

ARISTAS

MARCO GENERAL DE LA EVALUACIÓN



INEEd

Instituto Nacional de
Evaluación Educativa



Aristas

Evaluación Nacional
de Logros Educativos

Comisión Directiva del INEE: Alex Mazzei (presidenta), Pablo Cayota, Alejandro Maiche, Limber Elbio Santos, Marcelo Ubal y Oscar N. Ventura.

Director ejecutivo del INEE: Mariano Palamidessi

Directora del Área Técnica: Carmen Haretche

Director de Unidad de Evaluación de Aprendizajes y Programas: Juan Martín Soca

Los autores de este documento son Carmen Haretche, Beatriz Picaroni y Juan Martín Soca. Se elaboró a partir del trabajo del equipo de investigadores de la Unidad de Evaluación de Aprendizajes y Programas: Cecilia Alonso, Vanessa Anfitti, Laura Dodino, Cecilia Emery, Verónica Figueroa, Elizabeth García, Cindy Mels, Inés Méndez, Leonardo Moreno, Matías Núñez, Darío Padula, María Eugenia Panizza, Andrea Rajchman, Pilar Rodríguez y María Eugenia Ryan.

Corrección de estilo: Federico Bentancor y Mercedes Pérez

Diseño y diagramación: Diego Porcelli

Montevideo, 2018

ISBN: 978-9974-8673-0-7

© Instituto Nacional de Evaluación Educativa (INEE)

Edificio Los Naranjos, planta alta, Parque Tecnológico del LATU

Av. Italia 6201, Montevideo, Uruguay

(+598) 2604 4649 – 2604 8590

ineed@ineed.edu.uy

www.ineed.edu.uy

Cómo citar: INEE (2018), *Aristas. Marco general de la evaluación*, INEE, Montevideo.

En la elaboración de este material se ha buscado que el lenguaje no invisibilice ni discrimine a las mujeres y, a la vez, que el uso reiterado de /o, /a, los, las, etcétera, no dificulte la lectura.

ÍNDICE

Presentación.....	4
Los componentes de la evaluación	6
El contexto familiar y el entorno escolar.....	6
La convivencia y la participación.....	8
Las habilidades socioemocionales.....	10
Las oportunidades de aprendizaje.....	12
Los desempeños de los estudiantes en lectura y matemática.....	14
El reporte de resultados.....	25
Informes.....	25
Niveles de desempeño en lectura y matemática.....	25
Confidencialidad y anonimato de la información.....	26
Evaluación en centros educativos.....	26
Anexo 1. Diseño y validación de los instrumentos de Aristas.....	29
Anexo 2. Gestión del procedimiento y capacitación a los jueces para la determinación de puntos de corte.....	41
Bibliografía.....	46

PRESENTACIÓN

En Uruguay se realizan evaluaciones nacionales estandarizadas de aprendizajes desde hace más de dos décadas. Comenzaron a realizarse en educación primaria, a partir de 1996, mientras que en 1999 se realizó una evaluación en tercero de educación media que no tuvo continuidad en el tiempo¹. Estas evaluaciones fueron realizadas hasta el año 2013 por la Administración Nacional de Educación Pública (ANEP).

En 2008 la Ley General de Educación n° 18.437 creó el Instituto Nacional de Evaluación Educativa (INEEd), que comenzó a funcionar en 2013. En su artículo 115 establece que el INEEd “tendrá como cometido evaluar la calidad de la educación nacional a través de estudios específicos y el desarrollo de líneas de investigación educativas”. De esta manera, la evaluación del sistema educativo pasó a ser una responsabilidad de un actor autónomo de la institución prestadora del servicio.

Durante las últimas dos décadas, los países de América Latina han desarrollado sistemas de evaluación centrados, fundamentalmente, en la medición de desempeños en lectura y matemática y, en muchos casos, en ciencias. Pero en los últimos años nuevas dimensiones y capacidades están comenzando a ser exploradas y medidas, ampliando el espectro de lo que es pertinente para evaluar a los sistemas educativos en función de una concepción más integral de la calidad, y atenta al derecho a la educación de los estudiantes desde la educación inicial hasta el fin de la educación media.

Es en este contexto que se construye Aristas, la Evaluación Nacional de Logros Educativos. Es una evaluación del sistema educativo con carácter multidimensional, que no reduce la noción de calidad a los puntajes obtenidos por los estudiantes en evaluaciones estandarizadas externas. Esta evaluación considera los contextos en los que opera el sistema educativo y sostiene una preocupación por medir, describir y evaluar los distintos componentes —sean recursos, procesos o resultados— que hacen posible garantizar el derecho a la educación de los niños y adolescentes.

En 2015 el INEEd comenzó las tareas de diseño e implementación de Aristas. La primera evaluación nacional en primaria se realizó en 2017 y en educación media en 2018². A partir de allí, cada tres años, Aristas generará, analizará y presentará evidencia sistemática sobre las condiciones socioeconómicas y culturales de las familias de los estudiantes, la organización y el clima de trabajo escolar, la convivencia y participación, las habilidades socioemocionales, las oportunidades de aprendizaje y los desempeños de los estudiantes en lectura y matemática.

¹ Los informes de estas evaluaciones pueden consultarse en la sección “Evaluaciones nacionales de la ANEP” del sitio web del INEEd: <https://www.ineed.edu.uy/nuestro-trabajo/aristas.html>.

² Los años previos se realizaron los pilotos de cada evaluación.



Para construir esta observación multidimensional y compleja, Aristas consulta a diversos actores de la vida escolar. El conjunto de información que brinda permitirá dar a conocer los logros del sistema educativo y los puntos de vista de padres, docentes, directores y estudiantes relativos a las diversas realidades escolares, lo que servirá para debatir y reflexionar sobre los avances logrados y los problemas que persisten, con el fin de generar insumos útiles y válidos para el diseño de futuras políticas educativas. En síntesis, Aristas procura articular distintas dimensiones y perspectivas con el objetivo de producir información confiable y útil para movilizar debates y acciones públicas que garanticen el derecho a la educación de todos los estudiantes.

Aristas se propone:

- monitorear los logros del sistema educativo en sus distintas dimensiones;
- promover una conceptualización multidimensional de la calidad de la educación, sin reducirla a los desempeños en áreas cognitivas;
- brindar información útil para evaluar la garantía del derecho a la educación de los estudiantes y del derecho de los docentes a desempeñar su trabajo profesional en condiciones adecuadas;
- contribuir al diseño de políticas educativas basadas en evidencias; y
- ser un insumo para el diseño y la mejora curricular, brindando evidencias válidas a la operacionalización de lo que se espera que los estudiantes sean capaces de hacer al culminar cada ciclo escolar.

A partir de una muestra representativa de centros educativos públicos y privados del país, tanto urbanos como rurales, sin exclusiones por tamaño del centro, son consultados:

- estudiantes de tercero y sexto de primaria y de tercero de educación media y sus padres;
- maestros de tercero y sexto de primaria y profesores de tercero de educación media; y
- directores.

La información generada por Aristas se propone alimentar los debates públicos sobre educación, brindar evidencia útil para informar la política educativa y promover la reflexión y mejora pedagógica en los centros. Se dirige a diversos públicos: decisores de política educativa, investigadores y analistas de política, docentes, familias, estudiantes y la ciudadanía en general.

Los resultados de esta evaluación:

- buscan ser de utilidad para evaluar al sistema educativo en su conjunto, desde la educación primaria hasta el fin de la educación media;
- se presentan agregados según los estratos de muestreo o características de población relevantes que favorezcan comparaciones justas; y
- no brindan datos por centro, docentes o estudiantes.



LOS COMPONENTES DE LA EVALUACIÓN

Aristas se caracteriza por ser una evaluación abarcativa de los logros educativos, que propone superar la concepción que asocia la calidad educativa con los puntajes de estudiantes estimados en lenguaje, matemática o ciencias. Se articula en torno a un conjunto comprensivo de dimensiones de análisis: el contexto y el entorno escolar; las orientaciones de los estudiantes hacia la convivencia, ciudadanía y participación; las habilidades socioemocionales de los estudiantes; las oportunidades de aprendizaje de lectura y matemática que los docentes implementan en las aulas; y los desempeños de los estudiantes en lectura y matemática.

El contexto familiar y el entorno escolar

El componente de **contexto familiar y entorno escolar** procura recoger información sobre las características de los estudiantes, docentes y centros educativos. Este componente es relevante para dar cuenta de los contextos familiares e institucionales en los que se sustentan los procesos de aprendizaje.

El subcomponente de **contexto familiar** del estudiante refiere a las variables externas a la escuela que tienen efectos sobre los procesos y resultados educativos. Las dimensiones que se consideran en este subcomponente son: el estatus socioeconómico y cultural de los estudiantes, y el clima educativo del hogar. En la medida que se trata de un factor externo al centro educativo, estas dimensiones no pueden ser conceptualizadas como un logro del sistema, sino como un factor de contextualización que permite analizar la heterogeneidad de las familias de los estudiantes uruguayos.

El subcomponente de **entorno escolar** refiere fundamentalmente a variables organizativas y de clima escolar de cada centro. Dentro de este subcomponente se consideran las siguientes dimensiones: liderazgo escolar, capacidades profesionales del equipo docente, clima de aprendizaje, relaciones de confianza, e infraestructura y recursos del centro educativo.

Adicionalmente, el componente releva distintas variables de control: características sociodemográficas de estudiantes, docentes y directores, trayectoria educativa de los estudiantes, entre otras.

Medir y analizar estos subcomponentes es necesario por dos motivos fundamentales. En primer lugar, porque identificar factores extraescolares permite realizar interpretaciones justas de los desempeños de los estudiantes (NAGB, 2003) y aporta información necesaria para analizar la variación de los logros en el tiempo y la evolución de las brechas entre



poblaciones (IEE, 2009). En segundo lugar, porque dar cuenta de factores escolares vinculados a los desempeños es un camino fundamental para orientar la política educativa y contribuir a garantizar el derecho a la educación de todos los estudiantes.

TABLA 1.

DIMENSIONES GENERALES: COMPONENTE CONTEXTO FAMILIAR Y ENTORNO ESCOLAR³

CONTEXTO FAMILIAR	
Estatus socioeconómico y cultural	Barrio en que se ubica la vivienda
	Composición del hogar
	Cantidad de perceptores de ingresos
	Tipo de atención de la salud a la que tienen acceso los miembros del hogar
	Propiedad de la vivienda
	Cantidad de baños en la vivienda
	Bienes del hogar
	Nivel educativo de los padres o tutores del estudiante
Cantidad de libros en el hogar	
Clima educativo del hogar	Hábitos de lectura en el hogar
	Espacios y recursos disponibles para el estudio dentro del hogar
	Supervisión parental del estudio
ENTORNO ESCOLAR	
Liderazgo escolar	Liderazgo pedagógico e inclusivo del director y la influencia de los docentes en la toma de decisiones en la escuela
Capacidad profesional en el centro	Se centra en la disposición profesional y la existencia de una comunidad educativa. Refiere al grado de compromiso docente con el centro y su motivación laboral, así como a la existencia de una comunidad profesional en el sentido de la colaboración entre pares, el diálogo reflexivo de los docentes y la focalización en el aprendizaje
Clima de aprendizaje centrado en el estudiante	Incluye ítems referidos a la seguridad y el orden dentro del centro educativo y a la seguridad del entorno barrial en el que se inserta el centro. Asimismo, refiere a las relaciones de personalismo con los estudiantes, así como también al comportamiento de los estudiantes en clase
Relaciones de confianza en la comunidad educativa	Percepción que los diversos actores escolares tienen sobre las relaciones de confianza con los demás. Dichas relaciones proveen la base para el surgimiento de la colaboración entre pares, el diálogo reflexivo y la responsabilidad colectiva (comunidad profesional), también impacta en las disposiciones de los profesionales hacia el centro educativo (compromiso hacia el centro y disposiciones hacia la innovación). En un clima de confianza, los docentes tendrán una mayor disposición a abandonar el aislamiento del trabajo en el aula y a trabajar juntos en pos de la mejora del aprendizaje de los estudiantes
Infraestructura y recursos	Infraestructura educativa Recursos materiales Servicios disponibles en el centro educativo Recursos humanos disponibles

³ En el anexo 1.1.1. se describen los procesos de diseño y validación del componente contexto familiar y entorno escolar.



La convivencia y la participación

El sistema educativo es un escenario clave para la formación ciudadana. Por tanto, es relevante conocer y evaluar las normas y reglas —formales e informales— que modelan las relaciones interpersonales y el clima escolar en nuestros centros educativos. También las prácticas pedagógicas y de gestión que buscan promover las habilidades, las actitudes y los valores ciudadanos, y que hacen que la institución educativa sea vivida como un espacio que fomenta la participación, la convivencia y el respeto por los derechos humanos.

El trabajo sobre la convivencia como parte de la formación de ciudadanos opera fuertemente sobre el sistema de relaciones interpersonales que se desarrollan en los centros educativos y supone una preocupación por la calidad de los lazos sociales que establecen sus actores: estudiantes, docentes, padres y equipo directivo (Colombo, 2011). Aunque se constituye como escenario y medio para el aprendizaje, constituye un fin en sí mismo (Bentancor, 2004; Viscardi y Alonso, 2013). Donde las personas conviven, los conflictos son parte inherente de la interacción; en este sentido, las formas en que se aborde el conflicto en los centros pueden constituir una herramienta para el aprendizaje. A su vez, este entramado de relaciones repercute sobre el sentido de pertenencia del estudiante entendido como base, a la vez que consecuencia de su convivencia en el centro.

La concepción del estudiante como sujeto de derecho conlleva el compromiso del sistema educativo de generar y garantizar las oportunidades para su participación, además de ser una condición para lograr la convivencia democrática en los centros. Aparte de identificar la existencia e implementación de ciertos dispositivos o estructuras destinados a promover la participación de los estudiantes y de sus familias, evaluar la participación en el sistema educativo implica aproximarse a la medida en la cual los alumnos sienten que su voz u opinión están siendo tomadas en cuenta y pueden influir en la institución educativa (Conteri y Rodríguez, 2012).

Con la inclusión de este componente se pretende conocer la percepción de los distintos actores educativos sobre la convivencia y las prácticas participativas implementadas en los centros, explorar las condiciones que aportan a ellas y los desafíos que se presentan. En segunda instancia, se busca identificar factores que promueven una mejor convivencia y una mayor participación en los centros educativos.

Para educación primaria en particular se propone considerar la dimensión práctica de la formación ciudadana en las escuelas, poniendo el foco en la convivencia y la participación. En educación media se agrega el abordaje de los derechos humanos en el centro educativo.

Algunas preguntas vinculadas con este componente son las siguientes:

- ¿cómo son las relaciones interpersonales en los centros educativos (entre estudiantes, docentes, familias) y cuál es su encuadre institucional, tomando en cuenta la perspectiva de los distintos actores involucrados?;
- ¿cómo se implementan los dispositivos y espacios de participación que propone el sistema educativo, y cómo son percibidos por los actores involucrados?;



- ¿cuáles son los factores que promueven la convivencia y participación en los centros educativos?; y
- ¿cómo inciden la convivencia y la participación en los demás componentes evaluados por Aristas?

A partir de una revisión de la literatura y la sistematización de documentos normativos y orientadores de política, publicados y divulgados por las autoridades educativas nacionales (CEIP, 2016a; ANEP, 2015 y 2017; SNEP, 2016), el INEEd seleccionó y definió distintas dimensiones y subdimensiones.

TABLA 2.

DIMENSIONES GENERALES: COMPONENTE CONVIVENCIA Y PARTICIPACIÓN⁴

CONVIVENCIA	
Las relaciones interpersonales	La percepción de los distintos actores educativos acerca de la medida en la que las relaciones brindan apoyo, confianza, aceptación y respeto
Sistema de normas y reglas	El abordaje, la construcción y difusión de las normas y reglas de convivencia en el aula y el centro educativo
Resolución de conflictos	El abordaje para la resolución de conflictos en el centro, y la aplicación de estrategias que promueven el diálogo y la mediación
Sentido de pertenencia	El vínculo afectivo que tiene el estudiante con la institución educativa, que se manifiesta en sentimientos de conexión y compromiso con el centro escolar, y el involucramiento en actividades sociales y escolares
PARTICIPACIÓN	
Dispositivos formales	Dispositivos promovidos por el sistema educativo con el objetivo de brindar espacios para incentivar la participación democrática de estudiantes y familias en el sistema educativo
Actividades socioeducativas	Actividades gestionadas por el centro educativo (culturales, didácticas, deportivas, recreativas o solidarias), que involucran a estudiantes, sus familiares, funcionarios docentes, funcionarios no docentes, equipos de dirección del centro o miembros de la comunidad
Medios de comunicación	Herramientas a través de las cuales se comunica información relevante entre los diferentes integrantes de un centro educativo
Voz del estudiante	Percepción acerca de la medida en que las acciones de los estudiantes pueden influir en las características estructurales o el funcionamiento del centro educativo
Actitudes hacia la participación	Sentimientos, expectativas y creencias acerca de la participación estudiantil y de las familias en el centro educativo y su potencial para aportar a él
ABORDAJE DE LOS DERECHOS HUMANOS EN EL CENTRO EDUCATIVO	
Trabajo de inclusión en el centro*	La medida en que se abordan en el centro temas relativos al reconocimiento de las diferencias, la defensa y promoción de los derechos humanos, y las condiciones para su implementación
Integración de la diversidad*	Las actitudes de los estudiantes hacia la igualdad de género y la aceptación de la diversidad

* Se incluye solo en educación media.

⁴ En el anexo 1.1.2. se describen los procesos de diseño y validación del componente convivencia y participación.



Las habilidades socioemocionales

Las habilidades socioemocionales son herramientas cognitivas, emocionales y sociales que tienen como función la adaptación del individuo a su entorno y que facilitan el desarrollo personal, el relacionamiento social, el aprendizaje y el bienestar.

La inclusión de las habilidades socioemocionales en la evaluación educativa implica reconocer su importancia en los procesos educativos. Bajo una nueva concepción de conocimientos y aprendizajes fundamentales, las habilidades socioemocionales surgen como objetos de trabajo educativo en sí mismas, a la vez de ser reconocidas como promotoras del bienestar de los estudiantes en el centro educativo. Por lo tanto, es de interés conocer cómo se presentan estas habilidades en la población escolar y explorar los factores que las promueven. Asimismo, se considera importante indagar su compleja relación con factores contextuales y los procesos que ocurren en el centro educativo, para entender su potencial para la mejora de los aprendizajes y la convivencia escolar.

El INEEd se propone avanzar en el conocimiento de las características y distribución de estas habilidades en la población escolar. A la vez, se espera que esta evaluación contribuya con evidencias que permitan identificar e impulsar nuevos abordajes para la promoción de dichas habilidades en nuestro sistema educativo.

En educación primaria, el componente de habilidades socioemocionales se incluye solo para los estudiantes de sexto grado, dada la complejidad de abordar estas temáticas en cuestionarios de autorreporte con niños de menor edad (tercer grado).

Este componente contribuye a dar cuenta de las siguientes cuestiones:

- ¿cómo se presentan las habilidades socioemocionales en nuestros alumnos?;
- ¿se observan variaciones en su distribución en función de ciertas características individuales, sociodemográficas, contextuales y escolares?;
- ¿cuáles son las variables a nivel individual, familiar, contextual y del entorno escolar que inciden sobre estas habilidades?;
- ¿qué habilidades socioemocionales se relacionan con una mejor convivencia en el centro educativo y cómo?; y
- ¿cómo se relacionan estas habilidades con los desempeños en matemática y lectura, considerando su compleja interacción con otros componentes?

El INEEd seleccionó ocho habilidades socioemocionales para ser incluidas en su evaluación en sexto año de primaria, y cuatro más que se incluyen en la de tercero de educación media.



TABLA 3.

DIMENSIONES GENERALES: COMPONENTE HABILIDADES SOCIOEMOCIONALES⁵**MOTIVACIÓN Y AUTORREGULACIÓN DEL APRENDIZAJE**

Autorregulación metacognitiva	La conciencia y control de actividades cognitivas a través de: planificación, monitoreo (atención, revisión y cuestionamiento) y regulación (corrección continua de las actividades cognitivas durante la ejecución de una tarea)
Mentalidad de crecimiento	La creencia fundamental de que nuestras cualidades básicas pueden ser cultivadas a través del esfuerzo
Motivación intrínseca	La participación en una tarea como un fin en sí mismo, por razones tales como el desafío, la curiosidad, el dominio
Perseverancia académica	Compromiso (con tareas académicas, actividades o la escuela como ámbito social), foco y persistencia en la persecución de metas académicas, a pesar de obstáculos, dificultades y distracciones
Autoeficacia académica*	Autovaloración de las habilidades y aptitudes para dominar una tarea académica. Se aborda tanto para matemática como para lectura
Valoración de la tarea*	Percepciones del estudiante sobre qué tan interesante, importante y útil le parece la materia. Se aborda tanto para matemática como para lectura

HABILIDADES INTERPERSONALES

Empatía	Capacidad de entender y compartir el estado emocional de otro, y responder de forma compatible con él a través de la toma de perspectiva, el reconocimiento de emociones y su contexto
Habilidades de relacionamiento	Habilidades de comportamiento, socialmente aceptables, que permiten interactuar de forma efectiva con otros

HABILIDADES INTRAPERSONALES

Regulación emocional	Estrategias cognitivas para el manejo de la información emocional interna y la regulación de la expresión emocional. Las estrategias remiten a recursos como la capacidad de desviar la atención, capacidad de reformular o tomar una nueva perspectiva en una situación emocional o cambiar las reacciones del sistema emocional (como expresiones faciales, conductas, etc.)
Autocontrol	La habilidad para controlar reacciones impulsivas frente a situaciones positivas y negativas, así como cumplir con obligaciones y metas

CONDUCTAS DE RIESGO

Conductas internalizantes*	Problemas de conducta relacionados a manifestaciones de comportamientos ansiosos, depresivos y problemas somáticos
Conductas externalizantes*	Problemas de conducta relacionados con hiperactividad, agresividad y conductas oposicionistas

* Se incluye solamente en educación media.

⁵ En el anexo 1.1.3. se describen los procesos de diseño y validación del componente habilidades socioemocionales.



Las oportunidades de aprendizaje

Evaluar las oportunidades de aprendizaje implica atender las condiciones institucionales, los recursos de enseñanza y las prácticas pedagógicas que los centros ofrecen sistemáticamente a sus alumnos, es decir, la oportunidad de aprender el currículo establecido. En términos específicos, las oportunidades de aprendizaje constituyen una aproximación a la medición del currículo implementado; por tanto, su análisis es un insumo central para el diseño curricular.

Con el fin de brindar insumos para el diseño e implementación de políticas para la enseñanza y la formación docente, se considera pertinente incluir su descripción y análisis para:

- a. conocer la distribución y calidad de las oportunidades de aprendizaje que se ofrecen a los estudiantes a nivel nacional;
- b. diseñar pruebas de evaluación válidas que tomen en cuenta lo que los docentes enseñan en las aulas, considerando los temas, secuenciación e intenciones o expectativas (Schmidt, 1996); y
- c. analizar los desempeños del sistema educativo de acuerdo a las distintas oportunidades de aprendizaje que los propios docentes dicen ofrecer a sus estudiantes.

Esta información surge de las respuestas que brindan los docentes de los grupos considerados en la evaluación.

El componente oportunidades de aprendizaje busca aportar información sobre las siguientes cuestiones:

- ¿existen variaciones a nivel del país respecto a la cobertura curricular?, ¿se observan relaciones entre los logros de los estudiantes en las pruebas de matemática y lectura y la cobertura curricular reportada por los docentes?;
- ¿qué recursos utilizan los docentes en sus prácticas y con qué criterios los seleccionan?;
- ¿cuáles son los énfasis puestos en los diversos contenidos curriculares?;
- ¿cómo organiza y secuencia el docente las actividades que propone en el aula?, ¿qué complejidad cognitiva tienen las actividades que propone?;
- ¿cuánto tiempo pedagógico se dedica a trabajar matemática y lectura en las aulas?;
- las prácticas pedagógicas de los docentes, ¿varían por tipo de centro o por características de los docentes?; y
- ¿se observan relaciones entre los logros de los estudiantes en las pruebas de matemática y lectura y las prácticas de enseñanza reportadas por los docentes?

El INEEd seleccionó y definió cinco dimensiones para la evaluación de las oportunidades de aprendizaje en el sistema educativo uruguayo.



TABLA 4.

DIMENSIONES PARA LA EVALUACIÓN DE LAS OPORTUNIDADES DE APRENDIZAJE⁶

CARACTERÍSTICAS Y EXPERIENCIA DE LOS DOCENTES

Cualidades de los docentes que inciden en su desempeño y capacidad profesional

COBERTURA Y SECUENCIA

Grado de implementación de los contenidos curriculares presentes en las actividades que el docente propone a los estudiantes, así como su secuencia didáctica (inter e intragrado) y énfasis

TEXTOS Y RECURSOS

Disponibilidad efectiva de recursos escolares por parte del docente para la preparación de sus clases, para el trabajo de los alumnos en el aula o en el domicilio. Incluye los criterios para su selección y frecuencia de uso

PRÁCTICAS PEDAGÓGICAS

Acciones que realiza el docente en un contexto específico para la enseñanza de los contenidos y para cumplir con los objetivos propuestos

NIVEL DE EXIGENCIA COGNITIVA DE LAS TAREAS

Profundidad conceptual con que el docente desarrolla los contenidos que presenta a los estudiantes, involucrando procesos cognitivos desde los más simples a los más complejos

⁶ En el anexo 1.1.4. se describen los procesos de diseño y validación del componente oportunidades de aprendizaje.



Los desempeños de los estudiantes en lectura y matemática

La Evaluación Nacional de Logros Educativos está orientada al currículo. Dicha orientación al currículo prescripto significa un cambio con respecto a las evaluaciones nacionales estandarizadas de sexto año de primaria que se realizaron en el país desde el año 1996 hasta 2013⁷.

En Aristas las especificaciones de qué evaluar —especialmente en lectura y matemática— se construyen a partir de documentos curriculares en los que se describe lo que se espera que los estudiantes sean capaces de hacer (CEIP, 2008; ANEP, 2014; CEIP, 2016a y 2016b; CES, 2006a, 2006b, 2006c, 2007, 2010a, 2010b, 2010c y 2016; CETP, 2007a, 2007b, 2007c, 2007d, 2007e y 2007f). No obstante, se habla de orientación y no de alineación porque, si bien la política curricular uruguaya ha avanzado en la definición de lo que se espera que los estudiantes sean capaces de hacer al culminar ciertos ciclos educativos, no se ha alcanzado aún un grado de operacionalización en logros observables y contrastables que permita la alineación de la Evaluación Nacional de Logros Educativos en términos estrictos⁸.

Aristas evalúa las competencias de los estudiantes en lectura y en matemática, seleccionadas de los perfiles de egreso de cada grado escolar. Se entiende que la competencia es la puesta en acción de un saber, un saber hacer y un saber actuar a partir de la integración de distintos tipos de conocimientos, así como las habilidades, las actitudes y los valores puestos en juego en distintas situaciones y contextos concretos⁹. Se asigna un lugar preponderante al contexto que orienta la toma de decisiones y a la actuación de los individuos en sus interacciones con el escenario en que se desenvuelven en la vida personal, social y escolar.

Las competencias son un referente relevante para la evaluación del nivel de logro de los estudiantes. Ser competente no es una cuestión de todo o nada, sino un proceso que se desarrolla en el tiempo. La evaluación de las competencias posibilita disponer de información sobre el nivel alcanzado por los estudiantes en un determinado momento (Coll, 2007: 36).

⁷ Las evaluaciones de aprendizajes entre 1996 y 2013 no estuvieron alineadas con el programa oficial. Para esas evaluaciones se realizó una selección de competencias y contenidos que, en aquel momento, se consideraron imprescindibles en la formación de todos los estudiantes del país, independientemente de su origen social y familiar. Esto implicaba que no todos los contenidos evaluados en la prueba hubieran sido efectivamente enseñados. Dicha selección fue puesta a consideración de los docentes en noviembre de 1995. Los resultados de la encuesta aplicada, así como los resultados obtenidos en el operativo de validación de instrumentos, nutrieron los debates en el Grupo de Consulta, órgano creado por el Consejo Directivo Central (CODICEN) de la ANEP a propuesta de la Unidad de Medición de Resultados Educativos (UMRE) en marzo de 1996. Por razones metodológicas, para mantener la comparabilidad de los resultados obtenidos en el tiempo, este referente se mantuvo para todas las evaluaciones nacionales realizadas en sexto de primaria por la UMRE, la Gerencia de Investigación y Evaluación Educativa y por la Dirección de Investigación, Evaluación y Estadística, órganos de la ANEP que implementaron las evaluaciones entre 1996 y 2013.

⁸ A partir de los resultados de Aristas, de los procesos que se implementen luego de la definición del Marco Curricular de Referencia Nacional (ANEP, 2017) y de la interacción entre los resultados de la evaluación y las definiciones curriculares, será posible realizar una alineación que permita dar cuenta de la proporción de alumnos que alcanzan los desempeños esperados por la ANEP.

⁹ El uso de la noción de competencia implica una continuidad con lo definido en las evaluaciones nacionales en sexto de primaria. La UMRE en 1996 concibió el término “competencia” íntimamente vinculado al procesamiento de los contenidos programáticos no solo del grado que está siendo evaluado, sino también de los que lo anteceden. Estos criterios respaldaron todas las evaluaciones nacionales estandarizadas realizadas en sexto de primaria, desde 1996 hasta 2013.



En función de ello se concibe a la **competencia lectora** como la capacidad de construir significados a partir de una variedad de textos escritos, movilizando conocimientos, habilidades, estrategias, emociones y actitudes, a partir de la interpretación de información literal e implícita, y del establecimiento de relaciones intertextuales, intratextuales e hipertextuales.

Por su parte, la **competencia matemática** se concibe como la capacidad de resolver planteos matemáticos enmarcados en distintas situaciones, poniendo en juego conocimientos, habilidades, emociones y actitudes, involucrando el saber sobre los contenidos y el saber actuar intencionalmente con ellos (qué hacer, cómo, cuándo y por qué hacerlo). Para dar cuenta de su competencia, los estudiantes deben ser capaces de indagar matemáticamente sobre diferentes realidades, desarrollar estrategias, discutir su pertinencia, determinar el rango de datos que se necesitan para aprehenderlas, establecer relaciones entre ellos, manejar conceptos matemáticos aprendidos, analizar regularidades y patrones, generalizar, explicar, conjeturar, comunicar, disponer de distintas representaciones de los objetos de la asignatura, argumentar y defender posiciones propias y analizar la viabilidad de las de otros.

Tomando en cuenta que llegar a ser competente implica un proceso de desarrollo y que en nuestro país se observa una de las mayores desigualdades en los desempeños comparativamente con otros países, se diseñaron pruebas que incluyen ítems que cubren la máxima variabilidad posible de habilidades a medir en cada competencia a evaluar. Ello es muy pertinente para brindar evidencias válidas acerca de:

- lo que todos son capaces de hacer, considerando ambos extremos de la distribución; y
- lo que los estudiantes son o no son capaces de hacer en aquellos temas considerados en la selección realizada en la prueba entre los bloques temáticos de los perfiles de egreso (currículo teórico), pero que no aparecen como aquellos enfatizados en la enseñanza (currículo implementado).

En los marcos de lectura y matemática de educación primaria y educación media se puede ahondar sobre estas definiciones (INEEd, 2017a, 2017b, 2017c y 2017d).



Tablas de dominio de lectura

Tercer año de primaria

COMPETENCIA LECTORA			
El estudiante es capaz de construir significados en diversidad de textos escritos, con propósitos definidos, mediante la identificación de datos explícitos, la interpretación de información implícita, y del establecimiento de relaciones intratextuales, intertextuales e hipertextuales, para lo cual se movilizan conocimientos, habilidades, estrategias, emociones y actitudes.			
INTENCIONES	DIMENSIONES		
NARRAR EXPLICAR PERSUADIR	LECTURA LITERAL El estudiante reconoce significados explícitos en textos.	LECTURA INFERENCIAL (local y global) El estudiante reconoce significados implícitos en diversidad de textos, a nivel global, de párrafo, de enunciado y de oración.	LECTURA CRÍTICA El estudiante establece relaciones de sentido entre el contenido de textos generando opiniones.
	Reconoce elementos básicos de la situación de enunciación.	Reconoce el tema del párrafo o del enunciado. Resume la idea general del texto.	Reconoce la intencionalidad narrativa, argumentativa o explicativa del texto.
	Localiza información explícita.	Establece relaciones entre dos fuentes de un mismo texto.	Construye significados a partir de palabras claves.

Sexto año de primaria

COMPETENCIA LECTORA			
El estudiante es capaz de construir significados en diversidad de textos escritos, con propósitos definidos, mediante la identificación de datos explícitos, la interpretación de información implícita, y del establecimiento de relaciones intratextuales, intertextuales e hipertextuales, para lo cual se movilizan conocimientos, habilidades, estrategias emociones y actitudes.			
INTENCIONES	DIMENSIONES		
NARRAR EXPLICAR PERSUADIR	LECTURA LITERAL El estudiante reconoce significados explícitos en los textos.	LECTURA INFERENCIAL (local y global) El estudiante reconoce significados implícitos en diversidad de textos, a nivel global, de párrafo, de enunciado y de oración.	LECTURA CRÍTICA El estudiante establece relaciones de sentido entre el contenido de textos y genera opiniones.
	Reconoce elementos básicos de la situación de enunciación.	Reconoce el tema del párrafo o del enunciado.	Reconoce la intencionalidad narrativa, argumentativa o explicativa del texto.
	Localiza información explícita.	Resume la idea general del texto y elabora conclusiones.	Elabora opiniones acerca de los hechos y situaciones que plantea el texto.
	Reconoce la progresión tema-remata.	Relaciona información de los enunciados y párrafos. Jerarquiza información, datos o sucesos en diferentes formatos.	



Tercer año de educación media

COMPETENCIA LECTORA			
El estudiante es capaz de construir significados en diversidad de textos escritos, con propósitos definidos, mediante la identificación de datos explícitos, la interpretación de información implícita y del establecimiento de relaciones intratextuales, intertextuales e hipertextuales, para lo cual se movilizan conocimientos del mundo o conocimiento enciclopédico, habilidades, estrategias, emociones y actitudes.			
INTENCIONES	DIMENSIONES		
NARRAR DESCRIBIR PERSUADIR EXPONER	LECTURA LITERAL El estudiante reconoce significados explícitos en los textos.	LECTURA INFERENCIAL (local y global) El estudiante reconoce significados implícitos en diversidad de textos, a nivel global, de párrafo, de enunciado y de oración ¹⁰ .	LECTURA CRÍTICA El estudiante establece relaciones de sentido generando opiniones ¹¹ .
	Reconoce elementos básicos de la situación de enunciación.	Reconoce el tema del párrafo o del enunciado.	Evalúa e interpreta los hechos, situaciones o conceptos que plantea el texto.
	Localiza información explícita.	Resume la idea general del texto y elabora conclusiones.	
		Reconoce la intencionalidad narrativa, descriptiva, argumentativa o expositiva del texto.	Reconoce elementos complejos de la situación de enunciación.
	Reconoce la progresión temática.	Relaciona información de los enunciados y párrafos.	
			Jerarquiza información, datos o sucesos en diferentes formatos.



Tablas de dominio de matemática

Tercer año de primaria

COMPETENCIA MATEMÁTICA				
El alumno resuelve planteos matemáticos enmarcados en distintas situaciones, poniendo en juego conocimientos, habilidades, emociones y actitudes, involucrando el saber sobre los contenidos y el saber actuar intencionalmente con ellos (qué hacer, cómo, cuándo y por qué hacerlo).				
Dimensiones de la competencia matemática		Dimensión información	Dimensión aplicación	Dimensión comprensión
		AFIRMACIONES		
		El alumno reconoce información matemática básica, convenciones y representaciones de los objetos matemáticos. Es capaz de recordar, recuperar e identificar dicha información. (A)	El alumno usa sus conocimientos para ejecutar y aplicar rutinas matemáticas necesarias y procedimientos —algoritmos de cálculo, fórmulas matemáticas o trazados—. (B)	El alumno resuelve situaciones matemáticas para las cuales debe establecer relaciones, validar o elaborar procedimientos y validar afirmaciones. (C)
BLOQUES TEMÁTICOS	PERFILES	SUBAFIRMACIONES		
NUMERACIÓN (1)	Interpretar números racionales, reconociendo la equivalencia de distintas representaciones, apelando al valor posicional, regularidades del sistema de numeración decimal y a la relación de orden.	Reconoce distintas representaciones de números racionales. (1.A)	Establece relaciones de orden e identifica regularidades entre números naturales. (1.B)	Valida explicaciones sobre representaciones, ordenamiento de números racionales y regularidades del sistema de numeración decimal. (1.C)
OPERACIONES (2)	Resolver situaciones de adición, sustracción, multiplicación y división entre números naturales, teniendo en cuenta la relación entre sus términos y las propiedades de las operaciones.	Reconoce las operaciones básicas. (2.A)	Realiza adiciones, sustracciones, multiplicaciones y divisiones entre números naturales. (2.B)	Resuelve situaciones de cálculo entre números naturales teniendo en cuenta propiedades y relaciones entre los términos de las operaciones. (2.C)
MAGNITUDES Y MEDIDAS (3)	Resolver situaciones en las que se requiera medir y estimar cantidades de magnitud, considerando distintas unidades de medida.	Distingue las magnitudes y reconoce instrumentos y unidades que permiten la medición. (3.A)	Estima y compara cantidades de magnitud. (3.B)	Justifica procedimientos de estimación y medición. (3.C)
ESTADÍSTICA (4)	Identificar, registrar y utilizar información presentada en tablas o gráficos.	Identifica información presentada en tablas o gráficos. (4.A)	Relaciona información estadística en tablas, gráficos o lenguaje natural. (4.B)	Extrae conclusiones sobre información estadística. (4.C)
GEOMETRÍA (5)	Identificar, describir y clasificar figuras en función de distintas propiedades y representaciones.	Identifica figuras del plano y del espacio. (5.A)	Describe figuras geométricas. (5.B)	Clasifica figuras del plano y del espacio según sus propiedades. (5.C)



Sexto año de primaria

COMPETENCIA MATEMÁTICA				
El alumno resuelve planteos matemáticos enmarcados en distintas situaciones, poniendo en juego conocimientos, habilidades, emociones y actitudes, involucrando el saber sobre los contenidos y el saber actuar intencionalmente con ellos (qué hacer, cómo, cuándo y por qué hacerlo).				
Dimensiones de la competencia matemática		Dimensión información	Dimensión aplicación	Dimensión comprensión
AFIRMACIONES				
		El alumno reconoce información matemática básica, convenciones y representaciones de los objetos matemáticos. Es capaz de recordar, recuperar e identificar dicha información. (A)	El alumno usa sus conocimientos para ejecutar y aplicar rutinas matemáticas necesarias y procedimientos — algoritmos de cálculo, fórmulas matemáticas o trazados—. (B)	El alumno resuelve situaciones matemáticas para las cuales debe establecer relaciones, validar o elaborar procedimientos y validar afirmaciones. (C)
BLOQUES TEMÁTICOS	PERFILES	SUBAFIRMACIONES		
NUMERACIÓN (1)	Interpretar números racionales, reconociendo la equivalencia de distintas representaciones, apelando a las propiedades y regularidades del sistema de numeración decimal, a la relación de orden y a relaciones numéricas vinculadas a la divisibilidad.	1.1 Representaciones, regularidades y orden		
		Reconoce distintas representaciones de números racionales. (1.1.A)	Establece relaciones de orden e identifica regularidades entre números racionales (1.1.B)	Valida explicaciones sobre representaciones, ordenamiento de números racionales y regularidades del sistema de numeración decimal. (1.1.C)
		1.2 Divisibilidad		
		Reconoce relaciones de divisibilidad entre números naturales. (1.2.A)	Obtiene múltiplos y divisores de números naturales. (1.2.B)	Resuelve situaciones que implican relaciones de divisibilidad entre números naturales. (1.2.C)
OPERACIONES (2)	Resolver situaciones de adición, sustracción, multiplicación y división entre números racionales, teniendo en cuenta la relación entre sus términos, entre operaciones inversas y las propiedades de las operaciones.	2.1 Operaciones		
		Reconoce las operaciones básicas entre números racionales, teniendo en cuenta la relación entre operaciones inversas. (2.1.A)	Realiza operaciones entre números racionales, teniendo en cuenta la relación entre sus términos. (2.1.B)	Resuelve situaciones de cálculo entre números racionales teniendo en cuenta propiedades y relaciones entre los términos de las operaciones. (2.1.C)
		2.2 Proporcionalidad		
		Identifica la relación de proporcionalidad directa. (2.2.A)	Resuelve situaciones en las que interviene la proporcionalidad directa. (2.2.B)	Relaciona distintas representaciones de la proporcionalidad directa. (2.2.C)
MAGNITUDES Y MEDIDAS (3)	Resolver situaciones en las que se requiera medir perímetros, áreas y volúmenes y estimar cantidades de magnitud.	Reconoce distintas unidades de medida. (3.A)	Aplica procedimientos de medición y de estimación de cantidades de magnitud. (3.B)	Distingue la independencia entre magnitudes en una misma figura y relaciona unidades de medida e instrumentos que habilitan la medición. (3.C)
		ESTADÍSTICA (4)	Interpretar, organizar y registrar información presentada en tablas y/o gráficos.	Reconoce información estadística. (4.A)
GEOMETRÍA (5)	Identificar, describir y clasificar figuras en función de distintas propiedades y representaciones.	Identifica figuras del plano y del espacio. (5.A)	Establece relaciones entre la descripción de figuras del plano o del espacio y sus distintas representaciones. (5.B)	Clasifica figuras del plano y del espacio según sus propiedades. (5.C)

Tercer año de educación media

COMPETENCIA MATEMÁTICA				
El alumno resuelve planteos matemáticos enmarcados en distintas situaciones, poniendo en juego conocimientos, habilidades, emociones y actitudes, involucrando el saber sobre los contenidos y el saber actuar intencionalmente con ellos (qué hacer, cómo, cuándo y por qué hacerlo).				
Dimensiones de la competencia matemática		Dimensión información	Dimensión aplicación	Dimensión comprensión
AFIRMACIONES				
		El alumno reconoce información matemática básica, convenciones y representaciones de los objetos matemáticos. Es capaz de recordar, recuperar e identificar dicha información. (A)	El alumno usa sus conocimientos para ejecutar y aplicar rutinas matemáticas necesarias y procedimientos — algoritmos de cálculo, fórmulas matemáticas o trazados—. (B)	El alumno resuelve situaciones matemáticas para las cuales debe establecer relaciones, validar o elaborar procedimientos y validar afirmaciones. (C)
BLOQUES TEMÁTICOS	PERFILES	SUBAFIRMACIONES		
MAGNITUDES Y MEDIDAS (3)	Resolver situaciones en las que se requiera estimar y calcular medidas.	Reconoce relaciones o propiedades para el cálculo de medidas. (3.A)	Aplica relaciones o propiedades para el cálculo de medidas. (3.B)	Resuelve situaciones que implican utilizar relaciones métricas entre elementos de una figura. (3.C)
ESTADÍSTICA Y PROBABILIDAD (4)	Organizar e interpretar información estadística y probabilística presentada en distintos formatos.	4.1 Estadística		
		Reconoce información estadística explícita presentada en distintos formatos. (4.1.A)	Procesa y organiza información estadística. (4.1.B)	Toma decisiones basándose en la interpretación de información estadística. (4.1.C)
		4.2 Probabilidad		
		Reconoce fenómenos aleatorios y diferentes tipos de sucesos. (4.2.A)	Asigna probabilidades a sucesos. (4.2.B)	Toma decisiones basándose en la interpretación de la probabilidad de un suceso y sus propiedades. (4.2.C)
GEOMETRÍA (5)	Utilizar los elementos y propiedades de las figuras geométricas y de las funciones del plano en el plano para resolver situaciones.	Reconoce figuras, sus elementos y distintas representaciones. (5.A)	Establece relaciones entre figuras usando propiedades de las figuras o de las transformaciones. (5.B)	Resuelve problemas geométricos basándose en propiedades de las figuras o de las transformaciones. (5.C)
ÁLGEBRA (6)	Reconocer, utilizar e interpretar modelos algebraicos para resolver situaciones.	Reconoce diferentes representaciones de funciones (6.A)	Realiza cálculos algebraicos y numéricos asociados y usa patrones. (6.B)	Modeliza e interpreta situaciones usando enfoque algebraico. (6.C)
ARITMÉTICA (7)	Interpretar los números racionales, reconociendo y usando distintas representaciones, la relación de orden y las operaciones entre ellos y sus propiedades.	Reconoce distintas representaciones de los números racionales y de las propiedades de las operaciones. (7.A)	Establece relaciones de orden y calcula, usando números racionales. (7.B)	Resuelve y modeliza situaciones que implican el uso de los números racionales y la relación de proporcionalidad. (7.C)



Diseño de pruebas

Para que la información producida por una evaluación resulte pertinente y confiable a los decisores de política, docentes, familias, alumnos y ciudadanía en general, debe ser válida: el proceso técnico previo a la generación de información y conocimiento debe haber estado orientado a garantizar la validez de las interpretaciones de los resultados. Una evaluación es válida en la medida en que responde a los propósitos para los cuales fue elaborada y sigue procedimientos y técnicas que permiten llegar a medidas que den cuenta de lo que efectivamente se quiere medir.

Para ello el INEED lleva a cabo una serie de trabajos de validación y diseño de instrumentos¹⁰. Dicho proceso se realiza en dos etapas:

- en una primera etapa se revisan antecedentes, se preparan versiones de instrumentos y se pretestean en muestras pequeñas; y
- en una segunda etapa se lleva a cabo una aplicación piloto de estos instrumentos en una muestra que combina procedimientos de selección aleatorios y por conveniencia.

Este proceso permite recoger distintas evidencias de validez para la interpretación de los resultados (AERA/APA/NMCE, 2014; ETS, 2005, 2009 y 2014). Una vez que la evidencia y soporte teórico son considerados suficientes y consistentes entre sí, los instrumentos están listos para su aplicación definitiva y posterior análisis de datos orientado a la devolución de resultados a partir de interpretaciones válidas de acuerdo a las definiciones asumidas en los marcos conceptuales de la evaluación (AERA/APA/NMCE, 2014).

En el caso de las pruebas de desempeño en lectura y matemática el diseño, desarrollo y validación comenzó por un relevamiento de los antecedentes nacionales, los perfiles de egreso y el marco curricular nacional que está siendo elaborado por la ANEP, trabajo que fue realizado por docentes uruguayos seleccionados a través de un llamado abierto (ver anexo 1.2.1).

A partir de allí, se construyeron los marcos conceptuales de lectura y matemática, las competencias a evaluar y la tabla de dominios, los cuales fueron validados por un comité de revisores externos conformado por docentes y técnicos de la ANEP y sus desconcentrados y docentes de la Universidad de la República (Udelar) y de universidades privadas¹¹.

Una vez validados los marcos conceptuales de lectura y matemática, se selecciona — mediante un llamado público y abierto— a un grupo de docentes en actividad, los que son capacitados para elaborar los ítems de la prueba siguiendo especificaciones técnicas precisas.

Posteriormente los ítems son sometidos a doble revisión por los especialistas del Instituto y validados a través de una evaluación piloto. A partir de las evidencias recolectadas en la

¹⁰ En el anexo 1.2 se describe con mayor detalle el proceso de elaboración de las evaluaciones de lectura y matemática.

¹¹ Se realizaron también diversas consultas y grupos de discusión con expertos, docentes y estudiantes sobre los cuestionarios de habilidades socioemocionales, ciudadanía, oportunidades de aprendizaje y contexto y entorno escolar, además de entrevistas cognitivas con estudiantes.

aplicación piloto se construye la prueba para la evaluación definitiva. Aunque la mayoría de los ítems son cerrados de múltiple opción, se incluyen algunos abiertos.

Test basados en evidencias

Una buena evaluación debe sustentarse en instrumentos con alto grado de validez, de modo que se pueda establecer lo que los estudiantes saben y saben hacer en relación con lo evaluado. La validez se define como el grado en que la evidencia empírica y la teoría dan sustento a las interpretaciones de los resultados de una medición (AERA/APA/NCME, 2014: 11). Por lo tanto, de lo que se trata es de realizar interpretaciones de resultados válidas de acuerdo a la teoría y a la experiencia¹².

Para aumentar la validez y confiabilidad del constructo que se quiere medir, Aristas utiliza la metodología de test basados en evidencias y pruebas matriciales (anexo 1.2.2).

Para la construcción de pruebas que midan el desempeño en lectura y matemática Aristas adoptó la metodología de test basados en evidencias, desarrollada por el Educational Testing Service (ETS), que permite vincular de manera clara las preguntas específicas que componen la prueba con el marco de la evaluación. De esta forma se favorece la validez para la interpretación de los resultados (Mislevy, Almond y Lukas, 2003 y Zieky, 2014).

Pruebas matriciales

Aristas trabaja con pruebas matriciales, que se caracterizan por ser pruebas con una cantidad importante de actividades en las que no todos los estudiantes responden los mismos ítems. Se aplica una cantidad suficiente de ítems para relevar la mayor cantidad posible de evidencias con relación a las especificaciones de la prueba, lo que permite medir con mayor precisión la escala de desempeños.

Los cuadernillos de prueba se conforman por bloques de ítems que se intercalan de tal manera que posibilita vincular cada cuadernillo con los demás. Este diseño permite comparar las habilidades de los estudiantes en forma independiente del cuadernillo que hayan contestado. Todo este proceso aumenta la validez y la confiabilidad del constructo que se quiere medir, en este caso, la competencia lectora y la competencia matemática. La utilización de un diseño matricial es lo que permite elaborar una prueba que cubre adecuadamente todo el rango de habilidad.

Técnica de calibración de ítems, estimación de puntajes y equiparación entre evaluaciones

Se trabaja a partir de la Teoría de Respuesta al Ítem (TRI). Esta metodología permite superar dos limitaciones de la Teoría Clásica de los Tests (TCT)¹³. Una de las propiedades de la TRI

¹² La validez no concierne únicamente al desarrollo de los instrumentos de evaluación, sino también al uso que se hace de los resultados (AERA/APA/NCME, 2014:13). Los resultados deben interpretarse para aquello que fueron diseñados para medir.

¹³ Las limitaciones superadas son: a) las características de los ítems dependen del grupo de personas en el que se han aplicado y b) la puntuación de una persona depende del conjunto particular de ítems administrados.



es su invarianza en dos sentidos: de los ítems respecto a posibles diferentes distribuciones de la habilidad medida y de la habilidad medida a partir de diferentes conjuntos de ítems. Si se cumplen estas dos condiciones los valores de dificultad y de discriminación de los ítems serán los mismos, tanto si se han aplicado a un grupo de personas con alto nivel de habilidad como a un grupo con niveles bajos, y el nivel de habilidad de una persona puede ser obtenido a partir de conjuntos de ítems distintos.

Las pruebas están diseñadas para garantizar la comparabilidad en el tiempo entre distintas cohortes de estudiantes del mismo grado. Los bloques de anclaje se seleccionan luego de la aplicación definitiva y se resguarda su confidencialidad.

La **calibración de los parámetros de los ítems** que se utilizaron para estimar los puntajes de los estudiantes se lleva a cabo en varias etapas:

1. Depuración de estudiantes
 - a. Total de ítems contestados
 - b. Tiempo de realización de la prueba
 - c. Ítems contestados al azar
 - d. Estudiantes identificados con necesidades educativas especiales
2. Depuración de ítems por aspectos de teoría clásica
3. Depuración de ítems por comportamiento diferencial

Luego de depurados los estudiantes y los ítems, la calibración de ítems se realiza con el modelo de Rasch o de crédito parcial, dependiendo de la naturaleza del ítem. La estimación de los parámetros de los ítems se realiza por máxima verosimilitud marginal.

Para la calibración, las respuestas que fueron omitidas¹⁴ se las consideró como incorrectas, mientras que los ítems no abordados¹⁵ no fueron considerados en la calibración.

La estimación de puntajes de los alumnos se realiza mediante el modelo de Rasch (Bond y Fox, 2013) de un parámetro para el caso de ítems dicotómicos y el modelo de crédito parcial (van der Linden y Hambleton, 2013) para ítems abiertos. Para este proceso se utiliza el paquete TAM (Robitzch, Kiefer y Wu, 2017) del software R (R Core Team, 2017).

Los puntajes individuales de los estudiantes se obtienen por medio del estimador de verosimilitud ponderada (WLE por sus siglas en inglés). Se utiliza como insumo las respuestas a los ítems de los estudiantes y se fijan los parámetros de los ítems obtenidos en la etapa de calibración.

Los puntajes obtenidos asumen que la población proviene de una distribución normal con media cero y desvío uno, por lo tanto, los puntajes se transformaron linealmente a una escala con media 300 y desvío 50.

¹⁴ Ítems que el estudiante ve, pero por algún motivo no responde ninguna opción.

¹⁵ Ítems que el estudiante no llegó a ver al final de la prueba.

La aplicación de Aristas en los centros educativos

Los distintos instrumentos que componen la Evaluación Nacional de Logros Educativos son administrados en forma controlada por un aplicador del INEE, capacitado específicamente para tal finalidad. Su función es garantizar condiciones homogéneas en el desarrollo de la aplicación y la confidencialidad de las pruebas. Se trata de una condición indispensable para asegurar la validez de la información producida y para garantizar la comparabilidad de los resultados tanto entre los distintos centros que están siendo evaluados en un mismo momento, como en el tiempo¹⁶.

Salvo el cuestionario para las familias, que en primaria se responde en papel, los restantes instrumentos se administran en computadora y en línea. Ello implica algunas restricciones que son tenidas en cuenta en el marco conceptual y las especificaciones de cada una de las pruebas a aplicar (INEE, 2017a, 2017b, 2017c y 2017d).

Para la aplicación de Aristas se utiliza el parque de computadoras y la red Ceibal, con los refuerzos necesarios para cada centro, los cuales son previamente recabados. Cada aplicación es acompañada de un informe de campo que busca dar cuenta de la cobertura y de distintas incidencias que deban ser consideradas a la hora de analizar la información.

¹⁶ La comparabilidad en el tiempo depende, además, de la definición de un conjunto de ítems de anclaje entre dos evaluaciones de la misma población realizada en distintos momentos del tiempo.

EL REPORTE DE RESULTADOS

La presentación de los resultados de la Evaluación Nacional de Logros Educativos procura brindar evidencias válidas para el diseño e implementación de políticas, alimentar el debate pedagógico en los centros y proveer información de calidad para nutrir los debates públicos sobre educación.

Informes

Para cada aplicación de Aristas se presenta un informe general de resultados —acompañado de un resumen ejecutivo— donde se describe el contexto familiar y el entorno escolar, la participación y convivencia en los centros educativos, las oportunidades de aprendizaje que se brindan a los niños y adolescentes, las habilidades socioemocionales de los estudiantes, y sus desempeños en lectura y matemática.

En una segunda instancia, luego de la presentación del informe general, se realizan informes específicos por componentes e informes dirigidos a diversos actores, particularmente a docentes, directores y decisores de política.

Niveles de desempeño en lectura y matemática

Respecto de los desempeños en lectura y matemática, más que el reporte de un promedio nacional, Aristas enfatiza una devolución de resultados centrada en descripciones detalladas de lo que todos¹⁷ los estudiantes son capaces de hacer en diversos niveles de la habilidad. El anexo 2 detalla los procedimientos seguidos para elaborar las descripciones e identificar los puntos de corte en las escaladas de cada área y grado.

Se estiman puntajes en lectura y matemática. No se estiman puntajes para cada una de las dimensiones ni áreas de contenido definidas en las tablas de dominios. Sin embargo, pueden presentarse análisis descriptivos que ilustren los diversos grados de dificultad que para la población evaluada tienen las actividades de prueba relativas a las distintas dimensiones y contenidos contemplados en la tabla de dominios.

¹⁷ Se destaca la importancia de medir con poco error la cola inferior de la distribución, en donde se encuentran los estudiantes que no logran los desempeños básicos. Así se provee información acerca de lo que sí son capaces de hacer, de modo que a partir de allí se puedan diseñar estrategias de mejora. Igualmente relevante resulta no perder de vista lo que son capaces de hacer los alumnos de mayores logros.

Los resultados de los estudiantes se representan en una escala con promedio 300 y desvío estándar 50. Si bien todas las áreas de conocimiento tienen la misma escala, no se trata de puntajes comparables entre sí, ya que no cuentan con ítems comunes que permitan realizar comparaciones. Tampoco son comparables los resultados de una misma área entre tercero y sexto. Sí es posible monitorear en el tiempo los desempeños para cada una de las áreas y grados¹⁸.

Confidencialidad y anonimato de la información

El Instituto está comprometido con las obligaciones relativas a la reserva y confidencialidad de la información, según lo establecen la Ley de Derecho de Acceso a la Información Pública (n° 18.381), la Ley de Protección de Datos Personales y Acción de “Habeas Data” (n° 18.331) y sus decretos reglamentarios 484/009 y 232/010. El cumplimiento de estas leyes obliga al Instituto a velar por la confidencialidad y protección de los datos personales que recolecta —mediante su desidentificación— y a utilizarlos para sus estudios y evaluaciones resguardando siempre el carácter anónimo de la información, por lo cual está obligado a no brindar ninguna información que permita identificar personas. En ningún caso se darán a conocer datos por alumno o docente y no se construirán rankings de centros educativos.

Asimismo, la Ley General de Educación (n° 18.437) establece que cada organismo de enseñanza debe brindar al INEEd los medios necesarios para obtener la información que requiera la evaluación en la que participen los centros que están bajo su dependencia. Acto seguido, define que “la política de difusión de esta información resguardará la identidad de los educandos, docentes e instituciones educativas, a fin de evitar cualquier forma de estigmatización y discriminación” (artículo 116). Una vez difundidos sus informes de evaluación, el INEEd liberará las bases de datos desidentificadas.

Evaluación en centros educativos

Aristas ofrece la posibilidad de que todos los centros y docentes del país apliquen libremente una versión de prueba. Se trata de un insumo para nutrir la discusión docente sobre los procesos de aprendizaje de los estudiantes a partir de la comparación con los resultados nacionales y de otros centros de similar composición sociocultural. Este material incluye recomendaciones para promover la reflexión pedagógica entre los educadores respecto de lo que se busca evaluar con cada uno de los ítems.

La aplicación de esta prueba es voluntaria. Cada centro tiene libertad de aplicarla en los grados que considere pertinentes. Por ejemplo, aplicar la prueba de tercero en segundo, tercero y cuarto podría ser un insumo para aproximarse a la progresión de desempeños en el centro.

¹⁸ La evaluación se realiza cada tres años.

Esta prueba no persigue la evaluación de estudiantes, docentes ni centros. Podrá ser utilizada como una autoevaluación en los centros educativos, implementada localmente por sus actores. Este instrumento procura que directores y maestros puedan establecer relaciones entre el desempeño de sus alumnos y el que se ha observado a nivel nacional y por contexto sociocultural, así como alimentar la reflexión sobre las prácticas de enseñanza y desarrollar la cultura evaluativa a nivel institucional y su interacción con la evaluación nacional orientada al currículo.





ANEXOS

ANEXO 1.

DISEÑO Y VALIDACIÓN DE LOS INSTRUMENTOS DE ARISTAS

1.1. El contexto familiar y el entorno escolar, las oportunidades de aprendizaje, la convivencia y la participación, y las habilidades socioemocionales

Validar un instrumento de medición antes de su implementación definitiva es fundamental para poder interpretar adecuadamente los resultados. La validación consiste en recoger distintos tipos de evidencias que permitan revisar las definiciones conceptuales, identificar el grado de coherencia entre estas y los datos, de forma que los resultados sean interpretables y se favorezca su uso (AERA/APA/NMCE, 2014).

La validación técnica de un instrumento se concibe como un proceso de acumulación de evidencias que permitan realizar inferencias sobre su capacidad para medir lo que se pretende medir, justificando así las puntuaciones que se obtienen de su aplicación y la interpretación que de ellas se deriva. La validez constituye, entonces, la propiedad más importante y fundamental al momento de desarrollar y evaluar un instrumento (Abad y otros, 2011). Para los diferentes componentes de Aristas se realizaron diversas actividades de manera de acumular evidencias sobre la validez de los instrumentos construidos.

Si bien las etapas de validación de los distintos instrumentos para relevar estos componentes tuvieron algunas diferencias entre sí, en general se privilegió la validez de contenido a través de jueceo y la realización de grupos focales con docentes. También se realizaron diversas tareas de manera de generar evidencia sobre la validez de constructo (específicamente: validez de estructura de los constructos) y la validez de los instrumentos con relación al proceso cognitivo de respuesta de los informantes (Smith y Molina, 2011).

1.1.1. El contexto familiar y el entorno escolar

En esta fase del desarrollo de la evaluación del componente se le otorgó principal preponderancia a la validez de contenido. Esta se relaciona con la definición teórica de los constructos, con los aspectos relevantes que los componen y la selección de ítems para medir cada uno de ellos. Si los ítems miden de forma adecuada los aspectos teóricamente relevantes y estos describen de forma exhaustiva los constructos, entonces es posible afirmar que el contenido del instrumento es válido.

Para inferir la validez de los cuestionarios se realizaron varias actividades:

- a. Sistematización de la información relevada en los antecedentes. Se elaboraron matrices con las dimensiones, subdimensiones e ítems por cada uno de los cuestionarios y



luego se corroboró que las dimensiones y subdimensiones de interés relevadas en los antecedentes estuviesen incorporadas en los instrumentos, como recomiendan Prieto y Delgado (1996).

- b. Grupo de discusión con especialistas. Se realizó una jornada de discusión con investigadores del INEEd en la que se analizaron los ítems incluidos en los cuestionarios y se discutió sobre las dimensiones incorporadas.
- c. Grupos de discusión con docentes de educación primaria y media. Se seleccionaron centros educativos de Montevideo y del interior del país, y se realizaron grupos de discusión con docentes para intercambiar opiniones sobre los cuestionarios en relación con los temas incluidos, los ítems elaborados y los términos utilizados, entre otros temas.
- d. Entrevistas cognitivas a estudiantes de educación primaria y de educación media. Se llevaron a cabo entrevistas cognitivas individuales con estudiantes de sexto grado de educación primaria y de tercer grado de educación media de Montevideo y del interior del país para identificar el grado de comprensión de las preguntas y la forma en que los estudiantes formulan sus respuestas.

1.1.2. La convivencia y la participación

En el proceso de desarrollo y validación de los cuestionarios de convivencia y participación para primaria se enfatizó en la validez de contenido y de constructo a partir de las siguientes actividades:

- a. Sistematización de la información relevada en los antecedentes de evaluación a gran escala de la temática: se elaboraron matrices con las dimensiones, subdimensiones e ítems por cada uno de los cuestionarios. Estas matrices sirvieron como insumo para la selección y definición de los constructos a incluir en el marco, y como base de ítems para la operacionalización inicial de los constructos en un cuestionario propio.
- b. Entrevistas con informantes calificados con la finalidad de discutir sobre las dimensiones incorporadas y los ítems incluidos en los cuestionarios. Se realizaron entrevistas con académicos y referentes del sistema educativo con particular experiencia en la temática.
- c. Grupo de discusión con especialistas. Con la finalidad de discutir sobre los ítems incluidos en los cuestionarios, se realizó una jornada de discusión con investigadores del INEEd.
- d. Grupos de discusión con docentes. Se realizaron grupos de discusión con docentes de educación primaria y media, en Montevideo y en el interior del país, con el objetivo de intercambiar sobre los cuestionarios en relación con: los temas incluidos, la claridad lingüística (nivel de comprensión y nivel de adecuación sintáctica y semántica para el contexto nacional) y la pertinencia contextual (el grado en que refleja adecuadamente la variedad de realidades de los centros de educación primaria y media).
- e. Entrevistas cognitivas a estudiantes. Se realizaron entrevistas para examinar los mecanismos cognitivos que los sujetos ejecutan para responder a los cuestionarios, con el objetivo de determinar si los sujetos están realizando (o pueden realizar) las operaciones necesarias para que los investigadores puedan inferir la presencia, ausencia o nivel del constructo de interés (Smith y Molina, 2011). Las entrevistas fueron realizadas con estudiantes de sexto grado de primaria y de tercer grado de educación media de centros públicos y privados de Montevideo y del interior del país.



Cada una de estas instancias contó con un proceso de sistematización de la información relevada a partir del que se realizaron los ajustes pertinentes en los cuestionarios, definiendo así sus versiones para las aplicaciones piloto.

1.1.3. Las habilidades socioemocionales

Dado que se trata de un área poco explorada en la evaluación educativa, tanto a nivel internacional como nacional, el proceso de diseño del componente comenzó con un relevamiento de antecedentes internacionales y la consulta con expertos nacionales que trabajan en la temática, de manera de construir un marco de referencia pertinente y adecuado a la realidad educativa nacional. Durante el proceso de construcción del marco se mantuvo una consulta constante con alrededor de 20 expertos nacionales en el tema, con diversas adscripciones institucionales (Udelar, Universidad de Montevideo, Universidad Católica del Uruguay —UCU—, Universidad ORT Uruguay, Plan CAIF, Consejo de Educación Secundaria —CES— y ANEP).

Una vez que se contó con una definición teórica, operacionalizada en dimensiones y subdimensiones, se procedió a la operacionalización en ítems, a partir de instrumentos similares y validados, ya sea en Uruguay o a nivel internacional. Luego de seleccionados los ítems, cubriendo la totalidad del constructo definido, se inició un proceso de jueceo a través de un panel de nueve expertos que hicieron un juicio sobre la operacionalización realizada a partir de los criterios de claridad, relevancia y suficiencia de los ítems (Escobar-Pérez y Cuervo-Martínez, 2008), lo que permitió aumentar la validez de contenido de la primera versión del instrumento.

Una segunda etapa de la validación de contenido fue la realización de diversos grupos focales con maestros de educación primaria y profesores de educación media, que permitió valorar la adecuación contextual del instrumento. Por último, se realizaron entrevistas cognitivas con estudiantes de sexto de primaria y tercero de educación media, de manera de validar el instrumento con relación al proceso cognitivo de respuesta de los informantes (Smith y Molina, 2011).

En cuanto a la validez de constructo, específicamente en relación con la validez de la estructura latente de los constructos, se realizaron análisis factoriales a partir de los datos obtenidos en los estudios pilotos aplicados en sexto grado de primaria en 2016 y en tercer grado de educación media en 2017.

1.1.4. Las oportunidades de aprendizaje

En el año 2015 el INEE comenzó el proceso de construcción de los instrumentos para relevar las oportunidades de aprendizaje. Este proceso tuvo dos grandes etapas: a) una primera aproximación teórico-empírica y b) el diseño y validación técnica de los instrumentos.

La primera aproximación teórico-empírica se comenzó a desarrollar durante el año 2015, cuando se llevó a cabo un estudio exploratorio de tipo cualitativo en varias escuelas de educación primaria. Se realizaron observaciones de clase, análisis de los cuadernos de



los estudiantes y aplicación de un cuestionario exploratorio a los docentes de los grupos observados. Mediante los cuestionarios a docentes se relevaron algunas dimensiones que la revisión bibliográfica inicial sugería como pertinentes y se consultó específicamente sobre la adecuación de los tipos de actividades en el caso de ser incluidas en una prueba nacional. Además, se les solicitó que diseñaran y desarrollaran una actividad representativa de lo que suelen hacer para enseñar a resolver problemas de matemática o de lectura pensando en un tiempo de duración de aproximadamente 45 minutos. La observación de aula tuvo la finalidad de dar cuenta de las interacciones entre docentes y estudiantes a nivel general y frente a la realización de una tarea concreta. A partir de la revisión de los cuadernos de los estudiantes se buscó obtener información respecto a la cobertura curricular, la demanda cognoscitiva de las actividades y la retroalimentación del docente. El análisis de la información relevada por estas tres técnicas aportó insumos, entre otras cosas, para fundamentar la decisión de optar por consultar a los docentes.

En función del proceso realizado, el INEEd propone un enfoque próximo al abordaje del Trends in International Mathematics and Science Study (TIMSS) de la Asociación Internacional para la Evaluación del Rendimiento Educativo (IEA, por su sigla en inglés), mediante la consulta directa a docentes respecto de las características de una clase típica: el uso de textos y otros recursos, el tiempo destinado a la enseñanza de cada contenido, la secuenciación entre grados escolares, la profundidad cognitiva de las tareas que se les proponen a los alumnos y la planificación de la clase, entre otras¹⁹.

En la fase de diseño y validación técnica se otorgó principal preponderancia a la validez de contenido. La validez de contenido se relaciona con la definición teórica de los constructos, con los aspectos relevantes que los componen y la selección de ítems para medir cada uno de ellos. Si los ítems miden de forma adecuada los aspectos teóricamente relevantes y estos describen de forma exhaustiva los constructos, entonces es plausible afirmar que el contenido del instrumento es válido.

Se recogieron evidencias que permiten inferir la validez de contenido de los instrumentos a través de:

1. Juicio de expertos. Se trabajó con un grupo de expertos sobre la definición del constructo y las subdimensiones que se incluyeron. Este grupo estuvo conformado por los especialistas en lengua y matemática del INEEd, especialistas nacionales en evaluación y un especialista extranjero en evaluación de oportunidades de aprendizajes.
2. Construcción de grillas por ítem. Se realizó una matriz que contiene las dimensiones, subdimensiones, los ítems asociados a cada subdimensión y las distintas opciones de respuesta. Se siguieron las recomendaciones de Prieto y Delgado (1996) para la construcción de los ítems. Se elaboraron grillas para cada uno de los grados, niveles y áreas de interés. Esto permitió corroborar que todas las subdimensiones estuviesen representadas en los instrumentos.

¹⁹ Se optó por la consulta a docentes por varios motivos: en primer lugar, son quienes toman las decisiones en relación con la selección, secuenciación y adecuación de contenidos, prácticas pedagógicas y complejidad de las tareas, por lo tanto, son la fuente de información más apropiada para el objetivo. En segundo lugar, la decisión de consultar a los docentes como fuente de información coincide con la utilizada por la mayoría de las evaluaciones a gran escala, considerando los desafíos prácticos que demanda una evaluación en el nivel de los estudiantes (Schmidt, Cogan, y Solorio, 2017).

3. Aplicación del cuestionario a grupos de la población objetivo. En esta etapa se aplicaron tres técnicas:
- Grupos de discusión con docentes de educación media, tanto de Matemática como de Idioma Español y Literatura, donde se discutían las versiones de los cuestionarios y se buscaban sugerencias sobre la claridad de las consignas.
 - Entrevistas cognitivas a docentes seleccionados intencionalmente para representar a los distintos subsistemas, planes, modalidades y contextos educativos. Las entrevistas permitieron ajustar los instrumentos tomando en consideración los mecanismos cognitivos involucrados en el proceso de respuesta de los docentes.
 - Aplicaciones piloto. En el año 2016 se llevó adelante la aplicación piloto en educación primaria y en 2017 la aplicación piloto en educación media.

1.2. Las pruebas de lectura y matemática

1.2.1. Diseño y validación de las pruebas de lectura y matemática

La construcción de las pruebas se basó en la metodología de diseño de pruebas basado en evidencias (Mislevy, Almond y Lukas, 2003) del ETS. Entre los meses de febrero y mayo de 2016 los especialistas en lengua y matemática del INEEEd, los revisores externos de ítems y los redactores de ítems se capacitaron con personal del ETS en la metodología de diseño de pruebas basadas en evidencias.

La primera etapa de validación de este componente consistió en la definición del marco de referencia y la elaboración de la tabla de dominios a cubrir con la prueba. El marco de referencia y la tabla de dominios se sometieron a un jueceo de expertos en matemática y lengua, con la participación de especialistas con una adscripción institucional plural (Udelar, UCU, ANEP, Consejo de Educación Inicial y Primaria, CES, Consejo de Educación Técnico Profesional, Consejo de Formación en Educación, Proyecto de Apoyo a la Mejora de la Calidad de la Educación Inicial y Primaria en Uruguay, ProRazona, ProLee), para cada uno de los grados a evaluar (tercero y sexto de primaria y tercero de media).

Una vez definidas la tabla de dominios y las evidencias sobre las cuales realizar las tareas se comenzó con el proceso de producción de los ítems. Este proceso se hizo mediante la elaboración de pedidos de ítems específicos — que respondieran a cada una de las evidencias de la tabla— a los redactores de ítems (docentes de aula uruguayos). Para asegurar, de manera cualitativa, que los ítems diseñados medían las evidencias pretendidas se utilizó el juicio de expertos disciplinares para la revisión de cada uno de los ítems construidos por los redactores. Este proceso, conocido como *validez aparente* (Nevo, 1985), permitió aumentar la validez de cada uno de los reactivos diseñados. Para la realización de este procedimiento se siguieron lineamientos coincidentes con la metodología de desarrollo de pruebas basada en evidencias, la que provee una serie de criterios técnicos a los revisores de manera de evaluar cada uno de los ítems o tareas construidas. En la siguiente tabla se encuentran los criterios técnicos utilizados por los revisores para juzgar la validez de los ítems diseñados por los redactores.



TABLA A.1.1.

LINEAMIENTOS TÉCNICOS DE VALIDEZ DE LOS ÍTEMS

ELEMENTO POR ANALIZAR	OBJETIVO
CARACTERÍSTICAS GLOBALES DEL ÍTEM	
Correspondencia cabal con la evidencia de la cual deviene	Identificar la precisión conceptual de la correspondencia con la evidencia que antecede al ítem
Redacción accesible a la población objetivo	Conocer si el ítem será comprensible para la población a la cual se dirige
El contenido es relevante para la población objetivo	Identificar la coincidencia del contenido del ítem respecto a la información que recibe en su enseñanza
El contenido es socialmente adecuado para la población objetivo	Conocer si el ítem contiene símbolos o términos que puedan resultar ofensivos para algún sector de la población
La dificultad es acorde con la población objetivo	Identificar si el ítem corresponde al nivel de dificultad que la población objetivo recibe en su enseñanza
El contenido es equitativo para la diversidad sociocultural y geográfica de la población objetivo	Identificar si el ítem presenta algún elemento que pueda significar una ventaja para algún sector de la población objetivo
CARACTERÍSTICAS DE LA BASE DEL REACTIVO	
Está planteado en términos concretos	Observar si el ítem evita información innecesaria para responderlo correctamente. Se pretende analizar la concreción de los ítems con fines de evitar desgaste cognitivo de la población objetivo
Se redacta en términos positivos	Identificar si se evita el uso de términos negativos en su planteamiento que den lugar a confusiones lógicas que influyan en la consistencia de medida
La tarea que el estudiante debe realizar es explícita	Conocer si la ejecución solicitada no deja lugar a dudas para la población objetivo
PARA REACTIVOS DE OPCION MÚLTIPLE. CARACTERÍSTICAS DE LAS OPCIONES DE RESPUESTA	
La respuesta correcta es unívoca	Identificar si el planteamiento de las opciones de respuesta favorece a la disminución del error de medida
Las opciones de respuesta son excluyentes entre sí	Identificar si las opciones representan una relativa independencia entre ellas. Esto favorecerá la validez y la consistencia de medida
Los distractores son plausibles	Conocer si los distractores resultan viables de acuerdo con el nivel de formación y contexto de la población objetivo. El uso de las justificaciones en cada opción de respuesta deberá sustentar la plausibilidad
Las opciones están libres de pistas	Observar si existen pistas para responder correctamente el ítem en cuestiones de extensión, información innecesaria, entre otros aspectos
PARA REACTIVOS DE RESPUESTA CONSTRUIDA	
La instrucción/pregunta/enunciado es clara, precisa y no ambigua	Observar si el ítem es concreto con fines de evitar respuestas que no reflejen el conocimiento que se pretende medir
La pauta de codificación considera las generalidades de respuestas posibles	Identificar si la pauta de codificación considera, en la forma más abarcativa posible, el tipo de respuestas posibles de acuerdo con la población objetivo

Fuente: elaboración propia a partir de Haladyna, Downing y Rodríguez (2002).



Cada ítem fue sometido a doble revisión ciega por parte de un especialista del INEE y otro externo a la institución. Una vez terminado el proceso de elaboración de los ítems, se procedió al armado de bloques y cuadernillos mediante la metodología de bloques incompletos balanceados, de manera de realizar la aplicación piloto, tanto para educación primaria como para educación media, en una población lo suficientemente heterogénea.

Luego de realizada la aplicación piloto se seleccionaron los ítems para integrar las versiones definitivas de las pruebas. Se tuvo en cuenta el grado en que se cubría la tabla de dominios con ítems de diversas dificultades, así como aspectos específicos como su confiabilidad, funcionamiento diferencial y discriminación.

1.2.2. Desarrollo de test basados en evidencias

El principal aporte del desarrollo de test de acuerdo al modelo basado en evidencias, Evidence-Centered Design (ECD), radica en la generación de un marco coherente para el diseño y el desarrollo de las evaluaciones educativas. Dicho marco explica las relaciones existentes entre:

- las inferencias que el evaluador quiere hacer sobre el estudiante,
- lo que debe observarse para proporcionar las evidencias suficientes para hacer dichas inferencias y
- las características que deben tener las pruebas.

El diseño de especificaciones de pruebas a partir del modelo basado en evidencias es un marco para el desarrollo de pruebas que busca asegurar la validez mediante la alineación de los procesos y los productos de las pruebas con sus objetivos. Es una familia de prácticas del desarrollo de instrumentos diseñada para aclarar lo que mide una prueba y apoyar las inferencias hechas a partir de las evidencias.

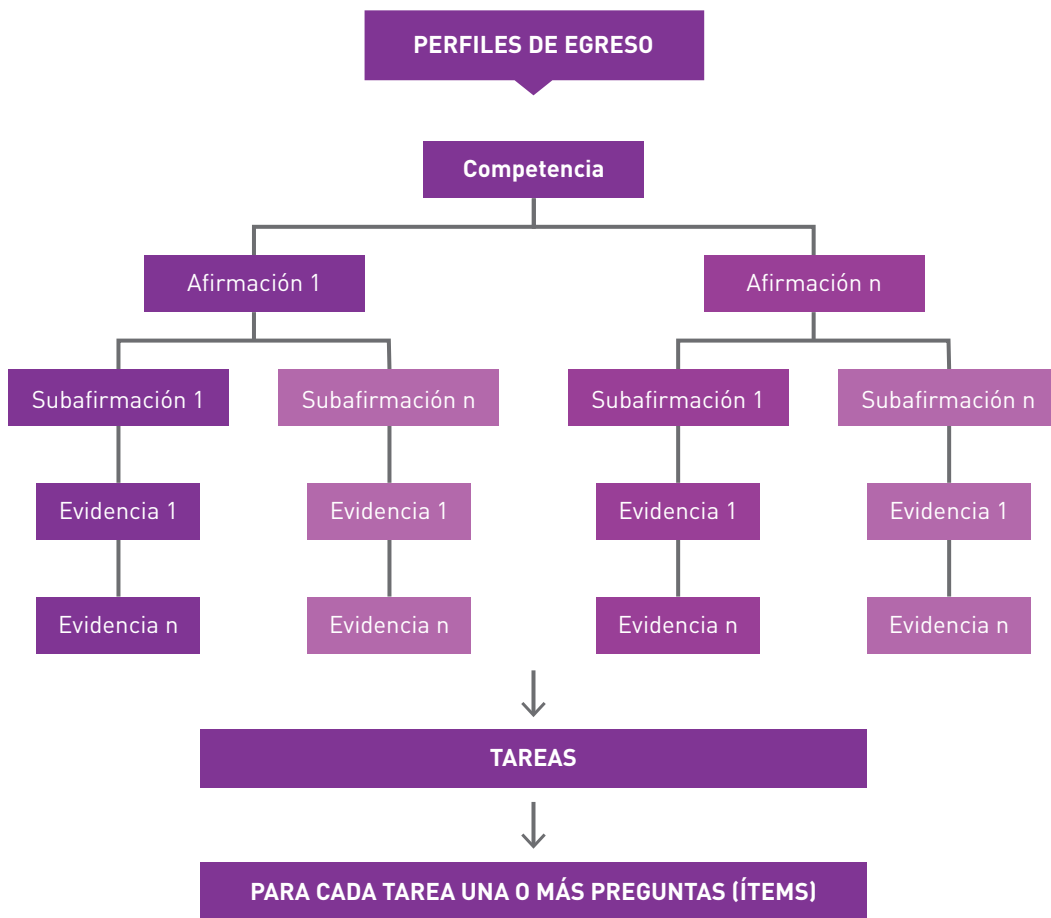
Esta metodología organiza el trabajo de diseño e implementación permitiendo vincular las definiciones teóricas con las respuestas de los estudiantes: meta o *perfil de egreso*, competencia, afirmaciones, evidencias, tareas, preguntas.

En Aristas se parte de un recorte (como en toda evaluación) de los documentos curriculares y de una definición de competencia formulada como una afirmación macro²⁰. De ella se derivan afirmaciones y de estas se desarrollan subafirmaciones y evidencias esperadas a partir de las cuales se diseñan tareas para conformar las preguntas que darán pie para observar el desempeño de los estudiantes. Este procedimiento facilita la coherencia entre las preguntas específicas de la prueba y su marco conceptual.

²⁰ Este proceso fue cometido del INEE junto a comités de referentes externos para cada área y grado.

FIGURA A.1.1.

ORGANIZACIÓN DEL DISEÑO DE ÍTEMS BASADOS EN EVIDENCIAS 1



Fuente: adaptación de Isabel Fernández, “Algunos aspectos relevantes de las evaluaciones en educación y principales resultados”, ICFES, presentación realizada el 15 de agosto de 2012 en Medellín.

TABLA A.1.2.

ORGANIZACIÓN DEL DISEÑO DE ÍTEMS BASADOS EN EVIDENCIAS 2

NIVELES	QUÉ ES	PROPÓSITO
PERFIL DE EGRESO	Expectativas de logro por nivel o ciclo en cada área de conocimiento	Promover una educación integral e inclusiva. Mejorar los aprendizajes de los estudiantes
COMPETENCIA	Puesta en acción de un saber, un saber hacer y un saber actuar a partir de la integración de distintos tipos de conocimientos, informaciones, habilidades, actitudes y valores puestos en juego en distintas situaciones y contextos concretos	Promover un enfoque de la enseñanza en la que los estudiantes comprendan los conocimientos y los utilicen efectivamente dentro y fuera de la escuela, en el marco de diferentes situaciones y contextos
AFIRMACIONES y SUBAFIRMACIONES	Enunciado sobre los conocimientos, habilidades y capacidades de los estudiantes que serán evaluados. Identifica el propósito de la prueba atendiendo a la pregunta: ¿qué se quiere decir sobre los estudiantes a partir de sus respuestas en una evaluación?	Comunicar el significado de los resultados de una evaluación
EVIDENCIAS	Enunciado que representa conductas o productos observables, con el que es posible verificar los desempeños a los que se refieren las afirmaciones. Responde a la pregunta: ¿qué tiene que hacer el evaluado que permita hacer inferencias sobre su desempeño?	Asegurar la pertinencia de una prueba, según sus características o restricciones. Para lograrlo, cada afirmación debe tener el número de evidencias suficiente para sustentarla
TAREAS	Enunciado que representa una descripción de una potencial familia de preguntas. Es específico y puede diferenciarse en su nivel de complejidad. Se refiere a lo que se pide a los evaluados que hagan en una prueba	Caracterizar la dificultad / complejidad de las preguntas. Dar las pautas para la construcción de preguntas
PREGUNTAS	Son los ítems a resolver por los alumnos. Dan cuenta de su desempeño	Observar el desempeño del alumno

Fuente: adaptación de Isabel Fernández, "Algunos aspectos relevantes de las evaluaciones en educación y principales resultados", ICFES, presentación realizada el 15 de agosto de 2012 en Medellín.

1.2.3. El banco de ítems

Los marcos conceptuales y las especificaciones técnicas de las pruebas de lectura y matemática fueron elaborados a partir de documentos curriculares. Se definieron dimensiones de cada competencia vinculadas a la complejidad cognitiva de las tareas y se tomaron en cuenta los contenidos curriculares. Para cada caso se establecieron afirmaciones, subafirmaciones y evidencias. De esta manera se llega a la tabla de dominios que operacionaliza cada competencia.

Las preguntas o ítems de las pruebas fueron elaborados por docentes capacitados en la metodología de test basados en evidencias. Para ello se elaboró una plantilla de diseño de ítem en donde se especificaban los requerimientos:



- área,
- grado,
- afirmación,
- subafirmación,
- evidencia y
- nivel de dificultad a criterio de experto.

Cada ítem pasa por un proceso de doble revisión ciega²¹, luego del cual se busca el acuerdo entre jueces respecto a las decisiones a tomar en relación a cada uno:

- aceptado sin modificaciones,
- aceptado con modificaciones:
 - menores, realizadas por los propios revisores,
 - mayores, realizadas por el autor del ítem²², y
- rechazado.

Todos los ítems, independientemente de su juicio, pasan a formar parte del banco de ítems (plataforma web). Esto permite dejar sentadas las valoraciones de los jueces y contar con la historia del ítem en términos de las modificaciones que se puedan haber realizado o no en cada etapa. Una vez aceptados, se especifica su estado: para revisión de edición y, posteriormente, para calibrar.

1.2.4. La validación empírica

Una vez diseñados los ítems se realiza una instancia de evaluación piloto cuyo objetivo es calibrar el metro de medida.

La muestra de las evaluaciones piloto es aleatoria, pero no representativa de la población nacional, sino que intencionalmente se busca contemplar la mayor heterogeneidad de situaciones posibles. Se busca obtener un número suficiente de respuestas para cada ítem de la prueba atendiendo la mayor heterogeneidad en términos educativos, socioeconómicos y lingüísticos con la menor cantidad de recursos.

La evaluación piloto no permite dar resultados porque el instrumento de medición aún no mide bien aquello que se busca relevar. En esta etapa se realiza una calibración psicométrica de los ítems²³ y, de acuerdo a los estadísticos obtenidos y la distribución de los ítems en la tabla de dominios, expertos en las asignaturas toman decisiones para el diseño de la prueba a aplicar en la evaluación definitiva. Los ítems seleccionados deben permitir medir adecuadamente toda la tabla de dominios.

²¹ Las duplas de revisores fueron integradas por un miembro del INEE y un revisor externo. Se siguen las recomendaciones para el diseño de evaluaciones planteadas por el ETS (2009).

²² En este caso el ítem reformulado volvió a pasar por el proceso de doble revisión ciega.

²³ La calibración es realizada por un equipo interno del INEE y por otro externo: el Departamento de Matemática y Aplicación del Centro Universitario Regional Este de la Udelar (CURE).



Una vez calibrados los ítems, sus parámetros pasan a formar parte del banco de ítems. Allí se registra en qué evaluación o evaluaciones fue utilizado, si se trata de un ítem de anclaje, liberado o ya calibrado.

1.2.5. El proceso en números (2016 y 2017)

CUADRO A.1.1.

CANTIDAD DE ÍTEMS Y DISEÑO DE CUADERNILLOS

		3° primaria		6° primaria		3° media	
		Lectura	Matemática	Lectura	Matemática	Lectura	Matemática
Ítems	Diseñados	340	368	421	492	383	437
	Calibrados	220	325	312	448	374	419
	Evaluación definitiva	108	195	144	224		
Diseño evaluación piloto	Ítems por bloque	9	12	12	15	14	15
	Bloques por cuadernillo	2	2	2	2	2	2
	Cuadernillos	22	25	24	28	29	28
Diseño evaluación definitiva	Ítems por bloque	9	13	12	16	14	15
	Bloques por cuadernillo	2	2	2	2	2	2
	Cuadernillos	12	15	12	14	18*	16*

* No son valores definitivos a la fecha. Pueden ser menos pero no más.

La tabla de dominios en la evaluación definitiva de primaria fue cubierta con la cantidad de ítems abiertos y cerrados presentados a continuación²⁴.

CUADRO A.1.2.

CANTIDAD DE ÍTEMS APLICADOS EN LA EVALUACIÓN DEFINITIVA DE PRIMARIA 2017 SEGÚN FORMATO DEL ÍTEM

		Cerrados	Abiertos	Total
Matemática	3°	180	15	195
	6°	210	14	224
Lectura	3°	96	12	108
	6°	132	12	144

En el cuadro A.1.3. se presentan los distintos tamaños muestrales de las aplicaciones de campo vinculadas a Aristas hasta 2017.

²⁴ No se presenta la distribución de los ítems para los dominios evaluados en media porque aún no se ha culminado el análisis del piloto y, por tanto, no se han tomado las decisiones finales sobre el diseño de la prueba a aplicar en octubre de 2018.

CUADRO A.1.3.

TAMAÑOS DE LAS MUESTRAS DE ARISTAS IMPLEMENTADAS HASTA FINES DE 2017

	Centros	3° primaria		6° primaria		3° media	
		Grupos	Estudiantes	Grupos	Estudiantes	Grupos	Estudiantes
Piloto de primaria	119	199	4.475	201	4.705		
Definitiva primaria	251	389	8.485	386	8.474		
Piloto de media	118					202	4.910

1.3. Características de la aplicación (2016 y 2017)

En octubre de 2016 se realizó la aplicación piloto en educación primaria y durante el año 2017 se llevó a cabo la aplicación piloto en educación media. Durante dicho período se llevaron a cabo análisis empíricos vinculando los datos con los constructos teóricos. A partir de ello se tomaron las decisiones correspondientes en vistas a contar con instrumentos válidos para la evaluación definitiva.

CUADRO A.1.4.

COMPONENTE DE LA EVALUACIÓN SEGÚN FUENTE DE INFORMACIÓN

	Alumnos	Familias	Docentes	Directores	Adscriptos
Contexto familiar	x	x		x	
Entorno escolar	x	x	x	x	x
Ciudadanía y participación	x	x	x	x	x
Habilidades socioemocionales	x				
Oportunidades de aprendizaje			x		
Lectura	x				
Matemática	x				

Los alumnos responden dos cuestionarios, uno sobre habilidades socioemocionales y otro sobre el resto de los componentes. Los docentes de educación primaria responden tres cuestionarios: dos de oportunidades de aprendizaje (uno sobre lectura y otro sobre matemática) y uno sobre el resto de los componentes. Los docentes de educación media responden dos cuestionarios: uno de oportunidades de aprendizaje (según la asignatura que dicta) y uno sobre el resto de los componentes. Los demás actores responden un único cuestionario.

Salvo la encuesta a las familias²⁵, el resto de los instrumentos son aplicados en una plataforma en línea.

²⁵ En educación primaria esta encuesta fue respondida en papel por los padres o responsables de los estudiantes participantes en la evaluación. En educación media resta analizar los datos del piloto para decidir si se realizará una encuesta a familias o se consultará a estudiantes (el trade-off entre maximizar la tasa de respuesta así como la calidad de las respuestas será el insumo básico para tomar la decisión). Se aplicó el cuestionario de familia y, a su vez, se incluyeron preguntas sobre contexto en el cuestionario a los estudiantes.

ANEXO 2.

GESTIÓN DEL PROCEDIMIENTO Y CAPACITACIÓN A LOS JUECES PARA LA DETERMINACIÓN DE PUNTOS DE CORTE

En el marco de un convenio de cooperación y asistencia técnica establecido entre el INEE y el Departamento de Matemática y Aplicaciones del Centro Universitario de la Región Este, dependiente de la Facultad de Psicología de la Udelar, el procedimiento de establecimiento de los puntos de corte fue coordinado por un equipo de especialistas en psicometría de ese departamento. Este equipo trabajó de manera coordinada con la dirección del Área Técnica del INEE, la dirección de la Unidad de Evaluación de Aprendizajes y Programas y el equipo de psicometría del INEE.

De acuerdo a las recomendaciones establecidas en los estándares para la evaluación educativa y psicológica (AERA/APA/NMCE, 2014: 107-109), se llevaron adelante jornadas de capacitación en las que se abordaron diferentes temáticas conceptuales:

- Introducción a los propósitos del establecimiento de estándares y de puntos de corte en pruebas de logro cognitivo.
- Clasificaciones de distintos métodos para el establecimiento de puntos de corte:
 - Métodos basados en sujetos
 - Métodos centrados en los tests
 - Métodos mixtos
 - Métodos holísticos y analíticos
- Aspectos a ser tenidos en cuenta para la elección del método de jueceo:
 - Propósitos del test
 - Nivel de complejidad del rasgo que se quiere evaluar
 - Complejidad del constructo y los productos a ser juzgados (formato del test)
 - Disponibilidad de recursos
- Centralidad de la definición conceptual de los niveles de desempeño y ejemplos de estos.
- Introducción a diversos métodos de jueceo:
 - Nedelsky
 - Ebel
 - Angoff: variaciones y Angoff extendido
 - Método de consenso directo
 - Método de grupos de contraste
 - Método de grupo límite
 - Método de captura política del juicio
 - Método de juicio analítico
 - Método de cuerpo de trabajo
 - Método Bookmark
 - Métodos de Hofstee y de Beuk
 - Método eCaT-Listening



Siguiendo algunas recomendaciones sobre la capacitación de jueces (Cizek y Bunch, 2007), durante la segunda sesión se realizaron ejercicios con diversos métodos de manera que los jueces se familiarizaran con los métodos, definiciones preliminares de los niveles de desempeño y con algunos ítems o actividades de las pruebas aplicadas.

Tanto para el área de lectura como de matemática, se realizaron ejercicios con el método de Bookmark y con el método de eCat-Listening (García y otros, 2013). En el área de matemática también se realizaron ejercicios utilizando el método Angoff extendido.

2.1. Cantidad y adscripción institucional de los jueces convocados para el establecimiento de puntos de corte

En el siguiente cuadro se especifica el número de jueces convocados para cada área y su adscripción institucional.

CUADRO A.2.1.

JUECES CONVOCADOS Y ADSCRIPCIÓN INSTITUCIONAL

	Lectura	Matemática
Internos INEEd	3	2
Externos (contrato INEEd)	2	1
Externos (DIEE-ANEP)	3	3
Total de jueces convocados	8	6

En total, se convocaron seis especialistas en el área de matemática y ocho en el área de lectura, con los que se comenzó a trabajar en diciembre de 2017 en una capacitación a cargo del CURE.

2.2. Procedimiento para el desarrollo de los puntos de corte: jueceo por especialistas

Una de las etapas críticas en el desarrollo de pruebas o test educativos es el establecimiento de puntos de corte en la escala de desempeño que permitan determinar cuándo se considera que un estudiante logra situarse por encima o por debajo de determinado nivel de conocimientos, competencias o habilidades (Dochy y otros, 2009). En la medida que Aristas se encuentra *orientada* y no *alineada* con el currículo, es importante explicitar que los puntos de corte establecidos no expresan las definiciones curriculares²⁶.

²⁶ Esto será posible una vez que ambos instrumentos estén alineados a partir de definiciones curriculares que representen la operacionalización de habilidades o conocimientos observables, así como el nivel mínimo o aceptable esperado por la ANEP. Una vez que se cuente con la definición curricular operacional y observable acerca de lo que se espera que los estudiantes sean capaces de hacer al egresar de los diversos ciclos o grados educativos, será necesario alinear la prueba al currículo. Esto implica atender a distintos procesos técnicos vinculados al "mantenimiento del instrumento de evaluación" (INEE, 2017: 48-49). Dado que el establecimiento de dichos puntos de corte cristaliza las reglas mediante las cuales se interpretan y utilizan los resultados de las evaluaciones, la validez de dichas interpretaciones depende de la validez de los procedimientos utilizados para su establecimiento (AERA/APA/NMCE, 2014). En el entendido de que la validez es un proceso de continua construcción, porque se orienta a las interpretaciones que se realizan de los resultados, es fundamental tomar en cuenta que para nuestro país Aristas constituye un insumo para retroalimentar el diseño curricular que incluye las definiciones operacionales necesarias para poder alinear la prueba al currículo y así brindar evidencias respecto al grado en el cual el sistema educativo está alcanzando lo que se propone como logros para sus estudiantes en lectura y matemática.



Si bien existen diversas metodologías para la determinación de los puntos de corte en la escala construida en una prueba, todos los procedimientos metodológicos involucran a personas (jueces) que realizan juicios (Cizek y Bunch, 2007). Para asegurar la validez del resultado, los jueces deben: a) ser capacitados sobre la lógica y la mecánica de los procedimientos a utilizar²⁷; b) ser especialistas en los contenidos evaluados en las pruebas; c) tener experiencia de trabajo con las poblaciones evaluadas y d) conocer las características del sistema que se está evaluando²⁸.

En el caso de Aristas, los jueces son docentes uruguayos, especialistas en las áreas evaluadas, capacitados en la lógica y mecánica de los procedimientos utilizados para establecer los puntos de corte en la escala de puntajes.

Dada la centralidad del proceso de establecimiento de los puntos de corte para la validez general de la evaluación, el INEEd seleccionó jueces que cumplieran con los siguientes criterios: a) ser docente especializado en las áreas evaluadas en Aristas: lectura o matemática y b) tener experiencia en el diseño de pruebas estandarizadas.

2.3. Selección del método de jueceo

Existe una multiplicidad de métodos de jueceo para el establecimiento de puntos de corte. Algunos se centran en las tareas de los tests, otros en los sujetos de la evaluación; algunos están diseñados para ser utilizados solo con ítems de respuesta cerrada y otros métodos admiten la utilización de ítems de respuesta construida.

Cizek y Bunch (2007) identifican al menos cinco cuestiones a tomar en consideración a la hora de decidir por un método u otro:

1. El propósito de las pruebas es el primer elemento a tomar en cuenta. Si se trata de una evaluación cuyo propósito es la selección de un determinado número de postulantes para un premio o para un puesto laboral, no se necesita más que un simple método de corte referido a norma.
2. El nivel de complejidad de los conocimientos o habilidades medidos por la prueba, así como de los ítems utilizados.
3. El formato de las pruebas. Algunos métodos solo pueden utilizarse con ítems de opción múltiple (Nedelsky, por ejemplo), otros pueden adaptarse para el procesamiento de ítems de respuesta construida o codificados a través de crédito parcial (Angoff extendido), mientras que otros métodos, como Bookmark, pueden utilizarse con ítems de respuesta cerrada o de respuesta construida, siempre que la razón de ítems de

²⁷ Para ello, un grupo de jueces de cada área —interno y externo al INEEd— es entrenado en la utilización de diversos métodos. A partir de la tabla de dominios de la evaluación, que tiene como referente los documentos curriculares nacionales, los especialistas del INEEd elaboran las descripciones de los diversos niveles y con ese insumo el grupo de jueces establece los puntos de corte en las escalas correspondientes.

²⁸ Los estándares para la evaluación educativa y psicológica, publicados por la American Educational Research Association, la American Psychological Association y el National Council on Measurement in Education establecen la importancia de seleccionar un jurado lo suficientemente grande y representativo de los especialistas en las áreas evaluadas, de manera de tener una razonable seguridad de que los resultados no tendrían grandes variaciones si el proceso se repitiera con otro conjunto de jueces (AERA/APA/NMCE, 2014: 101).

respuesta cerrada sobre los de respuesta construida sea lo suficientemente amplia (es decir, que la cantidad de ítems de respuesta construida sea bastante menor a los de respuesta cerrada).

4. El número de categorías de logro necesarias y la cantidad de ítems en las pruebas.
5. Los recursos temporales y presupuestales. Algunos métodos (Angoff, por ejemplo) necesitan de la emisión de muchos juicios por parte de los jueces, lo cual impacta en el tiempo y los recursos necesarios para la culminación del proceso.

Dados los propósitos de las pruebas aplicadas, su formato, el número de categorías necesarias, así como los recursos con que cuenta el INEE, se decidió utilizar los métodos de Bookmark y Cloud Delphi²⁹.

El método de Bookmark es uno de los más utilizados para el establecimiento de los puntos de corte en pruebas estandarizadas. Esto se debe a diversas razones:

1. se puede utilizar para evaluaciones de formato mixto en las que se utilizan tanto ítems cerrados como abiertos o de respuesta construida;
2. desde la perspectiva de los jueces, presenta una colección de tareas relativamente simple³⁰;
3. también es comparativamente simple para quienes llevan adelante las sesiones de establecimiento de los puntos de corte, dado que las tareas más importantes pueden realizarse de manera previa (descripción de los niveles de desempeño, programación de los aspectos computacionales del procedimiento); y
4. desde una perspectiva psicométrica el método tiene ciertas ventajas, dado que se basa en los modelos de teoría de respuesta al ítem (TRI), es decir, es un método que guarda fidelidad con los procedimientos de construcción y calibración de ítems (Cizek y Bunch, 2007: 159 y 160).

El método Cloud Delphi combina las técnicas desarrolladas por García, Abad, Olea y Aguado (2013) para la determinación de puntos de corte, con la teoría de toma de decisiones multiatributo difusa o modelo de nube (Cloud Delphi) para la toma de decisiones (Yang, Zeng y Zhang, 2012).

En cuanto al componente de toma de decisiones, se utiliza como método para estabilizar las opiniones de los jueces a través de la comparación de sus propios resultados de jueceo con los del resto de los jueces. Una vez que la opinión de cada juez se estabiliza (es decir que la iteración K del especialista j tiene muy poca diferencia con la iteración k-1) se aplica el método de García, Abad, Olea y Aguado (2013) para la determinación de los puntos de corte.

²⁹ Existe un consenso claro en que la elección del método de jueceo tiene una influencia importante en los puntos de corte resultantes (Dochy y otros, 2009).

³⁰ Para ilustrar la "simplicidad" de este método, Cizek y Bunch (2007) presentan el siguiente ejemplo: considerando una prueba que contenga 60 ítems cerrados y 4 abiertos y con la necesidad de establecer 4 niveles de desempeño, el método de Angoff implica que los jueces deban emitir 192 juicios (tres juicios probabilísticos para cada ítem), mientras que con el método de Bookmark solo deberán emitir tres juicios probabilísticos en total, aunque deban considerar el contenido de todos los ítems de la prueba.

El procedimiento para la aplicación del método de Cloud Delphi Ponderado es el siguiente:

- Al igual que el método de Bookmark la descripción de los niveles de desempeños es el referente conceptual con el que los jueces realizan sus juicios.
- Se toma una muestra aleatoria de ítems del banco de reactivos utilizado para la calibración de los ítems. Con esta muestra de ítems trabajan los jueces durante todo el procedimiento.
- Para la determinación de, por ejemplo, 5 niveles se trabaja con una escala continua de 0 a 6, donde los puntos 1, 2, 3, 4 y 5 representan los puntos de corte de los niveles. Se le solicita a los jueces que den intervalos en dicha escala donde cada uno de los ítems tiene una probabilidad de $\frac{1}{2}$ de ser contestado correctamente³¹.
- Esto se realiza para todos los ítems. Luego de la emisión del juicio, se presenta la información del juicio de cada uno de los jueces para cada uno de los ítems, en términos de la discrepancia o cercanía de cada juicio con respecto al juicio del resto de los jueces. A esto se le llama el modelo de nube.
- Esta iteración se continúa hasta que se estabilizan las opiniones de cada uno de los jueces (en donde la iteración k del especialista j tiene poca distancia con la iteración $k-1$).
- De manera posterior a la estabilización de los juicios se comienza con el cálculo de los puntos de corte:
 - Se calcula el punto medio de cada intervalo de cada ítem/especialista.
 - Luego, para cada ítem, se calcula la mediana de los puntos del paso anterior. Estos valores se utilizan para construir los pesos de cada ítem en la definición de los puntos de corte³².
 - Una vez que se tienen los pesos de cada ítem se estima la curva característica promedio para cada nivel, sumando cada curva ponderada por los pesos correspondientes de cada ítem.
 - Luego, se calculan las curvas conjuntas para cada nivel. Para el nivel 1 se toma la curva característica de dicho nivel. Para los niveles superiores se utiliza la productoria de las curvas de los niveles anteriores. Por ejemplo: la curva conjunta para el punto de corte del nivel 3 es el producto de la curva promedio del nivel 1, 2 y 3.
 - Estas curvas indican la probabilidad de contestar los ítems del nivel 3 (siguiendo con el ejemplo) dado que se contestaron correctamente los ítems de los niveles anteriores. Definiendo una probabilidad de contestar correctamente de $\frac{1}{2}$, se identifican los valores donde caen los puntos de corte para cada nivel.

³¹ Por ejemplo, si un juez considera que un ítem pertenece al nivel 1 y dentro de dicho nivel tiene una dificultad media, podría dar un intervalo de 1.3 a 1.6. Es decir que está dentro del nivel pero no está cerca del punto de corte. Pero si el juez considera que el ítem está en la frontera de lo descrito en los niveles 1 y 2, podría dar un intervalo de 1.9 a 2.2.

³² Por ejemplo, si para un ítem la mediana de los puntos medios de los intervalos que definieron los jueces fue de 3.5, entonces la distancia de ese ítem para el punto de corte del nivel 3 será de 0.5 (= 3 - 3.5). En cambio, el mismo ítem para establecer el punto de corte del nivel 2 será de 1.5 (= 2 - 3.5). Cuanto más grande es la distancia del ítem, menos influencia tendrá dicho ítem en la estimación del punto de corte del nivel específico. En este ejemplo, tendrá un peso mucho mayor para la determinación del punto de corte del nivel 3 que para la determinación del punto de corte del nivel 2.

BIBLIOGRAFÍA

- ABAD, Francisco José; OLEA, Julio; PONSODA, Vicente y GARCÍA, Carmen (2011), *Medición en Ciencias Sociales y de la Salud*, Síntesis, Madrid.
- AERA/APA/NCME (2014), *Standards for educational and psychological testing*, American Psychological Association, Washington D.C.
- ANEP (2014), *Aportes iniciales a la discusión sobre fundamentos y perfiles de la educación media básica*, ANEP, Montevideo.
- ANEP (2015), *Mapa de Ruta de Promoción de la Convivencia para Instituciones de Educación Media*, ANEP, Montevideo.
- ANEP (2017), *Documento Marco Curricular de Referencia Nacional: MCRN una construcción colectiva*, ANEP, Montevideo.
- BENTANCOR, Gabriela (coord.) (2004), *Proyecto Convivencia: el Centro Educativo como espacio de Aprendizajes*, UNICEF, Montevideo.
- BOND, Trevor G. y Fox, Christine M. (2013), *Applying the Rasch model: Fundamental measurement in the human sciences*, Psychology Press.
- CEIP (2008), *Programa para Educación Inicial y Primaria*, CEIP, Montevideo.
- CEIP (2016a), *Orientaciones de políticas educativas del Consejo de Educación Inicial y Primaria. Quinquenio 2016 – 2020*, CEIP, Montevideo.
- CEIP (2016b), *Documento Base de Análisis Curricular. Perfiles de Egreso*, CEIP, Montevideo.
- CES (2006a), *Idioma Español. Propuesta programática. Primer año ciclo básico. Reformulación 2006*, CES, Montevideo.
- CES (2006b), *Idioma Español. Segundo año ciclo básico. Propuesta programática - Reformulación 2006*, CES, Montevideo.
- CES (2006c), *Programa de Idioma Español. Tercer año ciclo básico. Reformulación 2006*, CES, Montevideo.
- CES (2007), *Programa de Literatura. 3° ciclo básico. Reformulación 2007*, CES, Montevideo.
- CES (2010a), *Programa de Matemática primer año. Ciclo básico. Reformulación 2006 – Ajuste 2010*, CES, Montevideo.
- CES (2010b), *Programa de Matemática segundo año. Ciclo básico. Reformulación 2006 – Ajuste 2010*, CES, Montevideo.
- CES (2010c), *Programa de Matemática tercer año. Ciclo básico. Reformulación 2006 – Ajuste 2010*, CES, Montevideo.
- CES (2016), *Expectativas de logro por asignatura y por nivel del Ciclo Básico del CES*, CES, Montevideo.
- CETP (2007a), *Programa de Matemática. Primer año de ciclo básico tecnológico*, CETP, Montevideo.
- CETP (2007b), *Programa de Matemática. Segundo año de ciclo básico tecnológico*, CETP, Montevideo.



- CETP (2007c), *Programa de Matemática. Tercer año de ciclo básico tecnológico*, CETP, Montevideo.
- CETP (2007d), *Programa de Idioma Español. Primer año, ciclo básico tecnológico*, CETP, Montevideo.
- CETP (2007e), *Programa de Idioma Español. Segundo año, ciclo básico tecnológico*, CETP, Montevideo.
- CETP (2007f), *Programa de Literatura. Tercer año de ciclo básico tecnológico*, CETP, Montevideo.
- CIZEK, Gregory J. y BUNCH, Michael B. (2007), *Standard setting: A guide to establishing and evaluating performance standards for tests*, Sage Publications, Thousand Oaks.
- COLL, César (2007), “Las competencias en la educación escolar: algo más que una moda y mucho menos que un remedio”, *Revista Aula*, 161, 34-39.
- COLOMBO, Graciela Beatriz (2011), “Violencia Escolar y Convivencia Escolar: Descubriendo estrategias en la vida cotidiana escolar”, *Revista Argentina de Sociología*, 8-9(15-16), 81-104, Buenos Aires.
- CONTERI, Carla y RODRÍGUEZ, Carmen, (2012), *En palabras de los estudiantes. Sistematización de los encuentros nacionales de estudiantes de secundaria*, UNICEF, Montevideo.
- DOCHY, Filip; KYNDT, Eva; BAETEN, Marlies; POTTIER, Sofie y VEESTRAETEN, Marlies (2009), “The effects of different standard setting methods and the composition of borderline groups: A study within a law curriculum”, *Studies in Educational Evaluation*, 35(4), 174-182.
- ESCOBAR-PÉREZ, Jazmine y CUERVO-MARTÍNEZ, Ángela (2008), “Validez de contenido y juicio de expertos: una aproximación a su utilización”, *Avances en Medición*, 6(1), 27-36.
- ETS (2005), *Guidelines for Constructed-Response and Other Performance Assessments*, Educational Testing Service.
- ETS (2009), *ETS Guidelines for Fairness Review of Assessments*, Educational Testing Service
- ETS (2014), *ETS Standards for Quality and Fairness*, Educational Testing Service.
- GARCÍA, Pablo Eduardo; ABAD, Francisco José; OLEA, Julio y AGUADO, David (2013), “A new IRT based setting method: Application to eCat-Listening”, *Psicothema*, 25(2), 238-244.
- HALADYNA, Thomas; DOWNING, Steven y RODRIGUEZ, Michael (2002), “A review of multiple-choice item writing guidelines for classroom assessment”, *Applied Measurement in Education*, 15(3), 309-334.
- IEE (2009), *Marco de la evaluación general de diagnóstico*, Instituto de Evaluación, Ministerio de Educación de España, Madrid.
- INEE (2017), *Criterios técnicos para el desarrollo, uso y mantenimiento de instrumentos de evaluación*, INEE, Ciudad de México.
- INEEd (2017a), *Aristas. Marco de lectura en tercero y sexto de educación primaria*, INEEEd, Montevideo.
- INEEd (2017b), *Aristas. Marco de matemática en tercero y sexto de educación primaria*, INEEEd, Montevideo.
- INEEd (2017c), *Aristas. Marco de lectura en tercero de educación media*, INEEEd, Montevideo.
- INEEd (2017d), *Aristas. Marco de matemática en tercero de educación media*, INEEEd, Montevideo.
- MISLEVY, Robert J.; ALMOND, Russell G. y LUKAS, Janice F. (2003), *A Brief Introduction to Evidence-centered Design*, ETS, Research & Development Division Princeton, Princeton.



- NAGB (2003), *Background Information Framework for the National Assessment of Educational Progress*, National Assessment Governing Board, Washington, D.C.
- NEVO, Baruch (1985), "Face validity revisited", *Journal of Educational Measurement*, 22(4), 287-293.
- PRIETO, Gerardo y DELGADO, Ana Rosa (1996), "Construcción de Ítems", en Muñiz, José (coord.), *Psicometría*, Universitas, Madrid.
- R CORE TEAM (2017), *R: A language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna.
- ROBITZSCH, Alexander; KIEFER, Thomas y WU, Margaret (2017), *TAM: Test analysis modules*, R package version 2.6-2.
- SCHMIDT, William H. (ed.) (1996), *Characterizing Pedagogical Flow. An Investigation of Mathematics and Science Teaching in Six Countries*, Kluwer Academic Publishers, Dordrecht.
- SCHMIDT, William; COGAN, Leland y SOLORIO, Michelle (2017), "The Missing Link- Incorporating Opportunity to Learn in Educational Research Analyses", en Son, Ji-Won; Watanabe, Tad y Lo, Jane-Jane (eds.), *What Matters? Research Trends in International Comparative Studies in Mathematics Education*, Research in Mathematics Education, Springer, Cham.
- SMITH, Vanessa y MOLINA, Mauricio (2011), *La entrevista cognitiva: guía para su aplicación en la evaluación y mejoramiento de instrumentos de papel y lápiz*, Universidad de Costa Rica, San José de Costa Rica.
- SNEP (2016), *Plan Nacional de Educación en Derechos Humanos*, SNEP, Montevideo.
- VAN DER LINDEN, Wim J. y HAMBLETON, Ronald K. (eds.) (2013), *Handbook of modern item response theory*, Springer Science & Business Media.
- VISCARDI, Nilia y ALONSO, Nicolás (2013), *Gramática(s) de la convivencia. Un examen a la cotidianidad escolar y la cultura política en la Educación Primaria y Media en Uruguay*, ANEP, Montevideo.
- ZIEKY, Michael J. (2014), "An introduction to the use of evidence-centered design in test development", *Psicología Educativa*, 20, 79-87.



aristas.ineed.edu.uy



INEEd
Instituto Nacional de
Evaluación Educativa



Aristas
Evaluación Nacional
de Logros Educativos