el medio ambiente.

CAPACIDADES DE CIENCIAS SOCIALES 5º/6º (NIVEL SECUNDARIO)

CAPACIDADES COGNITIVAS ESPECÍFICAS

Reconocimiento de hechos/datos	Identificar datos o hechos en un conjunto de información mediante la utilización de conocimientos que el estudiante posee.
Reconocimiento de conceptos	Identificar conceptos por medio de ejemplos, casos, atributos o definiciones, o bien, de reconocer ejemplos, casos, atributos o definiciones a partir de un concepto dado.
Interpretación/ Exploración	Obtener y cruzar información explícita o implícita a partir de la lectura comprensiva de distintos tipos de fuentes (textos, imágenes, mapas, gráficos, tablas, etc.).
Análisis de situaciones	Reconocer distintos tipos de relaciones-causales, de comparación, de contemporaneidad, de simultaneidad o seleccionar cursos de acción que requieren la aplicación de conceptos, hechos, datos o procesos previamente adquiridos.

CONTENIDOS DE CIENCIAS SOCIALES 5º/6º (NIVEL SECUNDARIO)

BLOQUES DISCIPLINARES	CONTENIDOS
Historia	<ul style="list-style-type: none"> Pueblos originarios de América y la Argentina. Expansión ultramarina europea. El sistema colonial hispanoamericano. La era de las revoluciones. Independencia de las colonias españolas en América. Construcción del Estado nacional argentino. Consecuencias de la crisis de 1929 a nivel mundial y su impacto en la Argentina. Peronismo. El mundo bipolar y la guerra fría.

BLOQUES DISCIPLINARES	CONTENIDOS
Geografía	<ul style="list-style-type: none"> • Configuración y dinámica del espacio geográfico mundial, americano y argentino. • Las bases naturales del mundo, de América y de la Argentina. • Ambientes y problemas ambientales en el mundo, en América y en la Argentina. • Estructura y dinámica demográfica y social del mundo, de América y de la Argentina. • Los espacios rurales y urbanos en el mundo, en América y en la Argentina.
Educación ciudadana	<ul style="list-style-type: none"> • Los derechos humanos. Derechos políticos, sociales, económicos y culturales. • La identidad individual y las identidades colectivas. • Golpes de estado en la Argentina. • Tipos de estado, tipos de democracia y formas de gobierno. • Sistemas de partidos políticos. • División de poderes. • La constitución nacional. Origen y reformas. Declaraciones, derechos y garantías. • La participación ciudadana. • Los medios de comunicación como formadores de opinión.

Posteriormente a la definición del marco general de referencia, se elaboró la estructura de la prueba a partir del diseño de la **tabla de especificaciones** en la que se indican los contenidos y las capacidades a evaluar así como el valor relativo de cada una.

Dentro de cada disciplina se evalúa la apropiación de contenidos y alcance de una capacidad o dominio de una habilidad. En todos los casos, los conocimientos y capacidades o habilidades son específicas de cada área disciplinar.

TABLA DE ESPECIFICACIONES - LENGUA 5º/6º AÑO (NIVEL SECUNDARIO)

CAPACIDAD/CONTENIDO	EXTRAER	INTERPRETAR	EVALUAR	TOTAL
Aspectos globales del texto	3 %	20 %	17 %	40 %
Aspectos locales del texto	23 %	33 %	4 %	60 %
Total	26 %	53 %	21 %	100 %

TABLA DE ESPECIFICACIONES - LENGUA 6º GRADO (NIVEL PRIMARIO)

CAPACIDAD/CONTENIDO	EXTRAER	INTERPRETAR	EVALUAR	TOTAL
Aspectos globales del texto	11 %	18 %	16 %	45 %
Aspectos locales del texto	22 %	25 %	8 %	55 %
Total	33 %	43 %	24 %	100 %

TABLA DE ESPECIFICACIONES - LENGUA 2º/3º AÑO (NIVEL SECUNDARIO)

CAPACIDAD/CONTENIDO	EXTRAER	INTERPRETAR	EVALUAR	TOTAL
Aspectos globales del texto	7 %	28 %	25 %	60 %
Aspectos locales del texto	4 %	32 %	4 %	40 %
Total	11 %	60 %	29 %	100 %

TABLA DE ESPECIFICACIONES - LENGUA 3º (NIVEL PRIMARIO)

CAPACIDAD/CONTENIDO	EXTRAER	INTERPRETAR	EVALUAR	TOTAL
Aspectos globales del texto	3 %	11 %	11 %	25 %
Aspectos locales del texto	36 %	38 %	1 %	75 %
Total	39 %	49 %	12 %	100 %

TABLA DE ESPECIFICACIONES - MATEMÁTICA 5º/6º AÑO (NIVEL SECUNDARIO)

CAPACIDAD/CONTENIDO	RECONOCIMIENTO DE CONCEPTOS	RESOLUCIÓN DE SITUACIONES EN CONTEXTOS INTRA Y EXTRA MATEMÁTICOS	COMUNICACIÓN EN MATEMÁTICA	TOTAL
Números	8 %	2 %	1 %	11 %
Funciones	8 %	12 %	8 %	28 %
Ecuaciones e inecuaciones	10 %	7 %	8 %	25 %
Geometría y medida	6 %	17 %	*	23 %
Estadística y probabilidad	4 %	7 %	2 %	13 %
Total	36 %	45 %	19 %	100 %

TABLA DE ESPECIFICACIONES - MATEMÁTICA 6º GRADO (NIVEL PRIMARIO)

CAPACIDAD/CONTENIDO	RECONOCIMIENTO DE CONCEPTOS	SOLUCIÓN DE OPERACIONES	RESOLUCIÓN DE SITUACIONES EN CONTEXTOS INTRA Y EXTRA MATEMÁTICOS	COMUNICACIÓN EN MATEMÁTICA	TOTAL
Números y operaciones	13 %	3 %	26 %	5 %	47 %
Geometría y medida	10 %	3 %	21 %	5 %	39 %
Estadística y probabilidad	3 %	1 %	7 %	3 %	14 %
Total	26 %	7 %	54 %	13 %	100 %

TABLA DE ESPECIFICACIONES - MATEMÁTICA 2º/3º AÑO (NIVEL SECUNDARIO)

CAPACIDAD/CONTENIDO	RECONOCIMIENTO DE CONCEPTOS	SOLUCIÓN DE OPERACIONES	RESOLUCIÓN DE SITUACIONES EN CONTEXTOS INTRA Y EXTRA MATEMÁTICOS	COMUNICACIÓN EN MATEMÁTICA	TOTAL
Números	8 %	6 %	6 %	6 %	26 %
Funciones	10 %	1 %	8 %	7 %	26 %
Geometría y medida	10 %	1 %	12 %	3 %	26 %
Estadística y probabilidad	3 %	* %	10 %	9 %	22 %
Total	31 %	8 %	36 %	25 %	100 %

TABLA DE ESPECIFICACIONES - MATEMÁTICA 3º (NIVEL PRIMARIO)

CAPACIDAD/ CONTENIDO	RECONOCIMIENTO DE CONCEPTOS	SOLUCIÓN DE OPERACIONES	RESOLUCIÓN DE SITUACIONES EN CONTEXTOS INTRA Y EXTRA MATEMÁTICOS	COMUNICACIÓN EN MATEMÁTICA	TOTAL
Números	17 %	11 %	31 %	17 %	76 %
Geometría y medida	10 %	*	11 %	3 %	24 %
Total	27 %	11 %	42 %	20 %	100 %

TABLA DE ESPECIFICACIONES - CIENCIAS NATURALES 5º/6º AÑO (NIVEL SECUNDARIO)

CAPACIDAD/ CONTENIDO	RECONOCIMIENTO DE CONCEPTOS	COMUNICACIÓN	ANÁLISIS DE SITUACIÓN	TOTAL
Seres vivos	8 %	8 %	24 %	40 %
Materia y energía	11 %	8 %	21 %	40 %
Medio ambiente	4 %	4 %	12 %	20 %
Total	23 %	20 %	57 %	100 %

TABLA DE ESPECIFICACIONES - CIENCIAS SOCIALES 5º/6º AÑO (NIVEL SECUNDARIO)

CAPACIDAD/ CONTENIDO	RECONOCIMIENTO DE HECHOS	RECONOCIMIENTO DE CONCEPTOS	INTERPRETACIÓN	ANÁLISIS DE SITUACIÓN	TOTAL
Historia	8 %	7 %	10 %	10 %	35 %
Geometría	1 %	8 %	17 %	7 %	33 %
Educación ciudadana	3 %	8 %	15 %	6 %	32 %
Total	12 %	23 %	42 %	23 %	100 %

El relevamiento de información realizado a partir de diciembre de 2015 con la colaboración de personal de la DINIECE, que hoy integran el equipo de la Secretaría de Evaluación Educativa, indica que durante la etapa de elaboración de los ítems (2014-2015), la DINIECE efectuó la convocatoria a los constructores de ítems: docentes en la disciplina con vasta experiencia áulica tanto en educación primaria como secundaria, algunos asimismo con experiencia previa en elaboración de ítems de opción múltiple; y también se convocó a lectores críticos (referentes pedagógicos y académicos de cada disciplina).

Una vez leídos, revisados y corregidos los ítems presentados por los constructores de ítems por parte del equipo pedagógico de la DINIECE fueron enviados a los lectores críticos, que son especialistas disciplinares que no participaron en la elaboración de los ítems por lo que pueden tener una mirada objetiva sobre la propuesta de prueba.

El personal de las jurisdicciones no tuvo acceso a los ítems, con lo cual el consenso que existe a nivel federal se estableció únicamente respecto a los criterios de evaluación. En el caso de Lengua, la búsqueda, revisión y selección de textos estuvo a cargo dicho equipo pedagógico.

3.3 PRUEBA PILOTO 2015

El armado de la prueba piloto se llevó a cabo de acuerdo a los criterios pedagógicos, disciplinares y psicométricos, en relación con la tabla de especificaciones incluida en el apartado anterior. La selección de provincias en las que se aplicó la prueba piloto se basó en función de sus niveles de desempeño general (bajo, medio y alto), respetando siempre los criterios de diferenciación (ámbito y sector de gestión).

Para esta instancia también se realizó el encadenado de bloques que componen la prueba. Antes de la implementación se armó un pre-diseño que se sometió luego a un proceso de revisión. Para cada área, se pilotearon 8 modelos constituidos por 8 bloques de 12 ítems cada uno, encadenados de modo tal de ocupar alternativamente la primera y la segunda posición en la prueba.

El piloto fue llevado a cabo en marzo de 2015. A partir del análisis psicométrico de sus resultados, se conformaron las pruebas definitivas.

3.4 AJUSTES 2016

En función de diagnósticos preliminares, a partir de diciembre de 2015 la Secretaría de Evaluación Educativa comenzó un proceso de análisis que tuvo como objetivo aumentar la validez del instrumento, lo que implicó una doble tarea: por un lado, eliminar ítems anómalos; por el otro, reducir el número total de ítems eliminando los abiertos de modo de contribuir a la obtención de una mayor tasa de respuestas y una política de devolución a tiempo de la información. Este proceso se realizó respetando las tablas de especificaciones originales y el balance en los niveles de dificultad. Cada instrumento (año y área evaluado) pasó de estar conformado por 90 ítems (bloques de 15) a 72 ítems (bloques de 12).

Los criterios tenidos en cuenta para la eliminación de ítems fueron:

- a) Que el ítem fuera anómalo, es decir, que mostrara valores inadecuados en los indicadores paramétricos (dificultad, infit, outfit, correlación Ptmea y discriminación).
- b) Que el contenido específico del ítem eliminado quedara representado por otros ítems en la totalidad de la prueba.
- c) Que la dificultad del ítem eliminado no generara una "laguna" en el rango de dificultad de la prueba (caracterizado a partir del procesamiento del piloto 2015).

De este modo, se buscó garantizar la validez de contenido del instrumento.

4. CUESTIONARIO DE FACTORES ASOCIADOS

Siguiendo la tradición de los ONE así como de las evaluaciones regionales e internacionales, en Aprender 2016 se administraron cuestionarios complementarios para recabar información sobre el contexto de aprendizaje, tanto a nivel escolar (historia escolar, recursos) como extraescolar (características sociodemográficas del hogar de los estudiantes y acceso y utilización de TIC).

OBJETIVO

El objetivo de los cuestionarios complementarios fue obtener información de contexto para analizar los aprendizajes de los estudiantes.

DESTINATARIOS

Los cuestionarios fueron respondidos por:

- a) Los estudiantes.
- b) Los docentes a cargo de todas las secciones evaluadas o, en el caso del nivel secundario, el profesor a cargo de la primera hora del primer día de evaluación.
- c) Los directores.

En las escuelas rurales con secciones plurigrado, se aplicó un instrumento específico a estudiantes y a directores.

CUESTIONARIOS COMPLEMENTARIOS: CRITERIOS Y DEFINICIONES

Los cuestionarios aplicados en las distintas asignaturas fueron diseñados con participación federal, en función de principios pedagógicos y metodológicos relativos a mediciones en gran escala. Se utilizaron dispositivos con ítems/preguntas de respuesta múltiple, ya que permiten un procesamiento objetivo y rápido. Las alternativas de cada ítem/pregunta buscaron contemplar las distintas realidades del país con sus diferencias regionales, provinciales y de carácter urbano y rural. Se realizaron sucesivas revisiones en distintas instancias, para garantizar que los ítems pudieran ser respondidos por estudiantes que habitan en distintos lugares del país, independientemente de cuál fuera su contexto más próximo. Se consideró no sólo el conocimiento de su realidad cotidiana, sino también el universo cultural al que acceden a través del estudio, las lecturas y los medios masivos de comunicación.

A continuación, se explicitan los principales criterios y definiciones que se han tenido en cuenta para la construcción de los cuestionarios complementarios.

Modelo y ejes temáticos. Definición del enfoque, ejes temáticos y destinatarios

Los cuestionarios complementarios parten del modelo clásico de factores asociados, introduciendo un conjunto de preguntas básicas seleccionadas con el objeto de construir indicadores e índices clave.

El desarrollo de los instrumentos contempló la exploración de algunos ejes y temáticas que el Ministerio de Educación y Deportes de la Nación (MED) considera estratégicos para elaborar un diagnóstico censal de la situación, con información actualizada y precisa que oriente el diseño de las intervenciones de política adecuadas. En la elección de estos ejes, se han priorizado aquellos aspectos y problemáticas susceptibles de ser modificadas desde el accionar de la política educativa y las instituciones escolares.

A tales efectos, el contenido y enfoque de los cuestionarios complementarios fue discutido y consensado con áreas clave del MED y de las jurisdicciones. En particular, se realizaron revisiones técnicas y acuerdos con la Secretaría de Innovación y Calidad Educativa: (i) el área de curriculum, (ii) el Instituto Nacional de Formación Docente y (iii) el equipo técnico de Educ.ar.

Revisión y sistematización de antecedentes para la operacionalización de dimensiones

Para la elaboración de los cuestionarios complementarios se recurrió a experiencias nacionales e internacionales. Se utilizaron tres conjuntos de recursos:

- i) Resultados de cuestionarios implementados en Argentina recientemente. Se accedió a las bases de datos de los resultados de cuestionarios implementados en el contexto de evaluaciones nacionales (ONE 2013) e internacionales (TERCE 2013 para nivel primario, PISA 2012 para nivel secundario).

Se realizó una selección de preguntas considerando aquellas que fueran pertinentes para las dimensiones definidas. En el caso particular de TERCE se contó con el apoyo virtual de consultores de OREALC/UNESCO, que asesoraron en el proceso de análisis y selección de las preguntas.

- ii) Como segundo recurso de referencia, fueron consultados instrumentos utilizados en estudios nacionales vinculados a algunas dimensiones en particular. Si bien no se contó con la posibilidad de analizar sus distribuciones de frecuencias, se consideró como antecedente de validación la experticia de las instituciones e investigadores a cargo de estos desarrollos. En particular se analizaron los siguientes instrumentos: Encuesta Nacional sobre Integración de TIC en la Educación Básica Argentina (UNICEF); Encuesta Nacional sobre Acceso y Uso de Tecnologías (INDEC); Instrumento de Autoevaluación de la Calidad Educativa (IACE); Indagación de Oportunidades Educativas (AEPT); Estudio de clima, conflictos y violencia en la escuela (UNICEF y FLACSO); Relevamiento estadístico sobre clima escolar, violencia y conflicto (Ministerio de Educación); Módulo Oportunidades educativas de las personas 5 a 17 años de la EPH 2005 (INDEC).

iii) Evaluaciones internacionales de las que Argentina no participa y evaluaciones nacionales de otros países. En forma complementaria a los recursos anteriores, se realizó un análisis y selección de bloques de preguntas utilizados en instrumentos aplicados en estudios internacionales o de otros países, de los cuales se pudo disponer de información que permitió evaluar su funcionamiento. De casos nacionales se tomaron como referencia instrumentos implementados en México, Colombia, Chile y Ecuador. Por último, se accedió también al material desarrollado por el Instituto de Investigación y Desarrollo Educativo de la Universidad Autónoma de Baja California, del cual fue posible acceder a informes con análisis de respuestas.

Formulación de preguntas y cuestionarios

Para el desarrollo de las preguntas de los cuestionarios, se elaboraron propuestas que tomaron como base los antecedentes antes mencionados y se aplicaron los criterios metodológicos adecuados para fortalecer la validez y fiabilidad de los instrumentos. Se prestó particular atención a los siguientes puntos:

- a) Que la extensión del instrumento fuera adecuada a la edad de los estudiantes. En 6° grado de primaria y para el nivel secundario se tomó como referencia la extensión del ONE 2013. Para 6° grado se tomó también como parámetro el estudio TERCE y para secundaria el estudio PISA 2012. En el caso de 3° grado se buscó un conjunto mínimo de preguntas que pudieran asumir un formato similar al utilizado en las pruebas.
- b) Se buscó que la utilización del lenguaje fuera adaptada a la edad de los estudiantes y a la terminología propia de los rangos etarios. Las preguntas se formularon como interrogantes directos.
- c) Se buscó proponer un formato visual amigable, presentado en forma despejada que facilitara su correcto completamiento sobre el que se incluyeron instrucciones al inicio.
- d) Se priorizó el uso de preguntas cerradas, algunas de opción múltiple y otras de categorías excluyentes. Se incluyeron preguntas abiertas sólo para la carga de valores numéricos y exclusivamente en los instrumentos para docentes y directores. Se contaba como antecedente la pérdida de información que requería de lectura OCR de la aplicación del ONE 2013.
- e) Se buscó implementar en forma estratégica el recurso de escalas de distancia semántica (Escala Lickert, Escala de Semántica Diferencial) atendiendo por un lado a las posibilidades analíticas que ofrecen este tipo de herramientas, y por otro lado utilizando con precaución el recurso para evitar el efecto desgaste. En el caso de las escalas Lickert, se incluyeron reactivos de sentido favorable y desfavorable a los fines de evitar sesgos. La escasez del marco temporal impidió la validación de los reactivos a través del método de jueces.
- f) El orden de las preguntas fue estructurado para organizar el cuestionario en bloques temáticos y facilitar el llenado, distribuyendo los temas de mayor peso simbólico para reducir el efecto de irradiación o el efecto reatrimiento.

Elaboración de cuestionarios definitivos según destinatarios

a. Marco de referencia y definiciones de variables, dimensiones e indicadores.

CUESTIONARIO A ESTUDIANTES	
EJES Y DIMENSIONES	INDICADORES
Características sociodemográficas	Sexo, edad, nacionalidad, máximo nivel educativo alcanzado por los padres, pertenencia a pueblo originario, condición de migración del hogar.
Indicadores de condiciones de vida del hogar	Hacinamiento, percepción de Asignación Universal por Hijo en el hogar, participación en actividades de cuidado en el hogar y condición laboral de los estudiantes, maternidad y/o paternidad de los estudiantes.
Acceso a TICS en el hogar	Propiedad y acceso a teléfono celular/smartphone, competencias y tipo de uso de las TICS, acceso a internet, edad de inicio en la utilización de TICS; propiedad y acceso a bienes tecnológicos en el hogar.
Trayectorias escolares	Asistencia a nivel inicial según edad, repitencia primaria, repitencia secundaria, asignaturas previas y desaprobadas, inasistencias.
Autoconcepto y oficio de estudiante	Autopercepción del desempeño en lectura, escritura y matemática.
Clima escolar y convivencia democrática	Trato con compañeros y profesores, discriminación y violencia en la escuela, bienestar emocional en la escuela.
Prácticas de enseñanza percibidas	Estrategias de enseñanza que fomentan competencias cognitivas, interpersonales e intrapersonales, formas de evaluación utilizadas en la escuela.
Acceso a TIC en la escuela	Acceso a computadora, frecuencia de uso en clase, modelo de acceso (laboratorio, aula móvil, netbook por estudiante), acceso a internet, calidad de acceso internet.
Enseñanza con TIC	Frecuencia de realización de actividades con computadoras, estrategias de enseñanza mediadas por TIC implementadas en la escuela, tipos de uso educativo fuera de la escuela.
Uso recreativo de TICS	Frecuencia y tipos ampliados de uso recreativo de computadora fuera de la escuela.
CUESTIONARIO A DOCENTES	
EJES Y DIMENSIONES	INDICADORES
Características sociodemográficas	Sexo, edad, antigüedad en la docencia.
Contexto y condiciones de vida	Acceso a PC, internet, celular/smartphone.
Características del cargo	Antigüedad en el cargo, horas que trabaja en la escuela, situación de revista, asignatura a cargo en la escuela.
Actividad docente en general	Enseñanza en otras escuelas, niveles, modalidades. Otra actividad no docente, cantidad de escuelas en las que enseña, total de secciones a cargo.
Formación	Máximo nivel educativo, título docente, acceso a capacitaciones del MED.
Información sobre la sección evaluada	Total de estudiantes, desaprobados del segundo trimestre, con dificultades de inasistencia.

EJES Y DIMENSIONES	INDICADORES
Percepción sobre el aprendizaje	Proyección del total de desaprobados del año escolar, representaciones en torno a las oportunidades educativas, representaciones sobre las causas del fracaso escolar y del abandono.
Percepciones y estrategias de enseñanza	Estrategias de enseñanza que fomentan el desarrollo de competencias cognitivas, interpersonales e intrapersonales, auto percepción en el rol docente.
Acceso a TIC en la escuela	Acceso a PC, frecuencia de uso en clase, modelo de acceso (laboratorio, aula móvil, netbook x estudiante), acceso a internet.
Estrategias de enseñanza mediadas por TIC	Actividades de enseñanza con TIC.
Criterios de acreditación	Evaluación de alternativas a la repetición de grado.
Clima escolar	Percepción general del clima escolar.
Percepción sobre la gestión directiva	Seguimiento del director a la enseñanza (observaciones, revisión de planificaciones, uso de información), transmisión y apropiación del proyecto institucional, trabajo en equipo, planificación conjunta, atención a estudiantes con dificultades.

CUESTIONARIO A DIRECTORES

EJES Y DIMENSIONES	INDICADORES
Características del docente	Sexo, edad, antigüedad en la docencia.
Acceso a TIC	Adquisición de competencias TIC.
Características del cargo	Antigüedad en el cargo, situación de revista en la escuela, horas de dedicación.
Actividad docente en general	Enseñanza en otros cargos, en otras escuelas. Otra actividad no docente.
Formación académica	Máximo nivel educativo, título docente.
Características de la oferta educativa	Tipo de gestión, niveles ofertados, orientaciones, cantidad de turnos, divisiones y estudiantes, actividades a contraturno.
Infraestructura escolar	Acceso a servicios, disponibilidad de espacios, existencia de biblioteca, existencia de computadoras, acceso a Internet, calidad de la conexión a internet.
Percepción sobre enseñanza	Percepciones sobre las oportunidades de aprendizaje.
Percepción sobre la institución	Evaluación general de la institución en torno a ejes clave (calidad, rendimiento, infraestructura, etc.).
Percepción sobre estudiantes y aprendizajes	Indicadores de abandono, de estudiantes con alta inasistencia y muchas asignaturas desaprobadas. Representación de las causas del fracaso escolar y del abandono, razones que motivan el cambio de escuela.
Clima institucional	Percepción sobre los vínculos con (y entre) estudiantes y profesores, situaciones de violencia escolar, clima de trabajo entre profesores.
Características de las tareas de gestión y seguimiento a la enseñanza	Tiempo dedicado a cada tarea directiva, seguimiento del director a la enseñanza (observaciones, revisión de planificaciones, uso de información), criterios de distribución de estudiantes entre secciones.
Trabajo en red	Experiencias de trabajo en red, encuentros e intercambios con otras escuelas.

Lecciones aprendidas sobre los cuestionarios

Las preguntas sobre factores asociados han permitido elaborar un conjunto de índices compuestos para caracterizar las condiciones de vida en los hogares de los estudiantes, las percepciones de los estudiantes sobre los propios procesos de aprendizaje en el aula, sus trayectorias escolares y sobre sus vivencias del clima escolar. La construcción de los índices contribuyó a evaluar la robustez de algunas de las preguntas sobre factores asociados; la cantidad de no respuestas, así como la cantidad de “no sabe” constituyen un dato ineludible al momento de esa evaluación. La prueba de la calidad de las respuestas ha llevado a reconsiderar tanto la forma de preguntar como la inclusión o no de las mismas en los próximos cuestionarios de Aprender.

Por ejemplo, el índice de nivel socioeconómico se construyó con las variables: nivel educativo de madre y padre, hacinamiento, percepción de la Asignación Universal por Hijo por hogar, acceso a internet y acceso a TICs. En el caso de 6° grado de primaria la pregunta sobre recepción de AUH tuvo bajo nivel de respuesta y alta incidencia de “no sabe”, por lo que no fue incluida en el índice para ese grado y no sería incluida en cuestionarios futuros. Otras preguntas como la de “nacionalidad” del estudiante y sus padres o de pertenencia a pueblos originarios serán reformuladas en acuerdo con las normas establecidas por organismos rectores de las estadísticas latinoamericanas, como CELADE.

Asimismo, se está trabajando en el fortalecimiento de preguntas que permitan obtener información más sólida sobre el aprendizaje en zonas rurales y, en particular, de estudiantes que se encuentran bajo el Programa de Educación Intercultural Bilingüe (EIB). A tal efecto, en diciembre de 2016 se realizó una jornada federal en la que con la participación de las provincias con mayores porcentajes de escuelas rurales e EIB se discutieron los cuestionarios de factores asociados y se propusieron mejoras.

5. UNIVERSO Y MUESTRA DE ESTUDIANTES DE APRENDER 2016

5.1. UNIVERSO DE APRENDER 2016

La Ley de Educación Nacional N° 26.206 establece 14 años de obligatoriedad escolar desde el nivel inicial hasta la finalización del nivel secundario. Así, la educación obligatoria comprende tres niveles: inicial (salas de 4 y 5 años, jardín de infantes), primario (cuya duración es de 6 o 7 años según la jurisdicción) y secundario (5 o 6 años, dependiendo de la duración de la educación primaria³).

Para dar cuenta de la calidad del sistema educativo, Aprender se concentra en la medición de los aprendizajes de los estudiantes que cursan regularmente 6° año del nivel primario y 5°/6° año de la educación secundaria, dependiendo del plan de estudios de cada jurisdicción. La evaluación Aprender 2016 se aplicó de manera censal a todos los estudiantes de estos años de estudio.

Al mismo tiempo, Aprender mide los aprendizajes de una muestra representativa de estudiantes que cursan 3° grado del nivel primario y el 2°/3° año del secundario. Ambos operativos –tanto el censal como el muestral– abarcan a la educación común sin considerar a quienes asisten a las escuelas de Educación Especial.

En la Tabla 1 se presenta la cantidad de establecimientos educativos y cantidad de estudiantes (matrícula) de los grados y años que participaron del operativo Aprender. Esta información proviene del Relevamiento Anual Escolar correspondiente a 2015 realizado por el MED, actualizado con los cambios realizados durante 2016, con la información brindada por las autoridades jurisdiccionales.

³ Artículo 18. La educación inicial constituye una unidad pedagógica y comprende a los/as niños/as desde los cuarenta y cinco (45) días hasta los cinco (5) años de edad inclusive, siendo obligatorios los dos (2) últimos años. Artículo 26. La educación primaria es obligatoria y constituye una unidad pedagógica y organizativa destinada a la formación de los/as niños/as a partir de los SEIS (6) años de edad.

Artículo 29. La educación secundaria es obligatoria y constituye una unidad pedagógica y organizativa destinada a los/as adolescentes y jóvenes que hayan cumplido con el nivel de educación primaria.

Artículo 31. La educación secundaria se divide en DOS (2) ciclos: UN (1) Ciclo Básico, de carácter común a todas las orientaciones y UN (1) Ciclo Orientado, de carácter diversificado según distintas áreas del conocimiento, del mundo social y del trabajo.

JURISDICCIÓN	NIVEL PRIMARIO				NIVEL SECUNDARIO			
	3° GRADO		6° GRADO		2°/3° AÑO		5°/6° AÑO	
	Escuelas	Matrícula	Escuelas	Matrícula	Escuelas	Matrícula	Escuelas	Matrícula
CABA	878	40.296	873	39.388	489	42.317	480	29.136
Buenos Aires	5.519	283.876	5.632	282.470	4.446	261.745	3.694	163.936
Catamarca	409	7.573	417	8.020	322	7.532	156	6.068
Córdoba	1.852	58.648	1.867	59.894	956	57.536	945	35.423
Corrientes	852	24.190	874	23.279	313	17.775	297	11.431
Chaco	1.062	24.521	1.117	26.621	337	23.294	332	14.724
Chubut	237	10.144	237	9.638	174	9.709	146	6.451
Entre Ríos	1.078	23.418	1.081	24.213	522	21.911	489	12.594
Formosa	574	12.833	570	13.413	323	11.093	255	8.354
Jujuy	393	12.707	401	13.443	202	15.207	177	9.842
La Pampa	191	5.678	187	5.778	154	5.794	134	3.683
La Rioja	310	6.905	317	7.117	221	6.636	105	5.226
Mendoza	852	33.031	798	29.286	391	28.982	369	17.333
Misiones	1.142	26.630	1.151	26.460	563	23.690	459	13.602
Neuquén	351	11.810	348	10.714	129	11.384	123	5.938
Río Negro	377	11.794	362	11.773	184	13.271	175	6.654
Salta	783	26.277	781	25.745	410	28.430	320	16.030
San Juan	397	15.267	397	14.006	275	12.064	134	6.893
San Luis	310	9.353	320	9.595	203	8.037	155	5.256
Santa Cruz	112	6.188	112	6.000	78	6.037	73	3.205
Santa Fe	1.544	52.613	1.569	58.198	892	53.612	858	32.138
S. del Estero	1.178	21.689	1.200	21.744	560	17.523	390	13.381
Tucumán	739	27.899	740	29.607	567	27.942	380	17.550
T. del Fuego	56	2.880	53	2.888	41	2.924	31	1.974
Total	21.196	756.220	21.404	759.290	12.752	714.445	10.677	446.822

Tabla 1: Establecimientos y matrícula correspondiente a grados/años abarcados por Aprender 2016.

5.2. MUESTRA DE APRENDER 2016⁴

Ámbito geográfico

Todas las jurisdicciones de la República Argentina.

Ámbito poblacional

Todos los estudiantes matriculados en 3° grado de la escuela primaria y de 2°/3° año de la educación secundaria.

⁴ Para un mayor nivel de detalle remitirse al documento "Diseño de muestra – Aprender 2016".

Tipo de muestreo

Se trata de una muestra estratificada y por conglomerados. Se selecciona una muestra estratificada de escuelas en cada nivel. Y en cada escuela se sub-selecciona una sección, en forma aleatoria y equiprobable. Todos los estudiantes de la sección seleccionada fueron encuestados.

Estratificación del universo de escuelas

La selección de las escuelas se hizo estratificando el universo de escuelas bajo estudio, por jurisdicción, (Aglomerado principal/ Resto), sector (Estatal/ Privado) y ámbito (Urbano/ Rural).

En la provincia de Buenos Aires se introdujo un criterio adicional de estratificación respecto a ONE 2013: partidos del GBA. A su vez, "partidos del GBA" se subestratificó por partido. Esto permitió mantener una muestra balanceada dentro de la provincia de Buenos Aires y dentro del Gran Buenos Aires.

Tamaños de muestra en Aprender 2016

En Aprender 2016 se evaluaron:

ESTUDIANTES EN 3º GRADO=	55.294
ESTUDIANTES EN 2º/3º AÑO=	57.093
SECCIONES EN 3º GRADO=	3.454
SECCIONES EN 2º/3º AÑO=	3.254

El tamaño de muestra en Aprender fue superior al de ONE 2013: mientras que en ONE 2013 en 3º grado se evaluaron en Lengua y Matemática 25.101 y 29.153 estudiantes; en Aprender 2016 en 3º grado se evaluaron 52.953 y 51.544, respectivamente. En tanto, en 2º/3º de ONE 2013 se evaluaron 34.817 estudiantes en Lengua y 32.822 en Matemática. En Aprender 2016 estas cantidades fueron 53.889 y 54.287 respectivamente.

6. NIVELES DE DESEMPEÑO DE LOS ESTUDIANTES

6.1 IDENTIFICACIÓN DE LOS PUNTOS DE CORTE

Los resultados de las pruebas no suelen expresarse en términos de promedios u otros indicadores de la distribución de puntajes sino en agrupamientos que definen niveles de desempeño diferenciados. Este formato es muy difundido y suele usarse para la presentación de resultados de la mayoría de evaluaciones de desempeño escolar a nivel nacional, regional e internacional.

Para definir los niveles de desempeño se requiere establecer en primer término la categorización deseada. Es decir, precisar la cantidad de niveles, su denominación y los criterios de diferenciación entre los mismos. En función de esos criterios se establecen, posteriormente, las líneas de corte.

Para establecer los puntos de corte, Aprender 2016 se apoyó en el método Bookmark o Marcador, de uso muy difundido en el análisis de resultados de pruebas estandarizadas de desempeño escolar. Una de sus principales ventajas es que se ha diseñado de manera consistente con la Teoría de Respuesta al Ítem (TRI).

El método Bookmark requiere el ordenamiento de los ítems según la medida de su dificultad que surge del modelo TRI. A partir de este ordenamiento, se establecen –de la manera que se describe más adelante– aquellos ítems que constituyen los límites de cada uno de los niveles de desempeño considerados. Estos ítems son denominados “marcadores”.

Un estudiante estaría clasificado en un determinado estrato “i” si su puntaje Theta es mayor o igual que la dificultad asociada al marcador de ese estrato “i” e inferior al del estrato “i+1”.

Para facilitar el trabajo de los talleres en los cuales se definieron los puntos de corte, Aprender optó por mantener la misma cantidad de niveles de desempeño y las mismas etiquetas utilizadas en los ONE 2010 y 2013. Por lo cual, para estos fines, los niveles de desempeño fueron tres (3) y las etiquetas de los niveles de desempeño se correspondieron con Bajo, Medio y Alto.

En el marco de la metodología Bookmark, los marcadores que fijan los niveles de desempeño de los estudiantes en una determinada prueba deben ser definidos por un grupo representativo de especialistas en la materia que debaten y construyen consensos al respecto. En Aprender 2016, esta tarea se le encomendó a un grupo conformado por 201 docentes en ejercicio de las 24 jurisdicciones del país, pertenecientes a escuelas estatales y privadas, de ambos géneros, de diferentes edades y años de experiencia en el cargo.

En la semana del 13 al 17 de febrero del 2017, se llevaron a cabo los talleres de trabajo Bookmark. Apoyados en la información correspondiente a un subconjunto de los ítems de cada prueba –ordenados según su nivel de dificultad– los grupos debieron examinarlos cuidadosamente para poder definir los marcadores. Se analizó la dificultad de los ítems, las operaciones cog-

nitivas que entran en juego para contestarlas de forma correcta, las alternativas de respuesta que pueden facilitar o complejizar la elección de la respuesta correcta, entre otras cuestiones.

El debate se llevó a cabo durante tres rondas sucesivas. Para cada ronda los docentes debieron completar un formulario donde consignaron de manera individual los ítems elegidos como marcadores. Esa información se utilizó luego para abrir el debate en la ronda siguiente y, al final, se eligieron como marcadores aquellos consignados por los docentes luego de la tercera ronda. El valor theta promedio de los ítems elegidos como marcadores en la última ronda es lo que define los puntos de corte de cada nivel.

En ciertos casos, puede ocurrir que exista mucha dispersión entre los valores theta de los marcadores elegidos por los docentes en la última ronda. En estos casos, se suele convocar a un grupo de expertos –en la disciplina que se evalúa– para poder tomar una decisión final respecto a los marcadores elegidos.

6.2 REDEFINICIÓN DE LOS NIVELES DE DESEMPEÑO

Con posterioridad a la definición de los puntos de corte, en el marco del taller Bookmark, se tomó la decisión de cambiar la denominación de los niveles de desempeño (Alto, Medio y Bajo) que pasaron a denominarse: Básico, Satisfactorio y Avanzado. Esto no implicó cambios en los marcadores que se habían fijado, trasladándose las líneas de corte anteriores a las nuevas denominaciones.

Se realizó luego una modificación adicional: el nivel Básico se subdividió en dos. Esta decisión respondió a la necesidad de agudizar el análisis de este grupo. Quedó definido entonces, por un lado, el nivel denominado Por debajo del nivel básico, que incluye a los estudiantes cuyo puntaje se distancia en más de un 25% respecto del punto de corte del nivel Satisfactorio. Por otro lado, el nivel Básico pasó a incluir a los estudiantes cuyos puntajes no superan esa diferencia. Esta subdivisión y el criterio utilizado es similar a los ajustes que suelen realizarse en las evaluaciones estandarizadas a nivel regional e internacional.

De esta forma, quedaron definidas cuatro categorías:

POR DEBAJO DEL NIVEL BÁSICO

BÁSICO

SATISFACTORIO

AVANZADO

Para cada uno de estos cuatro niveles de desempeño se elaboraron los descriptores de nivel que refieren con detalle a los saberes y capacidades que deben poseer los estudiantes para poder alcanzar cada uno de los niveles de desempeño establecidos⁵. Los contenidos de aprendizajes detallados para cada descriptor de nivel siguen los lineamientos definidos en los Núcleos de Aprendizaje (NAP) acordados a nivel nacional por las 23 provincias y la Ciudad Autónoma de Buenos Aires en el marco del Consejo Federal de Educación.

⁵ Ver apartados “Descriptores de los niveles de desempeño” y “Descripción de capacidades por nivel de desempeño” que se encuentran en las “Notas Técnicas” del documento Aprender 2016-Primer informe de resultados.

7. ESCALAMIENTO Y COMPARABILIDAD ENTRE LOS PUNTAJES TRI DE ONE Y APRENDER 2016

El modelo de Respuesta al Ítem, marco teórico de Aprender 2016 y de la mayoría de las encuestas de evaluación educativa estandarizadas de la actualidad, asigna un puntaje ('competencia') θ a cada estudiante, en base a las respuestas de un conjunto de ítems. Este puntaje θ es un número real (positivo, cero o negativo). Y cada ítem se presupone posee (por ejemplo, en el modelo a dos parámetros) dos números reales que lo caracterizan, uno mide su dificultad y otro su discriminación. Las escalas de medición de las competencias y la de los parámetros de los ítems son la misma, aunque esta escala es arbitraria en el sentido de que no posee un cero absoluto: a diferencia del puntaje en la teoría clásica se puede modificar la escala aplicando una transformación lineal, mientras se aplique la transformación correspondiente a los parámetros de los ítems.

Si aplicáramos en dos ocasiones sucesivas la prueba (a cierto año y para cierta disciplina) pero sin ningún ítem en común, estrictamente hablando sería imposible desde la perspectiva del TRI medir cambios en las competencias medias del conjunto de estudiantes. En la práctica esta situación a veces se relativiza considerando que si dos ítems se refieren a la misma competencia (dentro de la disciplina), y difieren por ejemplo en la redacción, o en el caso de Matemática en los números que sirven de ejemplo, poseen casi la misma dificultad y discriminación, entonces se los puede considerar como intercambiables. Es preciso recordar que lo que se busca es estimar la diferencia en las competencias medias de dos colectivos de estudiantes, no de un estudiante en particular.

Para permitir la comparación de las competencias medias entre encuestas estandarizadas sucesivas es práctica común mantener, de una evaluación a otra, un cierto número de ítems en común, denominados ítems de anclaje.

Los ítems en común, el bloque de anclaje como también se denomina, permite realizar lo que los psicometristas denominan escalamiento: llevar las escalas de medición de las competencias de dos evaluaciones a una escala en común, para poder así comparar las variaciones en los puntajes medios estimados. Los software de procesamiento psicométrico fijan en general una escala de competencias con media 0 y desvío standard 1. Cada encuesta luego ajusta esta escala a la media y desvío que desee (corrigiendo también los parámetros de los ítems). Por ejemplo, PISA fijó en su primera ronda una media de 500 y desvío de 100. Esta escala luego es ajustada en cada medición mediante los bloques de anclaje.

Entre ONE 2013 y Aprender 2016, en cada disciplina de cada año, hubo 12 ítems en común. Las diferencias en las respuestas a estos ítems de un relevamiento a otro es lo que permite realizar el proceso de escalamiento.

Como en ONE 2013 se aplicó un modelo a un parámetro y en Aprender 2016 uno a dos parámetros, el escalamiento se hizo reprocesando las bases de las respuestas de los estudiantes a ONE 2013 aplicando el mismo programa que se construyó para procesar Aprender

2016. Esto además tiene la ventaja de eliminar posibles fuentes de variación introducidas por diferencias en el tratamiento estadístico o informático de las bases (eliminación de patrones de respuesta aberrantes, imputaciones, etc.). En ambos procesamientos la base de partida es la que contiene las respuestas de los estudiantes al conjunto de ítems. En ONE 2013 esta base contiene las respuestas a 90 ítems, mientras que en Aprender 2016 a 72.

Esquema del proceso de escalamiento aplicado

- a) La base con las respuestas a ONE 2013 se procesa con un modelo TRI a dos parámetros. Esto proporciona para cada ítem k , un par de valores $(\hat{a}_{1k}, \hat{b}_{1k})$, donde el primer valor del par es la discriminación del ítem k , y el segundo valor la dificultad del ítem k .
- b) Se repite el proceso para la base de respuestas de Aprender 2016, obteniéndose para cada ítem k un par de valores $(\hat{a}_{2k}, \hat{b}_{2k})$. Se busca entonces que la serie a escalar tenga la misma media y el mismo desvío que la de referencia. Para esto se ajusta la siguiente ecuación:

$$\hat{b}_{1k} = \beta + \alpha \cdot \hat{b}_{2k}$$

pero no mediante una regresión sino por

$$\hat{\alpha} = \frac{ds(\hat{b}_{2016i})}{ds(\hat{b}_{2013i})}$$

$$\hat{\beta} = \hat{b}_{2016i} - \hat{\alpha} \cdot \hat{b}_{2013i}$$

solo con los ítems i del bloque de anclaje.

- c) La escala de ONE 2013 resultante del procesamiento de la base de respuestas de los estudiantes con el mismo programa aplicado a Aprender, se escaló con la transformación:

$$\hat{\theta}_{2013i}^* = \hat{\beta} + \hat{\alpha} \cdot \hat{\theta}_{2013i}$$

- d) Las escalas 2013 ($\hat{\theta}_{2013}^*$) y 2016 ($\hat{\theta}_{2016}$) son ahora comparables.
- e) Se calcula la media de los puntajes $\hat{\theta}_{2013}^*$ y $\hat{\theta}_{2016}$. La diferencia entre esas medias es la diferencia de competencia de los estudiantes de cierto año en cierta disciplina. Puede ser que esta diferencia no sea estadísticamente significativa, algo que probablemente suceda si los dos relevamientos están cercanos temporalmente.

8. CRITERIOS DE ORGANIZACIÓN DE LA APLICACIÓN TERRITORIAL: APRENDER 2016 Y OBJETIVO DEL CAMBIO.

La revisión de la experiencia de la forma en que se organizó la implementación territorial de los últimos operativos de evaluación permitió identificar ciertos problemas, que llevaron a pérdidas de información. En este sentido, uno de los principales esfuerzos de Aprender está puesto en la mejora sustantiva en la organización de la etapa de implementación territorial del operativo.

Una innovación cualitativa fue el impulso que la SEE le dio a la participación de docentes y directivos como aplicadores y veedores del operativo, algo que sucedía por primera vez a nivel regional, con el objetivo de acercar la evaluación y el uso de información a la práctica cotidiana de las escuelas y de estimular su incorporación a la reflexión cotidiana.

En esta línea, la SEE se propuso fortalecer las unidades jurisdiccionales de evaluación, como instancias clave tanto en la organización del operativo como enlaces con las escuelas en la promoción del uso de la información empírica para la toma de decisiones que conduzcan a la mejora. Para ello, se organizaron capacitaciones en Buenos Aires y a nivel regional, contando con el apoyo de universidades y expertos en evaluación.

En síntesis, las principales mejoras se orientaron a:

- i) Fortalecer la capacidad instalada en las jurisdicciones, tanto de los referentes ministeriales como de los docentes participantes de Aprender.
- ii) Hacer explícitos los procesos de contratación pertinentes.
- iii) Fortalecer las instancias de control, seguimiento y supervisión de todas las etapas del operativo.

El siguiente cuadro sistematiza los principales cambios realizados en la parte operativa y logística del Operativo Aprender 2016, respecto del ONE 2013.

APRENDER

Aplicadores y supervisores	Docentes en ejercicio y directores que rotan de escuela, a los cuales se les asignó una remuneración adicional por la tarea.
Capacitación en la aplicación	Capacitación ampliada, conformación de una red de evaluadores. Capacitaciones específicas para APRENDER 2016.
Distribución de materiales	Descentralizada en un conjunto amplio de cabeceras definidas por cada jurisdicción. Total de cabeceras: 653
Pliegos para contrataciones	Especificaciones detalladas con todos los requerimientos necesarios, incluyendo cláusulas de no cumplimiento.
Seguimiento del transporte de materiales	Sistema de <i>track and trace</i> de las cajas que permitirá conocer en tiempo real la situación de cada una. Código de barras en las etiquetas que permitirá un control más rápido de la recepción de las cajas.
Digitalización de las evaluaciones	Orden secuenciado de digitalización, asegurando un mayor control y la finalización del proceso en tiempo y forma. Acceso web en tiempo real a los instrumentos digitalizados, para un control periódico.
Impresión y lectura óptica	Campos de identificación de cuestionarios, escuelas, sección y estudiante pre impresos e imputados en un código de barras, que permite una rápida lectura óptica y reduce la pérdida de información por errores de los aplicadores.
Visitas de control	Se realizaron visitas periódicas para controlar la calidad de los materiales impresos.
Trabajo en equipo con las empresas de impresión y lectura óptica	Trabajo en equipo entre el equipo de la SEE, los diseñadores y las empresas de impresión y digitalización.
Veedores internacionales	Se contó con la evaluación externa de veedores expertos de México, España, Chile y Brasil, que visitaron escuelas el día de la implementación del operativo, en cuatro jurisdicciones diferentes.

Tabla 2: Criterios de organización de la aplicación territorial.

9. PROCESAMIENTO DE DATOS DE APRENDER 2016

El procesamiento de datos de Aprender 2016 fue liderado por el equipo de Metodología de la SEE.

9.1 DISEÑO DE LA BASE Y CARGA DE DATOS

La digitalización de los cuestionarios se realizó a través de la lectura óptica con escáner. El proceso consiste en escanear los cuestionarios y, por medio de un programa, detectar si las burbujas fueron completadas o no. En caso de que el programa no pueda determinar la respuesta marcada, se realiza un control manual asistido por computadora. En este control se verifican todas las preguntas en blanco o con multimarca.

Para facilitar la visualización y control de los cuestionarios, su escaneo y digitalización, la empresa digitalizadora diseñó un sistema online que permite ver las imágenes escaneadas en tiempo real junto con los datos codificados provenientes de la misma.

Esta herramienta permitió realizar el control de calidad del procesamiento de los cuestionarios. Cada día se controlaron los procesados el día anterior de forma diferenciada para cada tipo. Por medio de las tablas IRAM, se estableció una muestra de preguntas y los límites de aceptación para la misma, decidiéndose que se aceptarían hasta cuatro errores cada mil preguntas.

Además, se realizó un control suplementario para la totalidad de la base, verificando que no hubiera casos duplicados, casos incompletos y que las variables estuvieran dentro de los parámetros establecidos. Una vez realizado este control, se generaron las bases de datos a utilizar.

9.2 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. El objetivo de la depuración es eliminar de la base cuestionarios que puedan afectar la calidad de las estimaciones.

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.
- 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.

Se siguió este criterio estricto porque la devolución a nivel escuela exige un alto nivel de consistencia en los microdatos.

SÍNTESIS DE ELIMINACIÓN DE CASOS										
	3º GRADO		6º GRADO		2º/3º AÑO		5º/6º AÑO		5º/6º AÑO	
	L	M	L	M	L	M	L	M	CS	CN
Casos originales	56.183	56.263	519.616	528.610	56.280	56.369	266.651	289.162	263.874	285.470
Por más de 12 no respuestas se borran	2.420	3.931	13.714	11.522	1.506	1.191	3.151	5.397	10.620	8.756
Porcentaje de casos eliminados por más de 12 no respuestas	4,31 %	6,99 %	2,64 %	2,18 %	2,68 %	2,11 %	1,18 %	1,87 %	4,02 %	3,07 %
Estudiantes con patrones de respuesta atípico/aberrante	806	784	7.583	7.751	820	826	3.930	4.233	3.776	4.128
Porcentaje respuesta aberrante	1,43 %	1,39 %	1,46 %	1,47 %	1,46 %	1,47 %	1,47 %	1,46 %	1,43 %	1,45 %

Referencias: *M = Matemática; L = Lengua; CS = Ciencias Sociales; CN = Ciencias Naturales.

A pesar de tratarse de un porcentaje muy bajo, este tipo de comportamientos en los estudiantes nos pone el desafío de trabajar más fuertemente con los aplicadores y el conjunto de la comunidad educativa en la sensibilización acerca de la importancia de completar la evaluación.

9.2.1 TRATAMIENTO DE VALORES FALTANTES EN LAS PRUEBAS

A lo largo del cuestionario de un determinado estudiante puede encontrarse respuestas en blanco (*missing*) que pueden deberse a distintos motivos. Algunos pueden ser:

- No sabe la respuesta, sigue adelante con las siguientes preguntas para luego volver, pero este ítem queda finalmente en blanco.
- Se cansa. Todos los ítems, a partir de alguno de ellos en particular, quedan en blanco.
- Por un problema de impresión o diseño, un ítem es sistemáticamente saltado por los estudiantes con cierta probabilidad.

Hay varias alternativas para tratar las respuestas faltantes a los ítems. Algunas son:

- Un ítem en blanco está más cerca de una respuesta incorrecta que de una correcta y recomiendan considerarlo como respuesta incorrecta.
- 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 2016 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.

IMPUTACIÓN DE RESPUESTAS		
AÑO	MATERIA	PORCENTAJE DE REPUESTAS IMPUTADAS
3° grado primaria	Matemática	6,29
	Lengua	4,38
6° grado primaria	Matemática	2,42
	Lengua	2,03
2°/3° año secundaria	Matemática	3,15
	Lengua	1,56
5°/6° año secundaria	Matemática	2,15
	Lengua	0,89
	Sociales	2,46
	Naturales	2,85

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 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, salvo que haya menos de 5 ítems en blanco, en cada bloque, al final del cuestionario.

a) Eliminación de cuestionarios con menos del 50% de respuestas

En la práctica no hay unanimidad sobre cómo tratar a los cuestionarios con muy pocas respuestas contestadas (no blancos). En Aprender 2016, los cuestionarios con menos de 50% de respuestas son eliminados del cálculo. Al haber 24 preguntas en las pruebas Aprender 2016, se trata de cuestionarios con menos de 12 respuestas (correctas o incorrectas). El argumento para esto es que, con esa cantidad de ítems contestados, no es posible evaluar el desempeño del estudiante con cierta fiabilidad. 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) Eliminación de cuestionarios con patrones de respuesta atípicos

Respecto de este último punto, para determinar cuáles son los cuestionarios que siguen patrones de respuesta atípicos se calcula un estadístico denominado l . Este estadístico disminuye cuando el estudiante responde correcta o incorrectamente una pregunta que no se esperaba que conteste de esa manera, utilizando la función de log-verosimilitud, donde X_g es el *score* del ítem para el ítem g (en nuestro caso $X_g=1$ si la respuesta es correcta, 0 si es incorrecta):

$$l = \sum_{g=1}^k \{X_g \ln P_g(\theta) + (1 - X_g) \ln [1 - P_g(\theta)]\}$$

Se observa que si el estudiante contestó incorrectamente ($X_g=0$) la pregunta g , y posee un alto nivel de habilidad, aparece un sumando negativo con un gran valor absoluto: $1 - P_g(\theta)$ será cercano a cero.

El estadístico l_z es una versión estandarizada del estadístico l y está dada por:⁶

$$l_z = \frac{l - E(l)}{\sqrt{Var(l)}}$$

Donde la esperanza y la varianza de l está dada por:

$$E(l) = \sum_{g=1}^k \{P_g(\theta) \ln[P_g(\theta)] + [1 - P_g(\theta)] \ln[1 - P_g(\theta)]\}$$

$$Var(l) = \sum_{g=1}^k P_g(\theta)[1 - P_g(\theta)] \left[\ln \frac{P_g(\theta)}{1 - P_g(\theta)} \right]^2$$

Este estadístico se comporta ahora aproximadamente como una curva normal (0,1). Se consideraron como estudiantes con patrones de respuestas atípicas o aberrantes si el estadístico l_z es menor a -3 (menos tres). La probabilidad de que un estudiante al azar, que no contestó en forma malintencionada, caiga más allá de este valor es de alrededor del 0,1%.

⁶ Ver apartados "Descriptores de los niveles de desempeño" y "Descripción de capacidades por nivel de desempeño" que se encuentran en las "Notas Técnicas" del documento Aprender 2016-Primer informe de resultados.

9.2.2 TRATAMIENTO DE LA NO RESPUESTA EN LAS PREGUNTAS DE CONTEXTO

En Aprender 2016 se imputó mediante *hot-deck* la “no respuesta” siempre que ésta no supere cierto valor. Este método asigna un valor válido seleccionando al azar un donante entre las unidades con valores válidos que tenga el mayor grado de similitud con la unidad con el valor *missing*. Hay muchas variantes de este método, ya preprogramado en Stata. El aplicar esta técnica de imputación trae aparejado tres beneficios:

- i) Hay coherencia entre los cuadros generados a partir de diferentes variables de las bases de variables auxiliares.
- ii) El analista no debe preocuparse en eliminar los casos no válidos o realizar una imputación *ad hoc*.
- iii) Se reduce el posible efecto de sesgos introducidos por la “no respuesta”.

En ONE 2013 la “no respuesta” en las preguntas de contexto no tuvo ninguna corrección: cuando estas variables se cruzan con los niveles de desempeño, las tablas difieren en el total de casos válidos, dependiendo de la cantidad de valores *missing* en las variables de contexto.

⁹ Muñiz, José (1997). Introducción a la teoría de respuesta a los ítems. Ediciones Pirámide S.A. (65)

10. ANÁLISIS PSICOMÉTRICO DE LAS PRUEBAS

CÁLCULO DE LOS PARÁMETROS DE LOS ÍTEMS Y DE LAS COMPETENCIAS

Una vez recibida la base de datos con el conjunto de respuestas, se debe realizar la depuración y procesamiento de la misma para obtener:

- a) Para cada ítem i , las estimaciones de la discriminación a_i y su dificultad b_i
- b) Para cada estudiante j , la estimación de su habilidad (*rasgo latente*) θ_j

En la práctica hallaremos entonces \hat{a}_i, \hat{b}_i y $\hat{\theta}_j$. Como en toda estimación, es de esperar que estos estimadores presenten cierto grado de variabilidad (*varianza*).

Para obtener las estimaciones de las competencias de los estudiantes se aplica el método de máxima verosimilitud. Éste consiste en encontrar los valores de los parámetros que hagan más probable la matriz de respuestas obtenidas. La estimación se hace por aproximaciones sucesivas (iteraciones) que se detienen cuando los valores estimados de los parámetros convergen, esto es, cuando tras una iteración n no se producen cambios significativos en los valores estimados⁹.

En Aprender 2016 se utilizaron para el procesamiento de la base (depuración y cálculo de los parámetros) el software estadístico Stata Versión 14 y el software libre ICL (Bradley A. Hanson, 2002).

Eliminación de ítems de funcionamiento dudoso

Una vez estimados la dificultad y la discriminación de los ítems, se realiza una evaluación respecto del funcionamiento de los mismos. Se eliminaron ítems que presentaban una discriminación menor a 0.15. El proceso se complementó con el cálculo de la correlación biserial.

Verificación de los supuestos del modelo de respuesta al ítem

Una vez depurada la base, se debe verificar si los supuestos de la TRI (unidimensionalidad aproximada, independencia local y ajuste de los datos a la forma funcional de la probabilidad de respuesta) se cumplen. Para ello existe una serie de tests o pruebas, muchas de ellas gráficas, que las buenas prácticas recomiendan realizar. En el caso de Aprender 2016 se generaron informes que indican en qué medida estos supuestos se han cumplido.

En Aprender 2016 se testearon:

- Unidimensionalidad
- Independencia local
- Ajuste del modelo, en forma gráfica.

Unidimensionalidad

Uno de los supuestos requeridos para la implementación de la evaluación es la unidimensionalidad. Para corroborarlo se llevó a cabo un análisis factorial sobre los ítems de modo de demostrar que toda la variabilidad del modelo es explicada por un componente principal o dominante. Por ejemplo, el desempeño del estudiante en la evaluación de Matemática dependería principalmente de su habilidad en esta disciplina, teniendo menos influencia factores como la ansiedad, la motivación, capacidad de interpretación de textos, entre otros.

Para el desarrollo de este análisis es necesario codificar los ítems de forma dicotómica, donde 0 representa una respuesta incorrecta y 1 la correcta. Por tal motivo, no es factible utilizar la matriz de varianza-covarianza utilizada en este tipo de análisis teniendo que aplicar en su lugar la matriz de correlaciones tetracóricas.

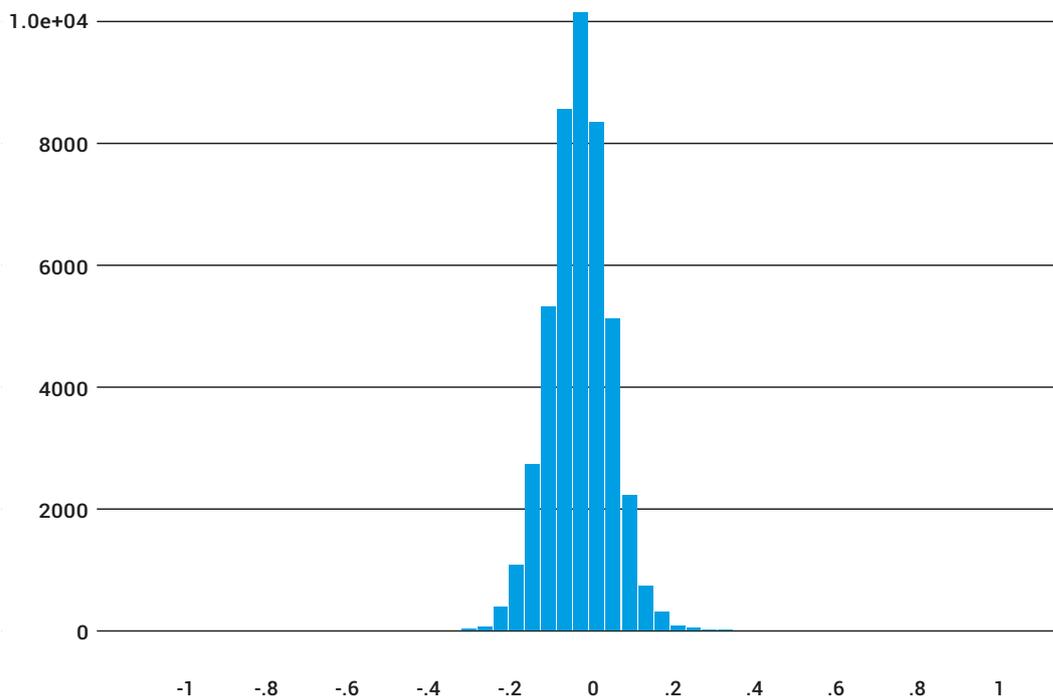
Independencia local

A grandes rasgos, la independencia local implica que el responder correcta o incorrectamente un ítem es independiente (estadísticamente) de la respuesta a otro ítem. Esto no se cumpliría, por ejemplo, si hay otra dimensión que interviene en el contestar bien o mal estos ítems.

Para esto, se agrupa a los estudiantes con habilidades similares y se calcula para cada grupo la matriz de correlación para ver si las respuestas a los ítems son independientes.

Los coeficientes de correlación tendrán que ser cercanos a 0 para que se cumpla el supuesto, como podemos observar en el siguiente ejemplo gráfico.

COEFICIENTES



Paso a un modelo logístico de dos parámetros

En ONE 2013, las respuestas a los ítems de las pruebas fueron definidas con un modelo logístico de un parámetro. En éste se supone que la probabilidad de respuesta correcta a un ítem I_i (pregunta), para un estudiante con un nivel de competencia θ_i está dada por

$$Prob(I_i = 1|\theta_i) = \frac{e^{(\theta_i - b_i)}}{1 + e^{(\theta_i - b_i)}}$$

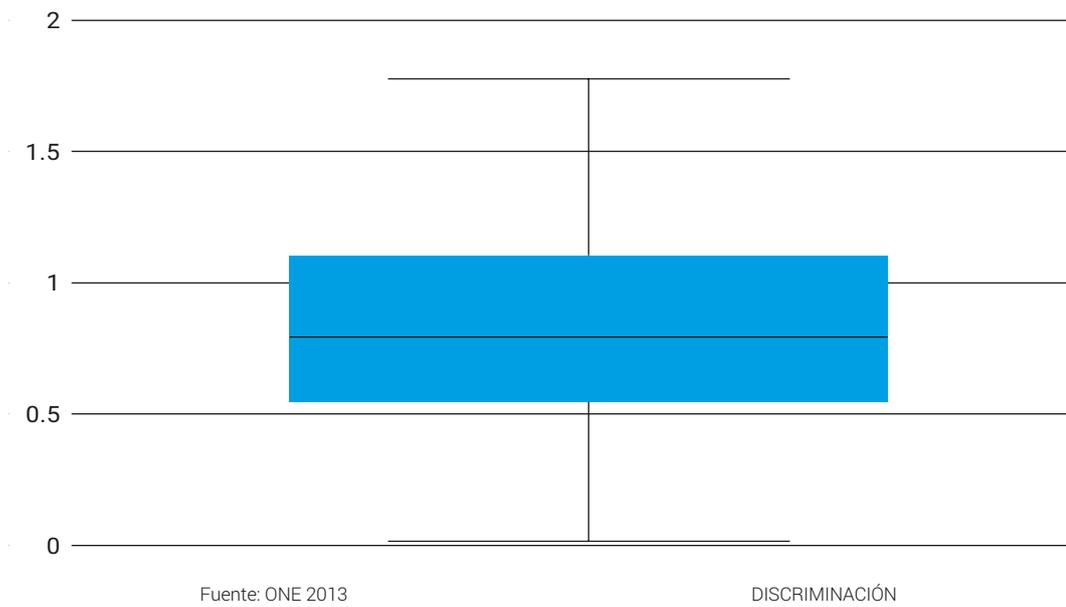
Este modelo presupone que todos los ítems poseen el mismo nivel de discriminación, lo único que varía de uno a otro es su nivel de dificultad que está dado por el parámetro b_i . Si este supuesto no se cumple, en general, las dificultades estarán mal estimadas con la consecuente mala estimación de las competencias. Hay otros modelos alternativos. El más sencillo es el logístico con dos parámetros que incorpora un segundo parámetro a_i (*discriminación*):

$$Prob(I_i = 1|\theta_i) = \frac{e^{a_i \cdot (\theta_i - b_i)}}{1 + e^{a_i \cdot (\theta_i - b_i)}}$$

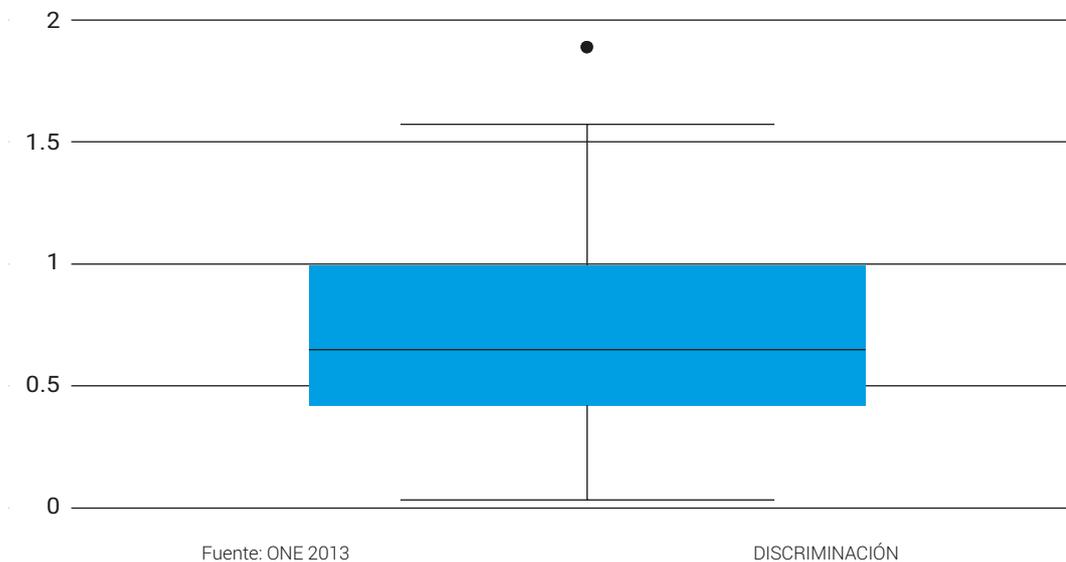
Este modelo es más flexible y permite incorporar diferencias en las discriminaciones de los ítems.

El análisis de los ítems en ONE 2013 verificó que las discriminaciones no eran iguales. Se presenta a continuación un gráfico de Box-Plot con las discriminaciones de los 90 ítems aplicados en ONE 2013 para 6° grado Lengua y Matemática.

DISTRIBUCIÓN DEL ESTIMADOR DE LA DISCRIMINACIÓN - 6º GRADO LENGUA



DISTRIBUCIÓN DEL ESTIMADOR DE LA DISCRIMINACIÓN - 6º GRADO MATEMÁTICA



Se observa en estos gráficos la dispersión en las discriminaciones, lo que confirma la conveniencia del paso a un modelo que tome en cuenta las discriminaciones de los ítems.

Análisis del funcionamiento de los ítems

Una vez hecho el primer cálculo de los parámetros de los ítems se debe evaluar cuáles no funcionaron correctamente y conviene eliminar. Este trabajo se realiza en conjunto entre los psicometristas y los especialistas en la redacción de ítems.

Como material para esta discusión, el equipo coordinador de Metodología genera una serie de informes donde se detallan los principales estadísticos y resúmenes de cada ítem:

- Proporción de blancos según orden del ítem, por modelo.
- Distribución de las respuestas, blancos incluidos.
- Distribución de las respuestas válidas.
- Coeficientes de correlación (ítem/ test).
- Proporción de respuestas válidas según nivel de competencia.
Se generan tres niveles de competencia 20%, 60%, 20%.
- Coeficientes de discriminación y dificultad.
- Gráfico de la curva ICC.
- Gráfico de la curva ICC más proporción de respuestas válidas según tramos de habilidad.

Con este material, los expertos deciden si hay ítems que conviene eliminar en la estimación de los parámetros y generación de los puntajes.

Si bien este proceso está documentado, no se publica para preservar la confidencialidad de los ítems.

Etapas del procesamiento para la obtención de puntajes y puntos de corte

El procesamiento de los instrumentos de evaluaciones estandarizadas hasta la obtención de los puntajes y puntos de corte a partir de los talleres Bookmark incluye, en síntesis, las siguientes fases:

- Depuración de las tablas relevadas, tanto de las pruebas como de las correspondientes a las variables de contexto.
- Ajuste preliminar del modelo y estimación de los parámetros de los ítems.
- Análisis del funcionamiento de los ítems: discriminación, funcionamiento diferencial, entre otros. Evaluación del ajuste del modelo.
- Estimación definitiva del modelo habiendo eliminado ítems con funcionamiento deficiente.
- Cálculo del puntaje estandarizado de los estudiantes, para cada año y disciplina.

- Determinación de los puntos de corte que determinarán los tramos de logro en cada una de las disciplinas y para cada uno de los años evaluados, en base a los resultados de los talleres Bookmark.

Los informes ejecutivos y los informes pormenorizados para los especialistas en las disciplinas, donde se detallan con gráficos y estadísticos específicos el funcionamiento de cada uno de los ítems fueron ya programados antes de la carga de los datos. Esto permite que antes de la finalización de la misma se puedan entregar informes periódicos sobre rendimiento de la muestra, no respuesta, funcionamiento de los ítems, etc., con un mínimo tiempo de procesamiento. Para ello se seleccionan muestras aleatorias estratificadas de los cuestionarios digitalizados, que son ponderadas para corregir distorsiones en las proporciones de cuestionarios que fueron llegando a las oficinas de digitalización.

11. REFERENCIAS

- Dikes, P., Tournois, J., Flieller, A., Kop, J.L. (1994). *La psychométrie*. Paris: Presses universitaires de France.
- Felipe Martínez Rizo. (2010). "Las pruebas de rendimiento y sus consecuencias", recuperado de: <http://www.mineducacion.gov.co/1621/article-241800.html>
- Haladyna, Thomas; Haladyna, Ronald y Merino Soto, César. "Preparación de preguntas de opciones múltiples para medir el aprendizaje de los estudiantes". En OEI-Revista Iberoamericana de Educación, s/f. Recuperado de: <http://rieoei.org/evaluacion5.htm>.
- Hambleton, R., H. Swaminathan y H. Rogers. (1991). *Fundamentals of Item Response Theory*. Newbury Park, CA: Sage Publications.
- Kolen, Michael J. y Brennan, Robert L. (2014). *Test Equating, Scaling, and Linking*. New York: Springer.
- Lacueva, A. (2014). *La determinación de la calidad escolar: ¿tecnocrática o democrática, parcial o integral?* Caracas: Ministerio del Poder Popular para la Educación.
- Laveault, D. y J. Grégoire. de Boeck. (2014). *Introduction aux Théories des Tests (3° Ed.)*. Bruselass: de Boeck Supérieur.
- Martínez Rizo, Felipe. (2003). "Evaluaciones de calidad educativa". En Iaies, Gustavo y otros: Evaluar las evaluaciones. Una mirada política acerca de las evaluaciones de la calidad educativa. Instituto Internacional de Planeamiento de la Educación, IIPÉ-UNESCO. Pp 137-138.
- Moreno, Rafael; Martínez, Rafael y Muñiz, José. (2004). "Directrices para la construcción de ítems de opción múltiple", *Psicothema*, 16 (3), pp 490-497.
- Ravela, Pedro. (2006). "Para comprender las evaluaciones educativas. Fichas didácticas", Santiago de Chile: Programa de Promoción de la Reforma Educativa en América Latina y el Caribe – PREAL.
- Ravela, Pedro et al. (2008). "Las evaluaciones educativas que América Latina necesita", *Revista Iberoamericana de Evaluación Educativa*, 1 (1).
- Raghunatham, T. (2016). *Missing Data Analysis in Practice*. Boca Ratón: CRC Press.
- Stufflebeam, D. L. (1968). *Evaluation as enlightenment for decision-making*. Columbus, OH: Evaluation Center, Ohio State University.
- Valliant, E., J. Dever y F. Kreuter. (2013). *Practical Tools for Designing and Weighting Survey Samples*. Singapur: Springer Nature.
- Wu, M., H. Tam, T-H y Jen. Springer. (2016). *Educational Measurement for Applied Researchers*. Singapur: Springer Nature.

