

Aprender 2017

CONSTRUCCIÓN

DE LAS PRUEBAS

CONTENIDO Y PROPIEDADES

PSICOMÉTRICAS

Aprender 2017

CONSTRUCCIÓN

DE LAS PRUEBAS

CONTENIDO Y PROPIEDADES

PSICOMÉTRICAS

AUTORIDADES

Presidente

Ing. Mauricio Macri

Ministro de Educación

Dr. Alejandro Finocchiaro

Jefe de Gabinete del Ministerio de Educación

Cdor. Javier Mezzamico

Secretaria de Evaluación Educativa

Prof. Elena Duro

Secretario de Gestión Educativa

Lic. Manuel Vidal

Secretaria de Políticas Universitarias

Mg. Danya Tavela

Secretaria de Innovación y Calidad

Sra. María de las Mercedes Miguel

COORDINACIÓN

Prof. Elena Duro

EQUIPO A CARGO DE LA ELABORACIÓN DEL DOCUMENTO

María Aranguren

María Elena Brenlla

Augusto Hoszowski

Agnés Laxague

Elisa Marilina Zayas

INTRODUCCIÓN

Una evaluación educativa nos brinda información clave sobre los logros y resultados de los estudiantes en un determinado momento. En este sentido, una evaluación es una foto panorámica o un corte transversal. Sin embargo, también es un proceso ya que implica la recolección sistemática de información y una interpretación de los resultados en función de ciertos criterios que permiten realizar un diagnóstico preciso del sistema educativo y que colabora a la toma de decisiones para acciones orientadas a las mejoras del sistema.

En noviembre de 2017, se aplicaron las pruebas Aprender en las áreas de Matemática y Lengua en 5°/6° año de la Educación Secundaria y en Ciencias Naturales y Sociales de 6° año de la Educación Primaria.

Dichos instrumentos fueron diseñados siguiendo los parámetros y condiciones estipulados para obtener resultados válidos y confiables. Para ello, se tuvieron en cuenta los lineamientos de los Estándares para la Evaluación Educativa y Psicológica (American Educational Research Association [AERA], American Psychological Association [APA] & National Council on Measurement in Education [NCME], 2014). En este documento se detalla cómo ha sido la construcción de cada una de las pruebas y se aportan evidencias acerca de su validez y fiabilidad. Primero se define qué son las pruebas referidas a criterio y su diferencia con las referidas a normas, luego se señalan los requisitos psicométricos de fiabilidad y validez que debe reunir una medida y, por último, los pasos seguidos para el diseño de las pruebas Aprender 2017.

PRUEBAS CRITERIALES Y PRUEBAS NORMATIVAS

En consonancia con las evaluaciones nacionales previas, las evaluaciones Aprender 2017 mantienen el mismo enfoque pedagógico que el utilizado en ocasiones anteriores. En este sentido, al igual que en Aprender 2016 y en los Operativos Nacionales de Evaluación (ONE) 2007, 2010 y 2013, se busca conocer los contenidos, destrezas y capacidades que los estudiantes logran dominar en cada una de las áreas y años evaluados. Este tipo de pruebas se denominan pruebas referidas a criterio y se diferencian de las pruebas referidas a la norma (Glaser, 1963).

En apariencia, las pruebas criteriosales y las pruebas normativas no muestran grandes diferencias ya que ambas evaluaciones suelen estar conformadas por ítems o ejercicios que siguen un formato similar y que requieren que el estudiante ponga en juego las mismas capacidades cognitivas en resolverlos (Leones, 2005). Sin embargo, existen diferencias sustanciales en la construcción y en las inferencias que es posible hacer a partir de cada una de ellas.

Una prueba normativa busca identificar cuál es la posición de un estudiante en relación con los demás estudiantes evaluados, mientras que una evaluación criterial busca conocer el dominio que tiene un alumno de ciertos criterios preestablecidos. Por ejemplo, en términos descriptivos, el desempeño de un estudiante en una prueba referida a criterio podría indicarse como “Enrique ha respondido correctamente al 75% de las preguntas de la prueba de Lengua” que difiere de “El desempeño de Enrique supera al 75% de los estudiantes que respondieron a la prueba de Lengua”, propio de las pruebas referidas a normas.

A partir de los cambios introducidos en la educación en las décadas pasadas -en los que se dio especial énfasis a los programas e intervenciones educativas-, fue necesario revisar las evaluaciones referidas a las normas de grupo ya que no resultaban apropiadas para conocer los rendimientos educativos ni para diagnosticar necesidades de intervención (Martínez Arias, 1995). Tal como lo ilustra el ejemplo anterior, ¿de qué sirve conocer la posición de un estudiante en su clase si no sabemos si logró alcanzar los objetivos básicos de la materia? Enrique podría superar al 75% de sus compañeros contestando correctamente unos pocos ítems más que ellos y aun así no haber logrado los objetivos propuestos en el programa o en la evaluación (por ejemplo, contestar correctamente al 50% de los ítems de la prueba).

En relación con lo anterior, cabe aclarar que las pruebas criteriosales han sido desarrolladas con el objetivo de corregir aquellas inadecuaciones de las pruebas normativas, particularmente, sus limitaciones para ser utilizadas en evaluaciones estandarizadas (Jornet Meliá & González Such, 2009). Así, las pruebas criteriosales permiten interpretar los resultados obtenidos por los estudiantes de acuerdo a ciertos criterios absolutos referidos al dominio que se está evaluando.

Las pruebas Aprender 2017 siguen la línea inaugurada en 2003 en nuestro país y se realizan bajo el enfoque de pruebas referidas al criterio. Ello implica que los dominios –sean

objetivos, capacidades, destrezas o competencias- están bien definidos, que los ítems los cubren apropiadamente y que si se establecen estándares de rendimiento (o puntos de corte) esto se realizará con la metodología apropiada. Las pruebas criterioles se conforman por ítems o reactivos que responden a los objetivos del aprendizaje. Como se indicó, los ítems incluidos en la prueba son un conjunto representativo de un dominio claramente definido (Hambleton & Swaminathan, 1978), permiten conocer la ubicación de un sujeto en un continuo representativo del dominio evaluado –y no respecto del grupo normativo- y se obtiene una interpretación directa de la puntuación obtenida: el desempeño que exhibe el estudiante indica su grado de competencia (Leyva Barajas, 2011).

EVIDENCIAS DE VALIDEZ Y FIABILIDAD DE UNA PRUEBA EDUCATIVA

En la construcción de una prueba educativa –al igual que en cualquier construcción de un test– se buscará que sus resultados sean válidos y confiables. Para ello será necesario contar con evidencias suficientes que demuestren que el instrumento utilizado funciona correctamente. En ocasiones es posible realizar algunos ensayos o pruebas piloto para corregir posibles errores en el instrumento y evitar errores en la medida.

El concepto de validez se utiliza para indicar el grado con el que un instrumento efectivamente mide aquello que se propone o debe medir (Martínez Arias, 1995). La fiabilidad, por su parte, es una propiedad psicométrica que hace referencia a la precisión de la medida (Cronbach, 1971). Un test es confiable cuando al repetirse su aplicación –en un período de tiempo relativamente acotado y en un mismo grupo de participantes– arroja resultados similares a los obtenidos en ocasiones previas (Cortada de Kohan, Macbeth & Alonso, 2008). A su vez, una prueba es válida cuando las inferencias realizadas a partir de ella son adecuadas, significativas y útiles (AERA, APA & NCME, 2014).

Actualmente, se sostiene que la validez y la fiabilidad no son atributos fijos o inmutables de un instrumento, sino que se trata de una serie de juicios que aportan evidencias acerca de qué tan apropiadas son las inferencias que puedan realizarse de una prueba en un determinado contexto o para un determinado grupo (AERA, APA & NCME, 2014; Covacevich, 2014). Esto se debe a que los estudios de validez y fiabilidad de un instrumento se efectúan sobre muestras delimitadas, con características propias, y respuestas y rendimientos concretos. Un instrumento puede ser válido para una población y no para otra. Por ejemplo, una prueba de razonamiento lógico diseñada para alumnos de 5° año no podrá reflejar de manera válida las habilidades de razonamiento lógico de un estudiante de 2° año. De ahí que en el área de la psicometría y de la investigación educativa, se hable entonces de evidencias de validez o evidencias de fiabilidad para un determinado grupo o determinado contexto.

Evidencias de validez

Teniendo en cuenta que la validez está relacionada con las inferencias que podrán realizarse a partir de los resultados de una prueba, para juzgar su calidad se deberán reunir varios indicadores y tipos de información (Covacevich, 2014). Estos indicadores y evidencias pueden ser obtenidos a través de distintas fuentes y enfoques. De acuerdo a las fuentes y/o enfoque utilizados, tendremos evidencias de distintos tipos de validez. Tradicionalmente, se suele hablar de evidencias referidas a la validez de contenido, referidas a la validez de criterio y evidencias referidas a la validez de constructo (Leyva Barajas, 2011).

Las evidencias de validez de las pruebas criterioles deberán ser analizadas en función del propósito para el cual son diseñadas. Básicamente, una prueba criterial busca describir el dominio que tienen los participantes de un área disciplinar y clasificarlos según niveles de desempeño de acuerdo a sus habilidades y capacidades (Hambleton, Swaminathan & Rogers, 1991). De ahí que se buscará, en primer lugar, reunir evidencias de la validez de contenido de la prueba.

El contenido de una prueba educativa estará delimitado por los contenidos y capacidades evaluadas de acuerdo a la disciplina y año. Por otra parte, las evidencias de validez de contenido pueden ser obtenidas en diferentes etapas: a) proceso de construcción de ítems; b) proceso de puesta a prueba de los ítems y; c) proceso de selección de los ítems que conformaran la versión final de la prueba. Cada etapa se caracteriza por presentar diversas fuentes de información.

En la primera etapa, se deberá conformar un grupo de itemistas y de supervisores (e.g. lectores críticos, expertos en psicometría y en gestión curricular) que atiendan a las tablas de especificaciones en donde se delimitan los contenidos y capacidades a evaluar para la confección, evaluación y edición de los reactivos. En segundo momento, se tendrán en cuenta los resultados y análisis de la prueba piloto. Por último, se realizará una selección y ajuste de los ítems para la prueba final. Cada una de estas etapas brinda información acerca de la validez de contenido de la prueba y también respecto de su fiabilidad. En el próximo apartado se indican los procedimientos llevados a cabo para analizar la validez de contenido de las pruebas Aprender 2017.

Evidencias de fiabilidad

En las pruebas criterioles, la confiabilidad es analizada a partir de la Teoría de Respuesta al Ítem (TRI) o Teoría del Rasgo Latente. Uno de los supuestos de la TRI es que existe una relación directa entre los resultados obtenidos en un ítem o en un conjunto de ítems, y el grado en que se posee la variable o rasgo evaluado (Hambleton et al, 1991; Leones, 2005, Secretaría de Evaluación Educativa [SEE], 2017). Es decir, que el acierto en un ítem o en un conjunto de ítems, implica el dominio de los contenidos y capacidades evaluadas. Así, el postulado principal de los modelos TRI es que existe una relación funcional entre el rasgo latente que miden los ítems y la probabilidad de acertar a éstos (Muñiz, 2010). Esta relación funcional se llama Curva Característica del Ítem (CCI). Hay una variedad de formas posibles funcionales y, por lo tanto, de modelos TRI en la literatura psicométrica (e.g. modelo logístico de un parámetro o modelo de Rasch, modelo de dos parámetros, modelo de tres parámetros).

En consonancia con la mayoría de las evaluaciones internacionales, las propias ONE desde 2005 y Aprender 2016, las pruebas Aprender 2017 han sido analizadas con base en la TRI. En la prueba Aprender 2016 y Aprender 2017 se optó por la logística de dos parámetros y en los pilotos de Aprender 2017 por la logística de un parámetro.

El uso de la TRI permite observar qué tan bien un ítem o un conjunto de ítems, discrimina entre estudiantes que tienen un alto y un bajo desempeño en las áreas evaluadas. Por otra parte, también se obtienen otros indicadores como el índice de dificultad y el índice de correlación biserial. Además, facilita una comparación más precisa de los resultados a través del tiempo (SEE, 2017).

Por último, dado que las pruebas criterioles tienen por objeto medir el dominio que tiene un estudiante de una determinada área o disciplina, la representatividad y relevancia de los ítems incluidos es parte fundamental del proceso. De ahí que obtener evidencias acerca de la validez de contenido de las pruebas sea un requisito clave para garantizar idoneidad de los resultados obtenidos.

Otros aspectos no menos relevantes que hacen a la fiabilidad de la prueba son los referidos a las condiciones estandarizadas de aplicación de los instrumentos. Si las condiciones de aplicación no son similares entre los diferentes grupos, los resultados no serán comparables. En este sentido, se deben considerar los tiempos de aplicación de la prueba, el entrenamiento de los aplicadores (la estandarización de las instrucciones para todos los participantes), y la facilidad para la puntuación y análisis de las respuestas obtenidas. Mientras más cuidados y sistematizados sean estos elementos, más resguardada estará la comparabilidad de los resultados obtenidos por distintos grupos de participantes.

En el próximo punto se indican las fases en la construcción de una prueba referida a criterio y qué procedimientos concretos se llevaron a cabo para el análisis de las evidencias de fiabilidad y validez de las pruebas Aprender 2017.

CONSTRUCCIÓN DE PRUEBAS REFERIDAS A CRITERIO. LAS PRUEBAS APRENDER 2017.

Existen criterios consensuados por la comunidad científica para el diseño de una prueba referida a criterios. En concreto, Hambleton et al. (1991) propusieron etapas que se resumen a continuación a la vez que se indican qué procedimientos concretos se llevaron a cabo en la construcción de las pruebas Aprender 2017.

1) Planteamientos iniciales

En esta etapa se plantean las hipótesis acerca de los propósitos y objetivos de las prue-

bas así como a qué grupos está dirigida, cómo se planificarán los tiempos y qué recursos humanos y materiales se requieren, a quiénes se convocará como expertos y cuál será la longitud y el tiempo para cada prueba.

Tabla 1
Propósito y caracterización de las pruebas Aprender 2017

Nombre	Pruebas Aprender 2017 Lengua y Matemática	Pruebas Aprender 2017 Cs. Naturales y Cs. Sociales
Propósito y objetivos	Evaluar capacidades y contenidos que debería alcanzar un alumno al fin del nivel secundario de acuerdo a los NAP	Evaluar capacidades y contenidos que debería alcanzar un alumno al fin del nivel primario de acuerdo a los NAP
Grupos	Estudiantes regulares de 5°/6° Nivel Secundario	Estudiantes regulares de 6° Nivel Primario
Cobertura	Censal	Censal
Tipo de ítems	Cerrados de opción múltiple	Cerrados de opción múltiple
Planificación	1° piloto 2° piloto Diseño gráfico	1° piloto 2° piloto Diseño gráfico
Identificar Expertos	Cuerpo colegiado de expertos de todo el país. Lectores críticos disciplinares internos. Expertos de Gestión curricular. Expertos de Institutos de Formación Docente.	Cuerpo colegiado de expertos de todo el país. Lectores críticos disciplinares internos. Expertos de Gestión curricular Expertos de Institutos de Formación Docente
Longitud y tiempo de aplicación	24 ítems 60 minutos por prueba	24 ítems 60 minutos por prueba

2) Revisión de los objetivos

En esta etapa se revisan críticamente los objetivos de evaluación, se selecciona el conjunto de objetivos, capacidades y contenidos que evaluarán las pruebas y se preparan las tablas de especificaciones de cada una.

Siguiendo estos lineamientos y con los antecedentes nacionales, las pruebas Aprender 2017 fueron construidas a partir de los Núcleos de Aprendizaje Prioritarios (NAP), los diseños curriculares jurisdiccionales y los consensos jurisdiccionales. Posteriormente y, al igual que en Aprender 2016, en función del marco general de referencia, se elaboraron las estructuras de cada una de las pruebas 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 a cada una de ellas.

La consecución de los objetivos para las pruebas Aprender 2017 estuvo a cargo del equipo pedagógico de la SEE quienes, en consulta e intercambio continuo con especialistas de otras áreas del Ministerio de Educación y del campo educativo en general, realizaron las definiciones de capacidades y contenidos, y establecieron los fundamentos teóricos sobre las que se basó la evaluación en cada disciplina.

Cabe destacar que el equipo de la SEE se encuentra conformado por docentes expertos –y actualmente, en ejercicio– en cada una de las áreas disciplinares principales (Matemática, Lengua, Ciencias Naturales y Ciencias Sociales).

A continuación se muestran las capacidades y contenidos delimitados para cada área disciplinar y año considerados para la elaboración de los ítems de las pruebas piloto Aprender 2017.

Evaluaciones en el nivel secundario

Matemática 5°/6° año de la Educación Secundaria

CAPACIDADES COGNITIVAS

a) RECONOCIMIENTO DE DATOS Y CONCEPTOS

Capacidad cognitiva de identificar datos, hechos, conceptos, relaciones y propiedades matemáticas, expresados de manera directa y explícita en el enunciado.

b) RESOLUCIÓN DE SITUACIONES EN CONTEXTOS INTRAMATEMÁTICOS Y/O DE LA VIDA COTIDIANA

Capacidad cognitiva de solucionar situaciones problemáticas contextualizadas, presentadas en contextos que van desde los intramatemáticos hasta los de la realidad cotidiana.

c) COMUNICACIÓN EN MATEMÁTICA

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.

Tabla 2

Tabla criterial Matemática 5°/6 año de la Educación Secundaria

BLOQUES	CONTENIDOS
NÚMEROS	Números racionales. Operaciones. Orden. Números irracionales. Representación en la recta. Operaciones.
FUNCIONES	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	Ecuaciones de 1° y 2° grado. Ecuaciones logarítmicas y exponenciales. Sistemas de dos ecuaciones lineales. Inecuaciones lineales.
GEOMETRÍA Y MEDIDA	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	Gráficos de barras, cartesiano, circular, pictogramas, cuadros, tablas. Frecuencia. Medidas de tendencia central. Variaciones, permutaciones y combinaciones. Probabilidad simple aplicada a situaciones numéricas.

Lengua en 5°/6° año de la Educación Secundaria

CAPACIDADES COGNITIVAS

a) EXTRAER:

Localizar información en una o más partes de un texto. Los lectores deben revisar, buscar, localizar y seleccionar la información. Deben cotejar la información proporcionada en la pregunta con información literal o similar en el texto y utilizarla para encontrar la nueva información solicitada.

b) INTERPRETAR:

Reconstruir el significado global y local; hacer inferencias desde una o más partes de un texto. Los lectores deben identificar, comparar, contrastar, integrar información con el propósito de construir el significado del texto.

c) REFLEXIONAR Y EVALUAR:

Relacionar un texto con su propia experiencia, conocimientos e ideas. Los lectores deben distanciarse del texto y considerarlo objetivamente. Deben utilizar conocimiento extra-textual (la propia experiencia, elementos proporcionados por la pregunta, conocimiento del mundo, conocimiento de la lengua, conocimiento de distintos géneros discursivos). Los lectores deben justificar su propio punto de vista.

Los contenidos evaluados en Lengua corresponden a cada una de las capacidades cognitivas para 5º/6º año.

Tipos de textos evaluados: cuentos breves de autores consagrados y textos expositivos o argumentativos tales como columnas de opinión, ensayos breves y textos académicos o de divulgación científica provenientes de revistas especializadas o manuales

Tabla 3

Tabla criterial Lengua 5º/6º año de la Educación Secundaria

BLOQUES	CONTENIDOS
EXTRAER	Información explícita en texto literario y no literario. Secuencia en texto literario y no literario. Resumen
INTERPRETAR	Tema en texto literario y no literario. Relaciones textuales. Procedimientos de cohesión. Elementos de enunciación en textos argumentativos: ensayos breves y notas especializadas. Características de personajes. Vocabulario. Información inferencial. Relación texto-paratexto.
REFLEXIONAR Y EVALUAR	Estructura textual. Recursos literarios y retóricos. Tipos de narradores. Tipologías y géneros discursivos.

Evaluaciones en el nivel primario

Ciencias Naturales 6° año de la Educación Primaria

CAPACIDADES COGNITIVAS

a) RECONOCIMIENTO DE CONCEPTOS

Incluye la identificación e interpretación de conceptos propios de las Ciencias Naturales. Involucra reconocer y distinguir características, identificar relaciones causa-efecto, relacionar explicaciones de fenómenos naturales con modelos científicos, clasificar y comparar.

b) COMUNICACIÓN

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). Incluye también las competencias lectoras en relación con el patrón temático y lingüístico propio de las Ciencias Naturales.

c) ANÁLISIS DE SITUACIÓN

Comprende la identificación, interpretación y análisis de evidencias, conclusiones, procesos y metodologías 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.

Tabla 4

Tabla criterial Ciencias Naturales 6° año de la Educación Primaria

BLOQUES	CONTENIDOS
SERES VIVOS	BLOQUES CONTENIDOS Estructuras y funciones: sostén. Estructuras y funciones: locomoción. Estructuras y funciones: nutrición. Estructuras y funciones: relación. Estructuras y funciones: reproducción. Célula. Diversidad: características, clasificación. Organismo-medio: adaptaciones morfofisiológicas. Organismo-medio: relaciones tróficas. Organismo-medio: las personas y el ambiente. Salud: composición química de los alimentos y dieta. Salud: prevención y cuidado.

LOS MATERIALES Y SUS CAMBIOS	<p>Materiales: clasificación.</p> <p>Materiales: conducción térmica, eléctrica.</p> <p>Materiales: magnetismo.</p> <p>Materiales: estados de agregación.</p> <p>Mezclas: clasificación.</p> <p>Mezclas: métodos de separación.</p> <p>Mezclas: soluciones.</p> <p>Cambios químicos: combustión y oxidación.</p> <p>Cambios químicos y físicos: reconocimiento.</p>
MUNDO FÍSICO	<p>Fuerzas y movimiento.</p> <p>Energía: formas, fuentes, clasificación.</p> <p>Energía: transformaciones.</p> <p>Energía: electricidad.</p> <p>Sonido: definición, características.</p> <p>Sonido: propagación, proceso de audición.</p> <p>Luz: reflexión, refracción, espejos.</p> <p>Luz: características, definición.</p> <p>Calor y temperatura: equilibrio térmico.</p>
TIERRA Y UNIVERSO	<p>Subsistemas terrestres: geósfera.</p> <p>Subsistemas terrestres: hidrósfera.</p> <p>Subsistemas terrestres: atmósfera.</p> <p>Subsistemas terrestres: biósfera.</p> <p>Sistema Solar: características del planeta Tierra.</p> <p>Sistema Solar: traslación, rotación, eclipses.</p>

Ciencias Sociales 6° año de la Educación Primaria

CAPACIDADES COGNITIVAS

a) RECONOCIMIENTO DE HECHOS Y DATOS

Contempla identificar hechos y datos en un conjunto de información mediante la utilización de los conocimientos que el estudiante posea.

b) RECONOCIMIENTO DE CONCEPTOS

Incluye 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.

c) INTERPRETACIÓN / EXPLORACIÓN

Comprende obtener y cruzar información explícita o implícita a partir de la lectura comprensiva de distintos tipos de fuentes (textos, imágenes, mapas, gráficos, tablas, etc.).

d) ANÁLISIS DE SITUACIONES

Incluye reconocer distintos tipos de relaciones -causales, de comparación, de contemporaneidad, de simultaneidad- o de seleccionar cursos de acción que requieren la aplicación de conceptos, hechos, datos o procesos previamente adquiridos.

Tabla 5

Tabla criterial Ciencias Sociales 6° año de la Educación Primaria

BLOQUES	CONTENIDOS
LAS SOCIEDADES Y LOS ESPACIOS GEOGRÁFICOS	Localización y representación del espacio geográfico. Espacios rurales y urbanos. Composición y dinámica de la población. Ambientes y problemas ambientales. Actividades económicas.
LAS SOCIEDADES A TRAVÉS DEL TIEMPO	Periodización de la historia nacional y latinoamericana. Sociedades indígenas: cazadores, recolectores y agricultores. Expansión ultramarina europea del siglo XV. Conquista y colonización de América. Revoluciones hispanoamericanas. Distintos proyectos de país en la primera mitad del siglo XIX. El proceso de construcción del Estado Nacional. Argentina aluvional.
LAS ACTIVIDADES HUMANAS Y LA ORGANIZACIÓN SOCIAL	Forma argentina de gobierno. División de poderes. Niveles político-administrativos. Soberanía argentina. Constitución Nacional. Participación ciudadana. Instituciones sociales y políticas. Normas básicas. Diversidad y multiculturalidad. Representaciones y prácticas sociales. Dignidad humana. Derechos Humanos. Responsabilidad estatal. Medios de comunicación masiva y Tecnologías de la Información y la Comunicación (TIC).

3) Redacción de los ítems

En esta fase se construyen y editan ítems con el objetivo de probarlos en el estudio piloto posterior. Los ítems incluidos en las pruebas de criterio pueden ser de respuesta abierta o ítems de respuesta cerrada y, en ambos casos, la cuestión esencial es delimitar con claridad el dominio y el criterio sobre la que se basa la prueba. Los ítems de respuesta abierta requieren que el estudiante construya o elabore la respuesta, mientras que los ítems de respuestas cerrada obligan a la selección de una o varias opciones de entre las propuestas.

Las evaluaciones educativas de gran escala suelen utilizar ítems de respuesta cerrada dado que su aplicación y corrección es más sencilla y más rápida (Covacevich, 2014). Además, las evaluaciones que utilizan ítems de opción múltiple pueden ser corregidas

mediante el uso de lectores ópticos, lo que no solo agiliza los procesos de corrección sino que también implica menores costos y evita las diferencias de criterio que podrían existir entre diferentes correctores para los ítems de respuestas abiertas (Livingston, 1999).

En el caso de las pruebas Aprender 2017 se utilizaron ítems de elección múltiple para las evaluaciones de Matemática y Lengua en 5°/6° año de la Educación Secundaria y de Ciencias Naturales y Sociales de 6° año de la Educación Primaria. Estos ítems están compuestos por un enunciado y cuatro opciones de respuesta predeterminadas, de las cuales solo una es la respuesta correcta (ver tabla 6).

Tabla 6
Tipos de ítems en las evaluaciones Aprender 2017

NOMBRE	AÑO Y NIVEL	TIPO DE ÍTEMS
Aprender 2017 Matemática	5°/6° nivel secundario	Ítems cerrados, opción múltiple
Aprender 2017 Lengua	5°/6° nivel secundario	Ítems cerrados, opción múltiple
Aprender 2017 Ciencias sociales	6° nivel primario	Ítems cerrados, opción múltiple
Aprender 2017 Ciencias naturales	6° nivel primario	Ítems cerrados, opción múltiple

Para la elaboración y redacción de los ítems se contrataron docentes externos, especialistas en las distintas áreas y años a evaluar. Algunos de los docentes externos contaban con experiencia previa en tareas similares y todos participaron de una capacitación en construcción de ítems dictada en la SEE a cargo del equipo pedagógico y los expertos en psicometría. A su vez también se les brindó una guía para la elaboración de ítems y la tabla de especificaciones de cada área y año. Por último, se indicó que los ítems redactados debían cubrir el continuo de los niveles de dificultad de la prueba: nivel Por debajo del nivel Básico, nivel Básico, nivel Satisfactorio y nivel Avanzado. La figura 1 ilustra la secuencia que se siguió para la elaboración de los ítems.

ENFOQUE DE LA EVALUACIÓN

- Diferencias entre pruebas referidas a normas y a criterios
- Criterios psicométricos para la elaboración de ítems

DEFINICIÓN DE LOS OBJETIVOS DE EVALUACIÓN

- Marco teórico
- Tablas de especificaciones/ determinación de dominios (objetivos, especificaciones y respuestas esperadas por nivel)

PAUTAS PARA LA ELABORACIÓN DE ÍTEMS

- Lineamientos internacionales para la elaboración de ítems en pruebas referidas a criterio
- Unidimensionalidad de los ítems
- Noción de coherencia, relevancia, suficiencia y claridad
- Niveles de dificultad de los ítems

Figura 1. Secuencia en la capacitación de la elaboración de los ítems.

4) Evaluación de la validez de contenido

Tal como se señaló, en una prueba referida a criterio es crucial el análisis de las evidencias de validez de contenido. Ello implica la identificación de un grupo de expertos en el contenido de las pruebas y en psicometría, el examen de la concordancia entre ítems, contenidos y objetivos de evaluación, la adecuación técnica y la revisión global de los ítems. A continuación se indican los procedimientos seguidos para el análisis de las evidencias de validez de contenido de las pruebas Aprender 2017. Cabe indicar que todos los actores involucrados en cada una de las etapas de construcción de la prueba firmaron acuerdos de confidencialidad a fin de garantizar la fiabilidad de la prueba.

Pruebas Lengua y Matemática (5°/6° nivel secundario) y de Ciencias Naturales y Ciencias Sociales (6° nivel primario)

Primera ronda de juicio de expertos

Una vez elaborados los ítems, y habiendo sido supervisados por el equipo pedagógico de la SEE, se procedió a la revisión por parte de: a) expertos en psicometría; b) lectores críticos y; c) gestión curricular. La figura 2 ilustra el circuito de esta primera ronda de juicio de expertos.

Los expertos en psicometría verificaron que la estructura del ítem fuera adecuada. Se examinó que la longitud de los enunciados fuera acorde a los parámetros recomendados, que las opciones de respuesta fueran similares en su nivel de dificultad, que fueran factibles, entre otros aspectos.

Por su parte, los lectores críticos y gestión curricular analizaron la claridad del enunciado y el grado de ajuste del ítem a los contenidos y capacidades enunciadas en los NAP. En cada una de estas instancias, se hicieron sugerencias para la mejora de los ítems que conformaron la versión final de las pruebas piloto 1 y 2.

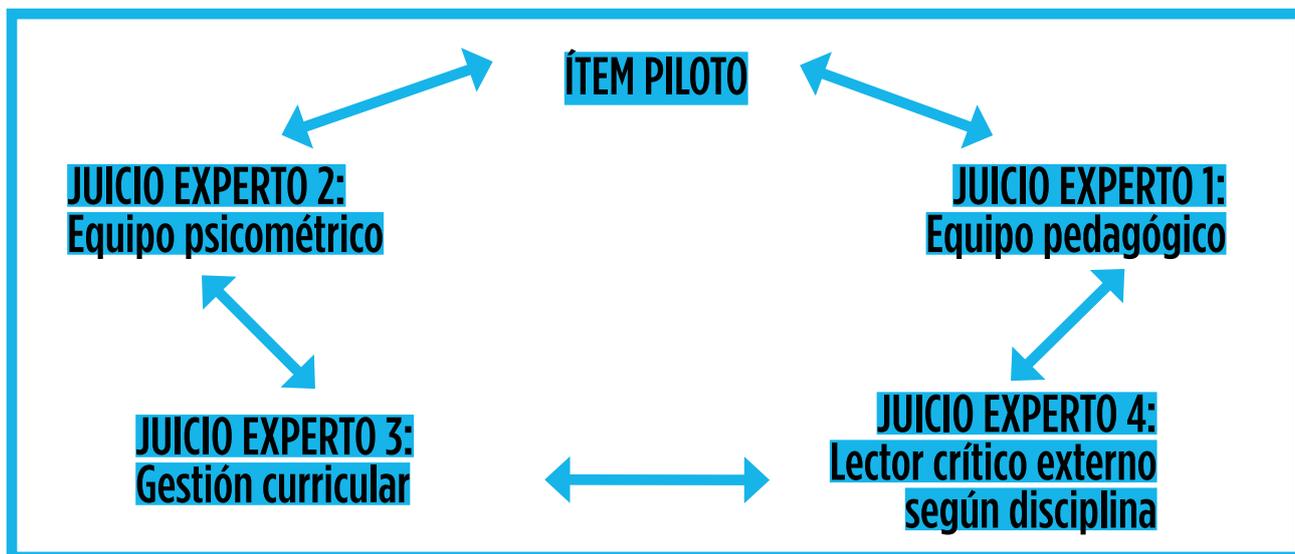


Figura 2. Juicio de expertos en la elaboración de los ítems.

Segunda ronda de juicio de expertos

Con el objetivo de fortalecer la validez de contenido de la prueba, se convocó a una “Jornada de Evaluación Federal de Lectores Críticos” que se llevó a cabo en mayo de 2017 en dependencias de la SEE.

En esta reunión, participaron 28 docentes de nivel primario y secundario en ejercicio con formación en el área de evaluación educativa pertenecientes a las jurisdicciones de CABA, Córdoba, La Pampa, Río Negro, Santa Fe y Salta.

Este equipo interdisciplinario trabajó sobre los ítems de primaria y secundaria, realizando una lectura crítica y aportando sugerencias para el fortalecimiento del carácter federal de la prueba. Se examinó la redacción de los ítems, las opciones y claves de repuesta, los contenidos y capacidades evaluados.

Se capacitó a los participantes para que dieran su juicio considerando las siguientes características: a) Claridad: si el ítem se comprende fácilmente, si es adecuado sintácticamente y semánticamente; b) Coherencia: relación lógica con la dimensión o indicador que está midiendo; c) Relevancia: análisis acerca de la importancia para la medición del desempeño y d) Suficiencia: análisis acerca de si los ítems bastan para obtener la medición y se los instruyó para que completasen una planilla en el que cada característica era evaluada en una escala de cuatro puntos (*nula, poca, bastante o mucha*).

Los juicios realizados por los docentes en la planilla de registro se volcaron a una base de datos Excel y SPSS. Luego de ello, se calcularon la cantidad y porcentajes de acuerdos (Ta) y desacuerdos (Td) entre jueces. Sobre esta base, se calculó el coeficiente de concordancia (b) entre ellos. Los resultados se interpretaron de acuerdo con los criterios de Martínez Arias (1995) y, como conclusión general, se observó que la mayor parte de los ítems obtuvieron coeficientes de acuerdo iguales o mayores al valor aceptable de 0.70.

Sobre la base de este estudio de juicio de expertos se revisaron, modificaron o reelaboraron los ítems observados. De esta manera los resultados de la Jornada de Evaluación Federal de Lectores Críticos proveyeron de un insumo fundamental para el logro de cada vez mejores pruebas Aprender. A partir de estos datos, se conformaron las versiones finales de la prueba piloto 1 y 2 de Aprender 2017.

5) Estudios piloto

Como ya se ha mencionado, el objetivo de los estudios piloto es analizar en forma empírica el comportamiento de los ítems que han sido elaborados bajo estándares consensuados y sometidos, previamente, a juicio de expertos. Los estudios piloto permiten identificar ítems débiles o defectuosos, determinar la dificultad de cada uno, conocer la capacidad para discriminar entre desempeños altos y bajos, fijar qué ítems conformarán el test, el tiempo promedio de resolución y calcular las intercorrelaciones entre ítems para examinar la homogeneidad de la prueba.

Para garantizar que los resultados del estudio piloto sean confiables, es necesario que la capacitación de los aplicadores sea uniforme y que la muestra del estudio piloto sea representativa de la población en estudio.

Capacitación de los aplicadores del estudio piloto de las pruebas de Lengua y Matemática (5°/6° año nivel secundario) y Ciencias Naturales y Ciencias Sociales (6° grado nivel primario)

Con el objetivo de garantizar las condiciones estandarizadas de aplicación de las pruebas piloto de Aprender 2017, la SEE –en colaboración con las jurisdicciones implicadas–, llevó a cabo diferentes jornadas de capacitación para los aplicadores. En dichas jornadas se presentaron los lineamientos generales del operativo, se familiarizó a los aplicadores con los materiales y con el formato de la prueba y se brindaron las pautas generales a mantener durante la aplicación (e.g. instrucciones y tiempo para completar la prueba). Dichas pautas fueron a su vez entregadas por escrito a través de una guía paso a paso y un documento elaborado a tal fin.

Las capacitaciones de aplicadores Aprender 2017 incluyeron información acerca de los objetivos, importancia del resguardo y confidencialidad de la evaluación, características generales de las pruebas, lineamientos para las actividades con los directores de las escuelas (presentación, selección de la sección, acuerdo para tomar decisiones respecto de cuestiones prácticas de la evaluación –por ejemplo, cómo proceder con los alumnos que terminan antes las pruebas; con los alumnos en integración y con los docentes o referentes jurisdiccionales para permanecer en el aula durante la evaluación), lineamientos para las actividades con los estudiantes (presentación, orden de las evaluaciones, tiempo, uso de la calculadora, apagado y guardado de celulares y dispositivos electrónicos, materiales) y relación con los coordinadores de la SEE.

De esta manera se garantizó que ambos pilotos fueran realizados bajo las mismas condiciones por todos los estudiantes que participaron en la evaluación. Por otra parte, todos los alumnos participantes tenían los mismos instrumentos para responder los cuestionarios (ej. lápiz, goma), la misma cantidad de tiempo, y el mismo tiempo de intervalo entre las diferentes áreas evaluadas.

Estudio piloto de las pruebas Aprender 2017 de Lengua y Matemática (5°/6° año nivel secundario) y Ciencias Naturales y Ciencias Sociales (6° grado nivel primario)

Con el objetivo de analizar el comportamiento psicométrico de los ítems elaborados para las pruebas Aprender 2017 se llevaron a cabo dos estudios piloto: el primero el 31 de mayo y 1 de junio y, el segundo, el 28 y 29 de junio de 2017. En cada uno, se pusieron

a prueba distintos conjuntos de ítems de las cuatro áreas evaluadas en los diferentes niveles.

En todos los casos las pruebas fueron realizadas en papel y completadas con lápiz. Tal como se ha señalado, la mayoría de las pruebas Aprender incluyen ítems de opción múltiple con cuatro opciones de respuesta. En las pruebas piloto, se utilizaron algunos cuadernillos que incluían 24 ítems por área y otros cuadernillos que incluían hasta 30 ítems por área. Todos los alumnos contaron con aproximadamente 60 minutos para contestar cada una de las áreas.

Piloto 1

El primer piloto se realizó los días 31 de Mayo y 1 de Junio de 2017. Se evaluaron 7427 estudiantes pertenecientes a 430 escuelas de 430 secciones distribuidas a lo largo de la provincia de Buenos Aires, CABA, Córdoba, Jujuy y Salta. Las áreas y años incluidos en la prueba fueron: Matemática y Lengua en 5°/6° año de la Educación Secundaria y Ciencias Naturales y Ciencias Sociales en 6° año de la Educación Primaria.

Tabla 7
Participación prueba piloto 1

Piloto parte 1	Escuelas	Secciones	Estudiantes
Primaria (6° grado)	148	148	2.939
Secundaria (5°/6° año)	282	282	4.488
Totales	430	430	7.427

23

En Matemática en 5°/6° año de la Educación Secundaria se evaluaron 8 modelos conformados por 2 bloques de 12 ítems cada uno (8 bloques - 8 modelos). La prueba piloto se conformó de 96 ítems en total distribuidos en 8 modelos de 24 ítems cada uno.

En Lengua en 5°/6° año de la Educación Secundaria se evaluaron 8 modelos conformados por 2 bloques de 15 ítems cada uno (8 bloques - 8 modelos). La prueba piloto se conformó de 120 ítems en total, distribuidos en 8 modelos de 30 ítems cada uno (4 textos literarios y 4 textos no literarios que se combinan en 8 modelos diferentes).

En Ciencias Naturales en 6° grado de la Educación Primaria se evaluaron 8 modelos conformados por 2 bloques de 12 ítems cada uno (8 bloques - 8 modelos). La prueba piloto se conformó de 96 ítems en total.

En Ciencias Sociales en 6° grado de la Educación Primaria se evaluaron 8 modelos conformados por 2 bloques de 12 ítems cada uno (8 bloques - 8 modelos). La prueba piloto se conformó de 96 ítems en total.

Piloto 2

El segundo piloto fue llevado a cabo el 28 y 29 de Junio de 2017. Fueron evaluados 2.637 estudiantes pertenecientes a 137 escuelas de 137 secciones distribuidas a lo largo de las provincias de Corrientes, Jujuy, La Rioja, Mendoza, Río Negro. Se evaluaron los mismos

años y áreas que en el primer piloto.

Tabla 8
Participación prueba piloto 2

Piloto parte 2	Escuelas	Secciones	Estudiantes
Primaria (6° año)	69	69	1.327
Secundaria (5°/6° año)	68	68	1.310
Totales	137	137	2.637

En Matemática en 5°/6° año de la Educación Secundaria se evaluaron 8 modelos conformados por 2 bloques de 15 ítems cada uno (8 bloques- 8 modelos), constituyendo una prueba piloto de un total de 120 ítems.

En Lengua en 5°/6° año de la Educación Secundaria se evaluaron 8 modelos conformados por 2 bloques de 15 ítems cada uno (8 bloques-8 modelos), constituyendo una prueba piloto de un total de 120 ítems distribuidos en 8 modelos de 30 ítems cada uno (4 textos literarios y 4 textos no literarios que se combinan en 8 modelos diferentes).

En Ciencias Naturales en 6° año de la Educación Primaria se evaluaron 6 modelos conformados por 2 bloques de 15 ítems cada uno (6 bloques - 6 modelos), constituyendo una prueba piloto de un total de 90 ítems.

En Ciencias Sociales en 6° año de la Educación Primaria se evaluaron 6 modelos conformados por 2 bloques de 15 ítems cada uno (6 bloques - 6 modelos), constituyendo una prueba piloto de un total de 90 ítems.

Análisis de los resultados de las pruebas piloto

Como se indicó en los documentos técnicos (SEE, 2017) el enfoque general para el análisis toma en cuenta la TRI, específicamente el modelo de dos parámetros que permite, para cada ítem, estimar no sólo su dificultad sino también su capacidad de discriminación. Este modelo de dos parámetros requiere de un mayor número de casos que el de un parámetro. Por ello y dado que se cuenta con una muestra para las pruebas pilotos Aprender 2017 (mientras que en la prueba definitiva Aprender 2017 se trata de un estudio censal), para el procesamiento de los datos del piloto, se utilizó un modelo de un solo parámetro, mientras que para el procesamiento de los datos de la prueba definitiva se utilizó un modelo de dos parámetros.

Los datos del estudio piloto fueron procesados por el equipo metodológico que se encargó del diseño y análisis de datos. Se calcularon los índices de dificultad relativa de los reactivos y, con el objeto de analizar la homogeneidad de los ítems en su conjunto, el coeficiente de correlación biserial.

Para el análisis de los resultados de las cuatro áreas (Matemática y Lengua 5°/6° año de secundaria, y Ciencias Naturales y Ciencias Sociales 6° grado de primaria) se consideró la distribución del nivel de dificultad de los ítems (que todo el continuo de dificultad estuviera adecuadamente representado), la correlación biserial y la distribución de las respuestas en las cuatro opciones correspondientes del cada ítem.

En cuanto al coeficiente de correlación biserial, en todos los casos se consideró como indicador psicométrico satisfactorio un valor ≥ 0.20 . Aunque para eliminar un ítem se consideró la correlación biserial en conjunto con otros indicadores psicométricos (e.g. el índice de dificultad del ítem). En función de ello, se seleccionaron los ítems con mejor comportamiento psicométrico y se eliminaron los ítems anómalos, vale decir aquellos que mostraran valores inadecuados en los indicadores paramétricos- cuidando que el contenido específico del ítem eliminado quedara representado por otros ítems en la totalidad de la prueba y que la dificultad del ítem eliminado no generara una "laguna" en el rango de dificultad de la prueba.

En los histogramas del Anexo se observa la distribución de los ítems según nivel de dificultad para cada una de las áreas y años evaluados en la prueba definitiva Aprender 2017. También se muestra la correlación entre la dificultad de los ítems en el piloto y la obtenida en la prueba definitiva.

6) Revisión de las pruebas y preparación de las formas definitivas de las pruebas Aprender 2017

A partir de los resultados de las pruebas piloto se seleccionaron los ítems que conformarían las pruebas Aprender 2017 para cada una de las áreas y año evaluados. En la selección de los ítems se tuvieron en cuenta los siguientes criterios:

- Aprobación de especialistas del equipo pedagógico de la SEE
- Aprobación de lectores críticos internos
- Aprobación por juicio de expertos externo y federal
- Aprobación por especialistas en gestión curricular
- Aprobación por especialistas en psicometría
- Indicadores psicométricos satisfactorios

A continuación se presentan las tablas de especificaciones en donde se detallan la cantidad y porcentaje de ítems según capacidad y contenidos para cada una de las áreas y año evaluados.

Tabla 9

Tabla de especificación de Matemática 5°/6° año de la Educación Secundaria

Capacidades/ Contenidos	Reconocer conceptos	Resolver problemas	Comunicación en matemática	Total
Números	Ítems: 5	Ítems: 6	Ítems: 5	16-ítems 22%
Funciones	Ítems: 2	Ítems: 9	Ítems: 7	18 ítems 25%
Ecuaciones e Inecuaciones	Ítems: 4	Ítems: 4	Ítems: 6	14 ítems 20%
Geometría y Medida	Ítems: 4	Ítems: 10	Ítems: 1	15 ítems 21%
Estadística y Probabilidad		Ítems: 7	Ítems: 2	9 ítems 12%
Total	15 ítems 21%	36 ítems 50%	21 ítems 29%	72 ítems 100%

Tabla 10

Tabla de especificación de Lengua 5°/6° año de la Educación Secundaria

Capacidades/ Contenidos	Extraer información explícita	Interpretar información sugerida	Reflexionar y evaluar sobre aspectos textuales	Total
Aspectos globales del texto (tema, secuencia, narrador, etc.)	Ítems: 4	Ítems: 17	Ítems: 12	33 ítems 46,5%
Aspectos locales del texto (cohesión, vocabulario, información explícita, etc.)	Ítems: 14	Ítems: 22	Ítems: 3	39 ítems 53,5%
Total	18 ítems 24,5%	39 ítems 54,5%	15 ítems 21%	72 ítems 100%

Tabla 11

Tabla de especificación de Ciencias Naturales 6° año de la Educación Primaria

Capacidades/ Contenidos	Reconoci- miento	Comunica- ción	Análisis de situación	Total
Seres vivos	Ítems: 8	Ítems: 7	Ítems: 18	33 ítems 46%
Materiales y sus cambios	Ítems: 2	Ítems: 2	Ítems: 10	14 ítems 19%
Mundo físico	Ítems: 4	Ítems: 1	Ítems: 7	12 ítems 17%
Tierra y universo	Ítems: 3	Ítems: 6	Ítems: 4	13 ítems 18%
Total	17 ítems 24%	16 ítems 22%	39 ítems 54%	72 ítems 100%

Tabla 12

Tabla de especificación de Ciencias Sociales 6° año de la Educación Primaria

Capacidades/ Contenidos	Reconocer hechos	Reconocer conceptos	Análisis de situación	Interpreta- ción	Total
Historia	Ítems: 3	Ítems: 2	Ítems: 4	Ítems: 15	24 ítems 33,33%
Geografía	Ítems: 3	Ítems: 4	Ítems: 5	Ítems: 12	24 ítems 33,33%
Educación Ciu- dadana	Ítems: 3	Ítems: 5	Ítems: 4	Ítems: 12	24 ítems 33,33%
Total	9 ítems 12,5%	11 ítems 15,25%	13 ítems 18,05%	39 ítems 54,16%	72 ítems 100%

7) Análisis del funcionamiento de los ítems de las pruebas finales Aprender 2017

Una vez recibida la base de datos con el conjunto de respuestas de Aprender 2017, se realizó 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 a_i, b_i y θ_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 estimados (Muñiz, 1997).

Una vez hecho el primer cálculo de los parámetros de los ítems se debe evaluar cuáles no funcionaron correctamente y conviene eliminar. Para ello, se evalúa la dificultad y la discriminación de los ítems entre otros indicadores. 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.

CONCLUSIONES

La evaluación educativa tiene el propósito principal de otorgar datos que contribuyan al fortalecimiento de las prácticas de enseñanza y a la planificación de estrategias de mejora de los aprendizajes. Para cumplir con este objetivo es menester contar con instrumentos válidos y confiables y documentación que respalde y dé cuenta de los procesos que han sido llevados a cabo para obtener dichos instrumentos y verificar su funcionamiento.

Cada proceso de evaluación educativa involucra distintas etapas. Una de las primeras y fundamentales es el diseño y construcción de las pruebas con las que se evaluará a los estudiantes. Como todo instrumento de medida, las pruebas tienen que pasar por controles de calidad en cuanto a su construcción para garantizar su uso idóneo. Las evidencias en cuanto a la validez y fiabilidad son cruciales ya que de ello depende que las interpretaciones y conclusiones llevadas a cabo en un análisis posterior sean representativas del dominio evaluado y sustantivos para la descripción del desempeño de los estudiantes.

Este documento tiene como propósito esencial comunicar cómo han sido construidas las pruebas Aprender 2017, brindar algunas evidencias acerca de su validez y confiabilidad y dar a conocer todos los recaudos y consideraciones metodológicas que se han tenido en cuenta en el proceso de construcción y diseño de la prueba a fin de garantizar resultados significativos y psicométricamente adecuados.

Entendemos que la mejora de los procesos de aprendizaje y de enseñanza depende de la articulación y del trabajo conjunto y colaborativo de los diferentes actores y sectores del sistema. Esperamos que la información aquí brindada sirva al entendimiento de nuestra tarea y nuestro compromiso con el sistema educativo argentino.

REFERENCIAS

American Educational Research Association (AERA), American Psychological Association (APA), & National Council on Measurement in Education (NCME) (2014). "Standards for educational and psychological testing". Washinton, DC: American Psychological Association

Cortada de Kohan, N., Macbeth, G., & López Alonso, A. (2008). "Técnicas de Investigación Científica". Buenos Aires, Argentina: Lugar.

Covacevich, C. (2014). "Cómo seleccionar un instrumento para evaluar aprendizajes estudiantiles. Nota Técnica". Banco Interamericano de Desarrollo. División de Educación (SCL/EDU). Obtenido el 29 de noviembre de 2017: <https://publications.iadb.org/bitstream/handle/11319/6758/C%C3%B3mo-seleccionar-un-instrumento-para-evaluar-aprendizajes-estudiantiles.pdf>

Cronbach, L. J. (1971). "Fundamentos de la exploración psicológica". Madrid: Biblioteca Nueva.

Glaser, R. (1963). Instructional technology and the measurement of learning out-comes: Some questions. "American Psychologist, 18", 519-521.

Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). "Fundamentals of item response theory" (Vol. 2). California: Sage.

Hambleton, R.K. & Swaminathan, H. (1978). Criterion-Referenced Testing and Measurement: A review of technical issues and developments. "Review of Educational Research, 40", 1-47.

Jornet Meliá, J. M. & González Such, J. (2009). Evaluación criterial: determinación de estándares de interpretación (EE) para pruebas de rendimiento educativo. "Estudios sobre Educación, 16", 103-123.

Leones, M. (2005). "Recorrido político y técnico-pedagógico en el proceso de elaboración, justificación y validación de los criterios de evaluación. Ministerio de Educación, Ciencia y Tecnología de la Nación". Dirección Nacional de Información y Evaluación de la Calidad Educativa (DINIECE). Obtenido el 29 de noviembre de 2017: http://repositorio.educacion.gov.ar:8080/dspace/bitstream/handle/123456789/95653/evaluacion_criterial_6abril06.pdf?sequence=1

Leyva Barajas, Y. E. (2011). Una reseña sobre la validez de constructo de pruebas referidas a criterio. "Perfiles Educativos, 33" (131), 131-154.

Livingston, S. A. (2009). Constructed-Response Test Questions: Why we use them; How we score them. Obtenido el 29 de Noviembre de 2017 en: https://www.ets.org/Media/Research/pdf/RD_Connections11.pdf

Martínez Arias, M. R. (1995). "Psicometría: Teoría de los test psicológicos y educativos".

Madrid: Síntesis.

Muñiz, J. (1997). Introducción a la teoría de respuesta a los ítems. España: Pirámide.

Muñiz, J. (2010). Las teorías de los tests: teoría clásica y teoría de respuesta a los ítems. "Papeles del Psicológico, 3"(1), 57-66.

Secretaría de Evaluación Educativa (2017). "APRENDER 2016. Notas Metodológicas. Serie de Documentos Técnicos/2". Ministerio de Educación y Deportes de la Nación. Obtenido el 30 de Noviembre de 2017 en: <http://www.bnm.me.gov.ar/gigal/documentos/EL005591.pdf>

ANEXO

En las figuras 3, 4, 5 y 6 se muestra la distribución de los ítems según nivel de dificultad para cada una de las áreas y años evaluados en Aprender 2017. Se observa que el rango de ítems de dificultad baja se encuentra representado (entre -2 y -1), aunque en todos los histogramas hay una leve asimetría hacia la derecha, lo que indica una mayor cantidad de ítems de dificultad mayor.

Distribución de dificultad de los ítems - Matemática
5°/6° año de secundaria

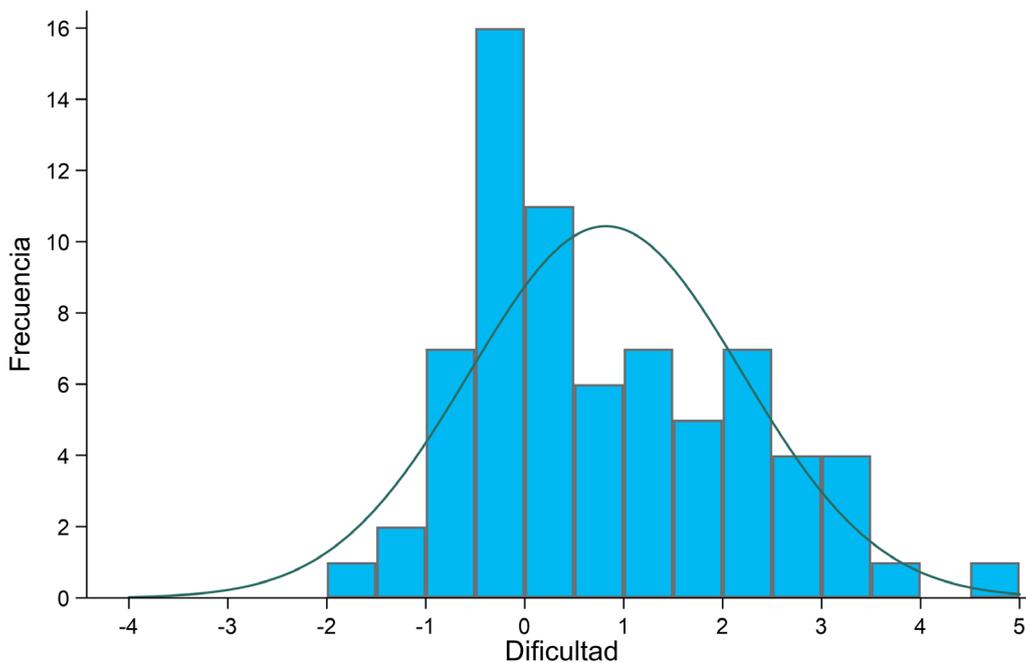


Figura 3. Histograma de la dificultad: Matemática 5°/6 año secundaria.

Distribución de dificultad de los ítems - Lengua
5°/6° año de secundaria

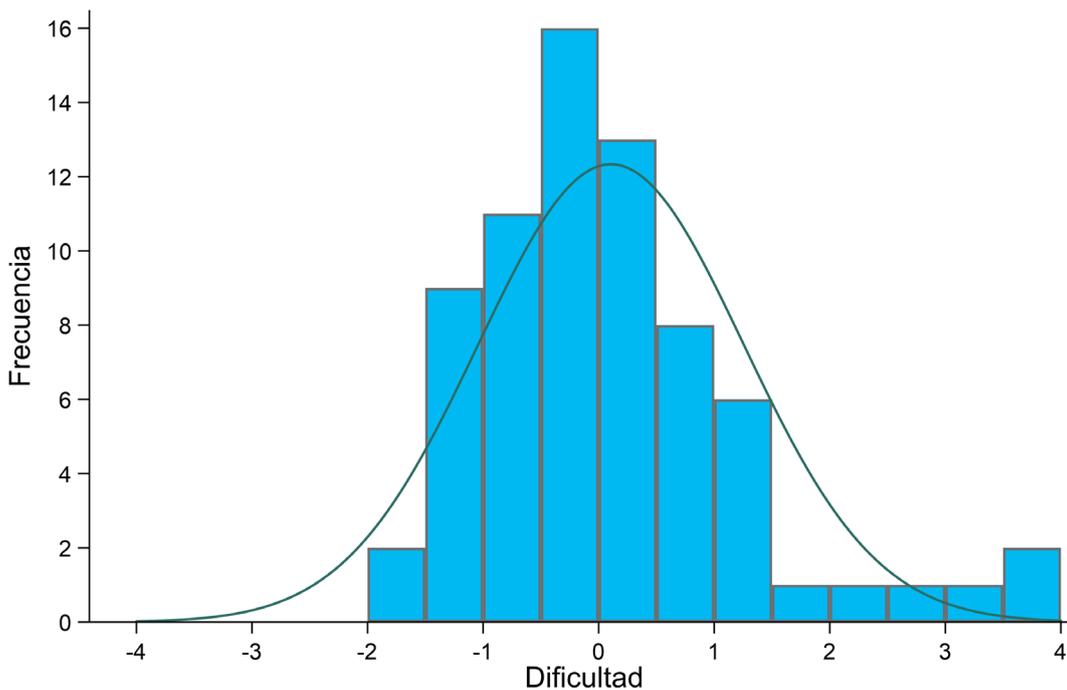


Figura 4. Histograma de la dificultad: Lengua 5°/6 año secundaria.

Distribución de dificultad de los ítems - Ciencias Naturales
6° grado de primaria

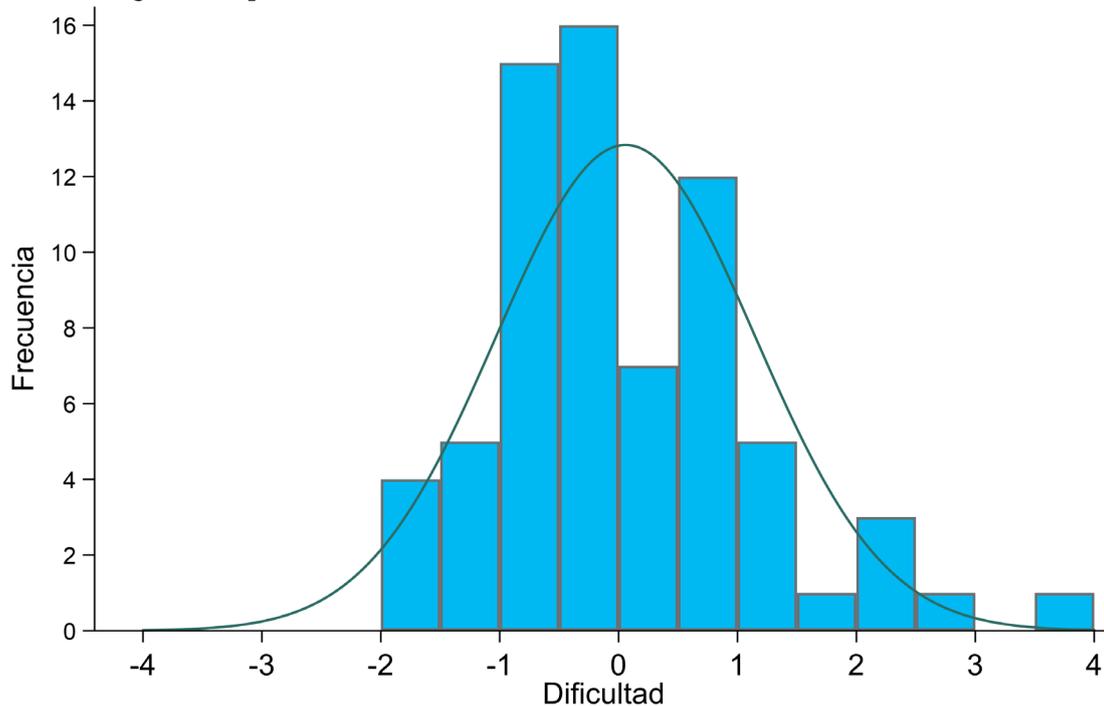


Figura 5. Histograma de la dificultad: Ciencias Naturales 6° grado de primaria.

Distribución de dificultad de los ítems - Ciencias Sociales
6° grado de primaria

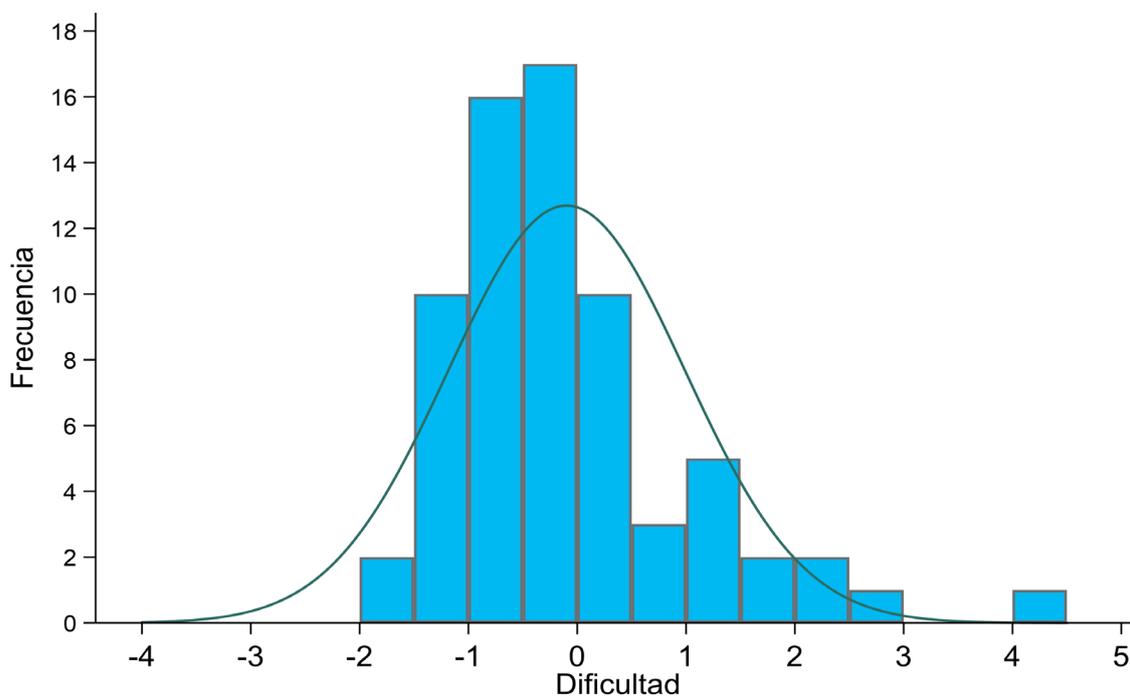


Figura 6. Histograma de la dificultad: Ciencias Sociales 6° grado primaria

En los diagramas de dispersión que se muestran a continuación, se puede observar la correlación entre la dificultad de los ítems en las pruebas piloto y la dificultad obtenida en la prueba definitiva. Estos gráficos evidencian una alta correlación en la dificultad observada por ambas evaluaciones en todas las áreas y años incluidos.

Comparación de los parámetros de dificultad entre la prueba piloto y Aprender 2017
5º/6º año- Matemática

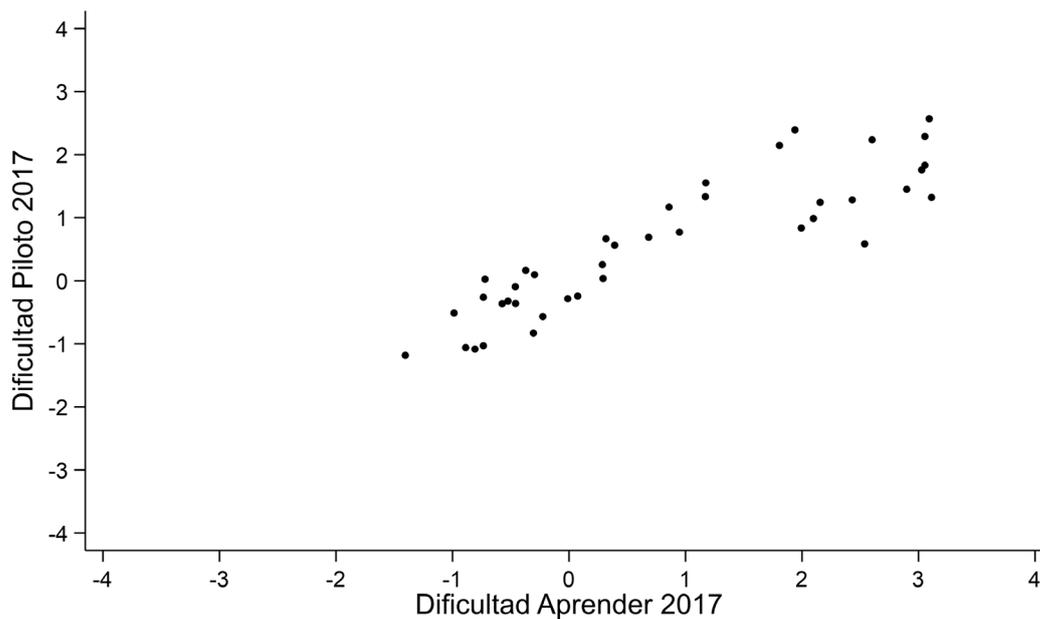


Figura 7. Comparación de la distribución de la dificultad de la prueba piloto y la definitiva de Aprender 2017 de Matemática 5º/6º año.

Comparación de los parámetros de dificultad entre la prueba piloto y Aprender 2017
5º/6º año- Lengua

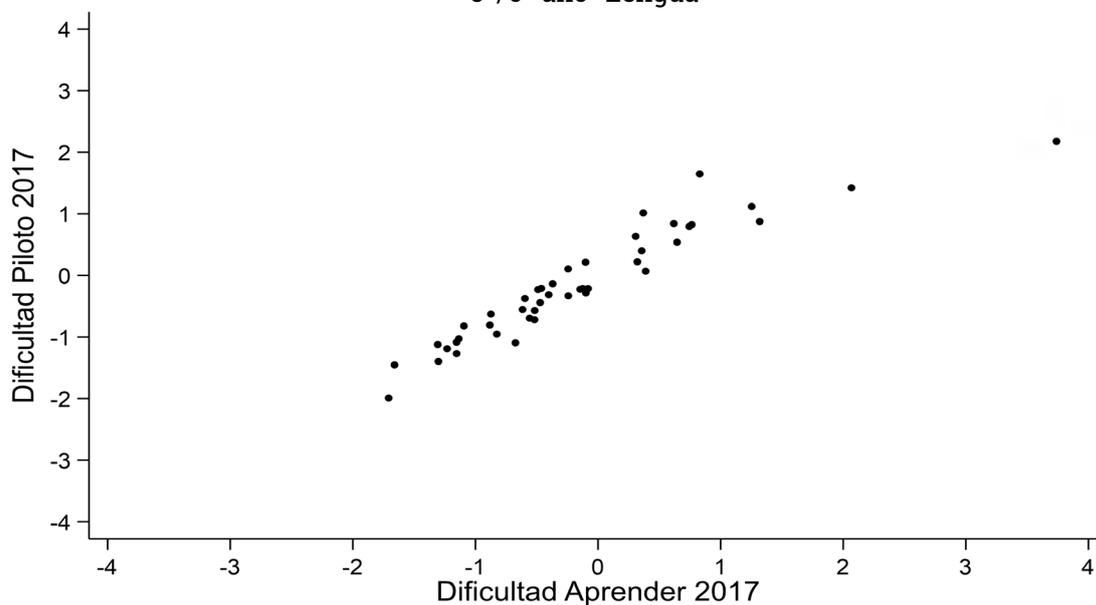


Figura 8. Comparación de la distribución de la dificultad de la prueba piloto y la definitiva de Aprender 2017 de Lengua 5º/6º año.

Comparación de los parámetros de dificultad entre la prueba piloto y Aprender 2017
6° grado - Ciencias Naturales

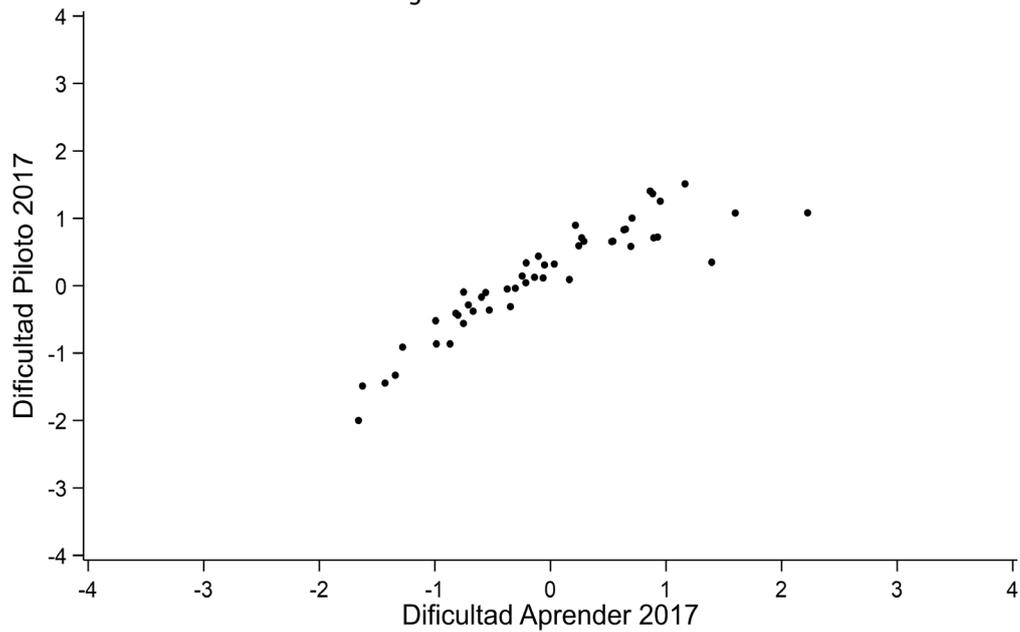


Figura 9. Comparación de la distribución de la dificultad de la prueba piloto y la definitiva de Aprender 2017 de Ciencias Naturales 6° grado.

Comparación de los parámetros de dificultad entre la prueba piloto y Aprender 2017
6° grado - Ciencias Sociales

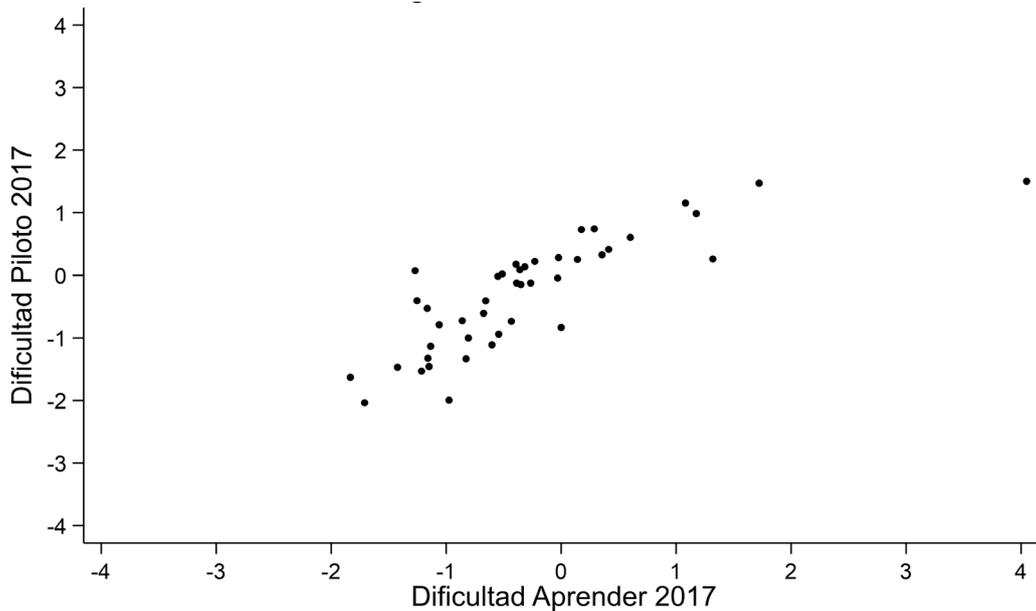


Figura 10. Comparación de la distribución de la dificultad de la prueba piloto y la definitiva de Aprender 2017 de Ciencias Sociales 6° grado.

