Agencia de Calidad de la Educación

Informe Técnico
Simce 2015

INFORME TÉCNICO
SIMCE 2015

Agencia de
Calidad de la
Educación

En el presente documento se utilizan de manera inclusiva términos como "el docente", "el estudiante" y sus respectivos plurales (así como otras palabras equivalentes en el contexto educativo) para referirse a hombres y mujeres. Esta opción se basa en la convención idiomática de nuestra lengua y tiene por objetivo evitar las fórmulas que aluden a ambos géneros en el idioma español ("o/a", "los/las" y otras similares), debido a que implican una saturación gráfica que puede dificultar la comprensión de lectura. Asimismo, a objeto de que este informe resulte funcional, todos los datos específicos (planillas, tablas, gráficos, bases de datos, etc.) han sido omitidos de la versión impresa y su acceso se encuentra reseñado en los apartados correspondientes.

INFORME TÉCNICO
SIMCE 2015

Agencia de Calidad de la Educación
División de Evaluación de Logros de Aprendizaje
Morandé 360
Santiago de Chile, 2018
www.agenciaeducacion.cl

ISBN: 978-956-9484-11-7

PRESENTACIÓN

El Sistema de Aseguramiento de la Calidad (SAC), liderado por el Ministerio de Educación (Mineduc) está compuesto por la Superintendencia de Educación (SUPEREDUC), el Consejo Nacional de Educación (CNED) y la Agencia de Calidad de la Educación. La misión del SAC es lograr mayor equidad y calidad en la educación. Para ello, cada uno de los organismos que lo componen cumple tareas específicas.

La Agencia de Calidad de la Educación tiene como propósito evaluar, orientar e informar a la comunidad escolar respecto de los aprendizajes de los estudiantes y los procesos de gestión escolar. En este contexto, una de sus principales tareas consiste en diseñar, implementar y aplicar un sistema de medición de los resultados de aprendizaje de carácter censal, referido a los objetivos del Currículo Nacional, de acuerdo a un Plan de Evaluaciones aprobado por el CNED.

Para dar cumplimiento a este propósito, la Agencia administra y aplica pruebas anuales estandarizadas a nivel censal en al menos un nivel de enseñanza básica y uno de enseñanza media: las pruebas Simce. Estas evaluaciones, que son obligatorias para todos los establecimientos educacionales reconocidos oficialmente por el Estado, miden logros de aprendizaje en diferentes áreas académicas y grados escolares. Junto con lo anterior, la Agencia aplica Cuestionarios de Calidad y Contexto Educativo que entregan información sobre los Indicadores de Desarrollo Personal y Social (IDPS) de los estudiantes. Con los antecedentes que otorgan ambos instrumentos, se busca informar sobre la calidad educativa desde una concepción más amplia que involucra no solo los logros de aprendizaje, sino también aspectos relevantes del contexto en que se producen tales aprendizajes.

Para continuar con los esfuerzos por compartir los procesos asociados a estas evaluaciones, presentamos el cuarto informe técnico que detalla cada uno de los procesos y aspectos técnicos que se consideraron para la correcta aplicación de las pruebas Simce 2015, desde la etapa de diseño de los instrumentos hasta la comunicación de sus resultados.

Ponemos este documento a disposición de investigadores, académicos, profesionales de la educación y la comunidad en general, con el propósito de proveer información relevante respecto de nuestros Macroprocesos, los que permiten la implementación y realización de la evaluación Simce. De esta forma buscamos colaborar en la difusión de la información y abrir camino para el análisis y buen uso de la misma, aportando así en el avance hacia una mayor calidad y equidad de nuestra educación.

Carlos Henríquez Calderón
Secretario Ejecutivo
Agencia de Calidad de la Educación

Contenido

Introducción 9

Capítulo 1: Diseño y construcción de las pruebas Simce 11

1.1 Definición del Marco de Referencia y formulación de especificaciones técnicas 13

1.2 Elaboración y validación de las preguntas 22

1.3 Armado y diagramación de las pruebas 24

1.4 Prueba muestral de Educación Física 26

Capítulo 2: Logística y aplicación de las pruebas Simce 29

2.1 Empadronamiento de establecimientos y estudiantes 32

2.2 Impresión, mecanizado y distribución del material de prueba 34

2.3 Aplicación de las pruebas 36

2.4 Entrega del material a los centros de captura 43

Capítulo 3: Procesamiento y análisis de datos 45

3.1 Elaboración y validación de la base de datos de pruebas y cuestionarios 47

3.2 Análisis psicométricos de las pruebas Simce 52

3.3 Análisis de datos agregados 61

Capítulo 4: Comunicación de resultados educativos 67

4.1 Documentos informativos previos a la aplicación 70

4.2 Informes de resultados 70

4.3 Otros documentos comunicacionales 72

Anexos 75

Anexo A: Pruebas experimentales Simce 2015 77

Anexo B: Corrección de preguntas de respuesta abierta 82

Referencias bibliográficas 86

Listado de Tablas

Tabla 1.1: Matriz de evaluación teórica de pruebas Simce Lectura 2015 14

Tabla 1.2: Matriz de evaluación teórica de la Prueba Simce Escritura 2015. Nivel 6° básico 16

Tabla 1.3: Matriz de evaluación teórica de pruebas Simce Matemática 2015. Ejes de habilidades 16

Tabla 1.4: Matriz de evaluación teórica de pruebas Simce Matemática 2015. Niveles 4° y 6° básico. Ejes temáticos 17

Tabla 1.5: Matriz de evaluación teórica de pruebas Simce Matemática 2015. Niveles 8°básico y II medio. Ejes temáticos 18

Tabla 1.6: Matriz de evaluación teórica de pruebas Simce Historia, Geografía y Cs. Sociales 2015. Ejes de habilidades 19

Tabla 1.7: Matriz de evaluación teórica de la Prueba Simce Historia, Geografía y Cs. Sociales 2015. Nivel 6° básico. Ejes temáticos 20

Tabla 1.8: Matriz de evaluación teórica de la Prueba Simce Historia, Geografía y Cs. Sociales 2015. Nivel II medio. Ejes temáticos 20

Tabla 1.9: Matriz de evaluación teórica de la prueba Simce Ciencias Naturales 2015. Nivel 8° básico. Ejes de habilidades 21

Tabla 1.10: Matriz de evaluación teórica de la prueba Simce Ciencias Naturales 2015. Nivel 8° básico. Ejes temáticos 22

Tabla 2.1: Códigos SIGE de establecimientos según tipo de enseñanza impartida 33

Tabla 2.2: Establecimientos y estudiantes evaluados en la aplicación censal de las pruebas Simce 2015 38

Tabla 2.3: Establecimientos y estudiantes evaluados en la aplicación muestral del Educación Física 2015. Nivel 8° básico 38

Tabla 3.1: Códigos de la variable de identificación por nivel escolar de la aplicación 2015 48

Tabla 3.2: Categorías de la variable de participación de estudiantes en la aplicación 2015 49

Tabla 3.3: Datos capturados en 2015 por nivel escolar y prueba aplicada 50

Tabla 3.4: Indicadores 2015 para preguntas de respuesta cerrada según Teoría Clásica e IRT 54

Tabla 3.5: Indicadores 2015 para preguntas de respuesta abierta según Teoría Clásica e IRT 54

Tabla 3.6: Puntajes de corte para la clasificación 2015 según Estándares de Aprendizaje 57

Tabla 3.7: Agrupación de análisis agregados de las pruebas regulares Simce 2015 62

Tabla 3.8: Clasificación de establecimientos por tamaño en la aplicación 2015 64

Tabla 4.1: Documentos comunicacionales de la aplicación Simce 2015 73

Tabla A1: Categorías de rendimiento Simce por prueba y nivel escolar en enseñanza básica 78

Tabla A2: Categorías de rendimiento Simce por prueba en II medio....78
Tabla A3: Estudiantes según nivel escolar en las PE2015 y porcentaje de la muestra teórica que representan....79
Tabla A4: Código y descripción de los campos de identificación de los ítems de las PE2015....80
Tabla A5: Código y descripción de los campos de datos métricos de los ítems de las PE2015....81
Tabla B1: Distribución de correctores de preguntas abiertas pruebas censales Simce 2015....82
Tabla B2: Cantidad de preguntas abiertas censales 2015....84

Listado de Figuras

Figura 3.1: Ejemplo de curva empírica del ítem....59
Figura 3.2: Ejemplo de gráfico de tendencia de puntaje a través del tiempo....65

Introducción

Las pruebas Simce censales evalúan el aprendizaje de todos los estudiantes del país, en relación con el Currículum Nacional, en diferentes asignaturas y niveles escolares, con el propósito de contribuir al proceso de mejoramiento de la calidad y equidad de la educación.

La información obtenida a través de los resultados de estas pruebas se pone a disposición de los establecimientos con el objetivo de que estos analicen y reflexionen sobre los aprendizajes alcanzados por sus estudiantes. Para ello, se informa tanto el puntaje promedio obtenido como la distribución de estudiantes en cada nivel de los Estándares de Aprendizaje definidos por el Ministerio de Educación, de manera que sea posible identificar desafíos y fortalezas que contribuyan al mejoramiento continuo de los aprendizajes de los estudiantes.

Para obtener los resultados se utiliza fundamentalmente el modelo de medición IRT (Teoría de Respuesta al ítem), pues este se centra en las propiedades de los ítems individuales, permitiendo establecer la probabilidad de contestar correctamente a cada pregunta, de acuerdo al nivel de habilidad del estudiante. En particular, el modelo IRT de tres parámetros, permite estimar la habilidad de un estudiante, basándose en la probabilidad de respuesta correcta según tres características propias de las preguntas: dificultad, discriminación y azar. Este modelo es ampliamente compatible con la evaluación en base a criterios o estándares y permite comparar puntajes en el tiempo y en cohortes diferentes, por medio de la aplicación de un proceso de equiparación estadística.

Actualmente, las pruebas Simce se aplican en los niveles 2°, 4°, 6° y 8° de enseñanza básica y en II año de educación media, abordando contenidos curriculares de las asignaturas de Lenguaje y Comunicación, Matemática, Ciencias Naturales e Historia, Geografía y Ciencias Sociales. Sin embargo, no todos los años se aplican todas las pruebas en todos los niveles. La aplicación de cada año está determinada por un Plan de Evaluaciones.

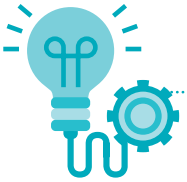
En 2015, en función del Plan Nacional de Evaluaciones Nacionales e Internacionales 2011-2015, correspondió aplicar un total de 13 pruebas censales: cinco de Lectura, una de Escritura, cuatro de Matemática, dos de Historia, Geografía y Ciencias Sociales, y una de Ciencias Naturales, distribuidas según se muestra en el siguiente cuadro:

Pruebas Simce censales aplicadas en 2015

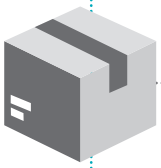
Pruebas Simce / Niveles evaluados	2° básico	4° básico	6° básico	8° básico	II medio
Lectura	✔	✔	✔	✔	✔
Escritura			✔		
Matemática			✔	✔	✔
Historia, Geografía y Ciencias Sociales			✔		✔
Ciencias Naturales				✔	

Adicionalmente, ese año se aplicó una prueba muestral de Educación Física en 8° básico.

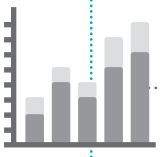
El ciclo evaluativo de las pruebas Simce 2015 consta de cuatro etapas que dan origen a los capítulos de este informe. La primera etapa comprende el diseño y construcción de los instrumentos, incluyendo la definición del Marco de Referencia y la formulación de especificaciones técnicas. Asimismo, supone la elaboración y validación de las preguntas; la selección definitiva de los ítems que formarían parte de la aplicación censal; los procesos requeridos para establecer comparabilidad entre las mediciones de distintos años y, el ensamblaje y diagramación de los cuadernillos. En el segundo capítulo se describe la etapa que contempla las operaciones de campo, incluyendo los procesos de empadronamiento, imprenta, distribución y aplicación censal de las pruebas. En el tercer capítulo se aborda el procesamiento y análisis de los datos, desde la captura de las respuestas hasta los análisis psicométricos y agregados que se realizan con los resultados. Finalmente, en el cuarto capítulo se describe el proceso de comunicación de resultados de las pruebas.



Capítulo I:
Diseño y construcción de las pruebas Simce



Capítulo II:
Logística y aplicación de las pruebas Simce



Capítulo III:
Procesamiento y análisis de datos

Capítulo IV:
Comunicación de resultados educativos

Capítulo I : Diseño y construcción de pruebas Simce 2015

La construcción de las pruebas Simce censales se lleva a cabo mediante un proceso que dura aproximadamente un año y medio, y en el que se resguarda el cumplimiento de estándares internacionales, como los definidos en el documento Standards for Educational and Psychological Testing (AERA, APA y NCME, 2014).

Este proceso, cuyo flujo se inicia con la definición del Marco de Referencia y las especificaciones técnicas, continúa con la elaboración y validación de las preguntas, y culmina con el armado y diagramación de las formas de prueba definitivas que se aplican a los estudiantes.

En este capítulo se describe cada una de las etapas de construcción de pruebas censales Simce 2015 y, en el último apartado, se detalla también la construcción de una prueba de carácter muestral que se aplicó ese año en 8° básico, en la asignatura de Educación Física.

1.1 Definición del Marco de Referencia y formulación de especificaciones técnicas

El primer paso en la construcción de las pruebas Simce consiste en definir el Marco de Referencia, esto es, los documentos que delimitan los constructos a evaluar y que otorgan validez a las evaluaciones, según indican los estándares internacionales (AERA, APA y NCME, 2014; Case, Jorgensen y Zucker, 2004).

Considerando que las pruebas Simce son referidas al Currículo Nacional, el Marco de Referencia está dado por este. En particular, en las pruebas 2015 de los niveles de enseñanza básica 2°, 4° y 6°, el Marco de Referencia lo constituyen las Bases Curriculares establecidas mediante el Decreto N.° 439/2012 (Mineduc, 2011a), y en las pruebas de los niveles 8° básico y II medio, el Marco de Referencia es el Marco Curricular ajustado en 2009, derivado de los decretos supremos de educación, números 254 y 256 (Mineduc, 2009c).

A partir de este Marco de Referencia se generan las especificaciones técnicas de las pruebas, esto es, los documentos que establecen las características particulares de diseño de los instrumentos, y que buscan garantizar cobertura y comparabilidad en la medición de los aprendizajes curriculares. Las especificaciones técnicas las elabora el Departamento de Construcción de Pruebas (DCP) de la Agencia de Calidad y las revisa la Unidad de Currículo y Evaluación (UCE) del Ministerio de Educación.

Las especificaciones técnicas de las pruebas distinguen y definen ejes temáticos (o de contenidos) y ejes de habilidades (o de dominios cognitivos) y además establecen las matrices de evaluación teóricas, que señalan los porcentajes de preguntas con los que serán representados los ejes en la prueba. Estos porcentajes se definen a partir de un análisis de énfasis curriculares que realizan en conjunto profesionales de la Agencia y de la UCE.

1.1.1 Pruebas de Lectura

En las pruebas de Lectura 2015, en todos los niveles evaluados, se distinguen tres ejes de habilidades: localizar, interpretar y relacionar, y reflexionar, cuya matriz de evaluación teórica se muestra en la Tabla 1.1.

El eje **Localizar** agrupa las tareas de lectura que el estudiante debe emplear para operar con elementos explícitos de los textos, tales como las habilidades de identificar, discriminar y extraer información explícita de las lecturas.

El eje **Interpretar y relacionar** incluye las tareas de lectura relativas a construir significados implícitos, estableciendo conexiones entre elementos explícitos. En este eje se encuentran las habilidades de inferir información, interpretar lenguaje figurado y reconocer relaciones causales, entre otras.

El eje **Reflexionar** incorpora las tareas de lectura relativas a la capacidad de confrontar distintos aspectos del texto (formales y/o de contenido) con conocimientos previos, experiencia personal, conocimiento de mundo y lecturas anteriores, permitiendo al lector examinar el texto y evaluarlo críticamente. En este eje se incluye, por ejemplo, la capacidad de emitir opiniones sobre lo leído y de determinar el aporte de elementos gráficos al sentido de un texto.

Los ejes antes señalados se evalúan a partir de textos literarios y no literarios. Los primeros se entienden como aquellos que tienen una intención estética de base y constituyen un testimonio de la experiencia humana, que abre un espacio a la reflexión de aspectos personales, valóricos y sociales; los segundos se definen como aquellos que tienen una intención comunicativa focalizada en la conexión entre el autor y el lector, es decir, en la transmisión de un mensaje determinado. Adicionalmente, en cada uno de estos tipos de textos, se distinguen subcategorías: en los textos literarios se incluyen textos narrativos, líricos y dramáticos; y en los no literarios se distinguen textos informativos, persuasivos e instruccionales. Los textos informativos, a su vez, se presentan en tres tipos de formatos: continuo (formado por oraciones que se organizan en párrafos), discontinuo (formado por una serie de listas, gráficos o diagramas, entre otros), y mixto (conformado por textos de formato continuo y discontinuo).

Tabla 1.1

Matriz de evaluación teórica de pruebas Simce Lectura 2015

Ejes de habilidades	2° básico	4° básico	6° básico	8° básico	II Medio
Localizar	35%	25%	25%	15%	15%
Interpretar y relacionar	45%	45%	45%	45%	45%
Reflexionar	20%	30%	30%	40%	40%

1.1.2 Prueba de Escritura

En la prueba de Escritura de 6º básico 2015 se evalúan, a través de textos escritos por los estudiantes[1], habilidades asociadas a la competencia de expresión escrita, tales como la capacidad de producir textos que tengan un propósito comunicativo claro, una adecuada organización de las ideas, que sean coherentes, que desarrollen las ideas acabadamente y que incorporen un vocabulario preciso y variado, junto con emplear correctamente las estructuras gramaticales y los recursos de la lengua que permiten transmitir los mensajes con claridad y facilitar la comprensión por parte del lector.

En esta prueba se miden ocho habilidades que se describen a continuación:

1. Adecuación al propósito comunicativo: capacidad de escribir un texto adecuándose a la intención comunicativa solicitada.
2. Organización de las ideas: habilidad de seleccionar un esquema global adecuado para sistematizar las ideas del escrito.
3. Coherencia textual: capacidad de seleccionar ideas y de ordenarlas en una secuencia que mantiene las relaciones de sentido entre ellas, tanto a nivel local como global.
4. Desarrollo de ideas: habilidad para desarrollar con suficiente detalle las ideas relevantes del texto, de modo que el lector logre una comprensión acabada del mensaje.
5. Uso de vocabulario preciso y variado: capacidad de utilizar las palabras de modo de comunicar las ideas de manera clara y precisa.
6. Cohesión: habilidad para relacionar las ideas mediante recursos de cohesión apropiados y suficientes, tales como conjunciones, preposiciones, pronombres relativos, etc.
7. Organización de las ideas mediante puntuación: habilidad para separar las ideas y organizarlas dentro del texto, utilizando correctamente los signos de puntuación, de acuerdo con las reglas de la lengua.
8. Aplicación de normas de ortografía literal y acentual: capacidad para aplicar correctamente en la escritura de palabras, las normas de ortografía literal y acentual establecidas por la Real Academia Española, con el fin de permitir una lectura fluida del texto.

La matriz de evaluación teórica de la prueba de Escritura, que se muestra en la Tabla 1.2, se estructura sobre la base de dos objetivos de evaluación que incorporan las habilidades antes mencionadas y que se distinguen en función del tipo de texto:

- Objetivo de Evaluación 1: producir textos que tengan como propósito narrar, que presenten una organización textual acorde a este tipo de texto (situación inicial, desarrollo -con un conflicto claro- y desenlace), que sean coherentes, que desarrollen sus ideas acabadamente y utilicen un vocabulario preciso y variado, además de utilizar los recursos de la lengua para cohesionar sus ideas y facilitar la comprensión por parte del lector.

[1] La prueba de Escritura 2015 incluyó algunas preguntas de opción múltiple, pero tras evaluar su comportamiento se determinó no utilizarlas en la corrección y puntuación de la prueba porque presentaron problemas de dimensionalidad. En 2016 se prevé realizar un "estudio puente", en el que una muestra de estudiantes rinda la prueba 2015 con todas las preguntas y el resto lo haga sin las preguntas cerradas, con el objeto de resguardar la comparabilidad entre años.

- Objetivo de Evaluación 2: producir textos que tengan como propósito informar, que presenten una organización textual acorde a este tipo de texto (planteamiento de un tema y de ideas subordinadas a este), que sean coherentes, que desarrollen sus ideas acabadamente y utilicen un vocabulario preciso y variado, además de utilizar los recursos de la lengua para cohesionar sus ideas y facilitar la comprensión por parte del lector.

Tabla 1.2

Matriz de evaluación teórica de la Prueba Simce Escritura 2015. Nivel 6° básico

Objetivos de evaluación	6° básico
Objetivo de evaluación 1	50%
Objetivo de evaluación 2	50%

1.1.3 Pruebas de Matemática

En las pruebas Simce de Matemática 2015, en todos los niveles evaluados, se distinguen tres ejes de habilidades: Conocimiento, Aplicación y Razonamiento, cuya matriz de evaluación teórica se muestra en la Tabla 1.3.

El eje Conocimiento comprende el conocimiento factual de conceptos, propiedades y procedimientos matemáticos, así como sus representaciones simbólicas. En este eje se incluyen, entre otras habilidades, las de recordar definiciones, fórmulas y teoremas, métodos de cálculo y procedimientos de resolución de ecuaciones; reconocer propiedades; localizar datos, y clasificar elementos según propiedades dadas.

El eje Aplicación involucra las habilidades relativas a la selección de estrategias adecuadas para resolver un problema y la aplicación de operaciones, procedimientos, etc. Supone, asimismo, la capacidad de generar representaciones equivalentes para comunicar cantidades, operaciones y relaciones dadas, y de representar e interpretar información y datos en diagramas, tablas y gráficos.

El eje Razonamiento incluye el razonamiento intuitivo e inductivo necesario para descubrir patrones y regularidades, para realizar generalizaciones y para buscar estrategias, representaciones y modelos que permitan resolver problemas no rutinarios. Supone también la habilidad de realizar deducciones basadas en reglas y supuestos específicos, de diferenciar afirmaciones erróneas de verdaderas, de argumentar la validez de un resultado obtenido y de comunicar ideas y conclusiones.

Tabla 1.3

Matriz de evaluación teórica de pruebas Simce Matemática 2015. Ejes de habilidades

Ejes de habilidades	4° básico	6° básico	8° básico	II Medio
Conocimiento	50%	50%	45%	50%
Aplicación	35%	35%	45%	30%
Razonamiento	15%	15%	10%	20%

Respecto de los ejes temáticos, dependiendo del nivel escolar, se definieron distintos contenidos.

Así, en las pruebas de Matemática de 4° y 6° básico, cuya matriz de evaluación teórica se muestra en la Tabla 1.4, se distinguen cinco ejes temáticos: Números y operaciones, Patrones y álgebra, Geometría, Medición y Datos y probabilidades.

El eje **Números y operaciones** incluye conceptos básicos de números y algoritmos, tales como el sistema posicional, los algoritmos de la adición, sustracción, multiplicación y división de números naturales, y las propiedades, operaciones y establecimiento de conexiones entre números racionales (fracciones, decimales y porcentajes).

El eje **Patrones y álgebra** incorpora las relaciones entre números, formas, objetos y conceptos; el cambio de una cantidad en relación con otra; las ecuaciones e inecuaciones; las reglas que generan los patrones; el transporte de los patrones de una forma de representación a otra; y el modelado de situaciones sencillas que impliquen las cuatro operaciones.

El eje **Geometría** abarca el reconocimiento, la visualización y la comparación de figuras de dos y tres dimensiones; el conocimiento de sus características y propiedades; y el reconocimiento y la descripción de su movimiento a través de la reflexión, la traslación y la rotación. Incluye asimismo conceptos relativos a la estructura del espacio, los elementos del entorno y los sistemas de coordenadas informales.

El eje **Medición** incluye medición del tiempo y cuantificación y medición de objetos (ancho, largo, alto, peso, perímetro, superficie, volumen, ángulos, etc.), con unidades estandarizadas y no estandarizadas. Supone también la discriminación de las unidades de medición apropiadas al contexto y el reconocimiento de herramientas específicas de medición, en virtud del objeto a ser medido.

El eje **Datos y probabilidades** incluye el registro, clasificación y lectura de información dispuesta en tablas y gráficos presentes en la vida cotidiana, así como la construcción de tablas y gráficos, y la formulación de conclusiones a partir de los datos. Además, el eje incorpora cálculo e interpretación de la media aritmética y conceptos básicos sobre probabilidades, como la capacidad de describir la posibilidad de ocurrencia de un evento aleatorio y la comparación de probabilidades de distintos eventos, sin cálculo.

Tabla 1.4

Matriz de evaluación teórica de pruebas Simce Matemática 2015. Niveles 4° y 6° básico. Ejes temáticos

Ejes temáticos	4° básico	6° básico
Números y operaciones	45%	45%
Geometría	20%	15%
Patrones y álgebra	5%	15%
Datos y probabilidades	10%	15%
Medición	20%	10%

Por su parte, en las pruebas de Matemática de 8° básico y II medio, cuya matriz de evaluación teórica se muestra en la Tabla 1.5, se distinguen cuatro ejes de contenido: Números, Álgebra, Geometría y Datos y azar.

El eje **Números** incluye cantidades y números, operaciones aritméticas (adición, sustracción, multiplicación y división), potencias, raíces y propiedades y relaciones de los conjuntos numéricos (números naturales, enteros, racionales y reales).

El eje **Álgebra** involucra símbolos de representación y operación de cantidades; expresiones de relaciones generales y abstractas de aritmética y medición; lenguaje matemático abstracto; ecuaciones y sistemas de ecuaciones, y el concepto de función, tipos de funciones y su representación gráfica.

El eje **Geometría** abarca la imaginación espacial; el conocimiento de objetos, figuras y cuerpos geométricos básicos y sus propiedades; las relaciones entre formas de dos y tres dimensiones; la construcción y transformación de figuras; la medición de figuras planas y cuerpos; y las transformaciones isométricas, los vectores y la geometría cartesiana.

El eje **Datos y azar** incorpora diversos tipos de tablas y gráficos; los conceptos de muestra y población; el análisis de datos, utilizando herramientas como las medidas de tendencia central, de posición y dispersión, y la teoría de probabilidades, con todos los conceptos asociados a ella como, por ejemplo, espacio muestral, Modelo de Laplace, principios aditivo y multiplicativo y la Ley de los Grandes Números.

Tabla 1.5

Matriz de evaluación teórica de pruebas Simce Matemática 2015. Niveles 8° básico y II medio. Ejes temáticos

Ejes temáticos	8° básico	II medio
Números	35%	30%
Geometría	25%	20%
Álgebra	20%	30%
Datos y azar	20%	20%

1.1.4 Pruebas de Historia, Geografía y Ciencias Sociales

En las pruebas de Historia, Geografía y Ciencias Sociales, en los dos niveles evaluados en 2015 (6° básico y II medio), se distinguen tres ejes de habilidades: Conocimiento y comprensión, Aplicación y Razonamiento, cuya matriz de evaluación teórica se muestra en la Tabla 1.6.

El eje **Conocimiento y comprensión** implica la capacidad del estudiante de recordar, definir, describir y comprender conceptos, hechos, fenómenos y procesos históricos, geográficos y sociales; así como de extraer información de diversas fuentes propias de las disciplinas sociales.

El eje **Aplicación** involucra las habilidades relativas al uso de conceptos y conocimientos históricos, geográficos y sociales, para situar elementos en el tiempo y en el espacio, clasificar y secuenciar información o eventos y transferirlos a situaciones nuevas. Asimismo, supone la capacidad de interpretar información presentada en diversas fuentes.

El eje **Razonamiento** incluye diversas tareas de pensamiento crítico respecto de los procesos y fenómenos de la realidad histórica, geográfica y social. Supone, entre otras habilidades, las de comparar, analizar y evaluar información referida a las Ciencias Sociales, y las de utilizar y problematizar conceptos y situaciones.

Tabla 1.6

Matriz de evaluación teórica de pruebas Simce Historia, Geografía y Ciencias Sociales 2015. Ejes de habilidades.

Ejes de habilidades	6° básico	II medio
Conocimiento y comprensión	60%	55%
Aplicación	15%	15%
Razonamiento	25%	30%

Respecto de los ejes temáticos, dependiendo del nivel escolar, se definieron distintos contenidos.

Así, en la prueba de 6° básico, cuya matriz de evaluación teórica se muestra en la Tabla 1.7, se distinguen tres ejes temáticos: Historia, Geografía y Formación ciudadana.

El eje Historia refiere a una visión panorámica de la historia de Chile, considerando aspectos de continuidad y cambio de la sociedad, desde el proceso de descubrimiento y conquista hasta el quiebre y recuperación del sistema democrático. Entre otros, se incluyen conocimientos sobre los factores que hicieron posible el proceso de expansión europea y las consecuencias de dicho proceso para Europa y América; la caracterización de los ámbitos económico, social y cultural de la Colonia (con énfasis en la relación entre españoles y mapuches); antecedentes y desarrollo del proceso de emancipación nacional; algunos hitos del proceso de organización republicana de Chile; aspectos relativos al desarrollo cultural, económico y social del siglo XIX; la progresiva democratización de la sociedad chilena durante el siglo XX; las diferentes visiones existentes respecto del quiebre democrático de 1973 y de la dictadura militar que se instauró; y la recuperación de la democracia a fines del siglo XX, incluyendo la visión de consenso que existe respecto de su valor como sistema político.

El eje Geografía involucra contenidos sobre geografía de Chile. En particular se incluyen las grandes zonas de Chile y sus paisajes (desde los puntos de vista físico, demográfico y económico); los elementos constitutivos del territorio nacional y las regiones político-administrativas; los recursos y ambientes naturales del país (considerando las oportunidades y dificultades que estos presentan a las personas para vivir y desarrollarse); los principales riesgos naturales que afectan al territorio nacional, así como las diferentes estrategias para enfrentarlos; y el uso de herramientas geográficas para ubicar, caracterizar y relacionar diferentes elementos del espacio y para orientarse, usando categorías de ubicación absoluta y relativa.

El eje Formación ciudadana se centra en el desarrollo de conocimientos que permiten participar activa y responsablemente en la sociedad democrática. Entre los contenidos que se incluyen, se cuentan: los derechos, deberes y responsabilidades de las personas; elementos fundamentales de la organización democrática de Chile, como la Constitución Política; organización comunitaria y soluciones a los problemas comunes; y conocimientos sobre las actitudes que los estudiantes deben desarrollar y demostrar en su vida diaria, como el trabajo en equipo, el esfuerzo, la honestidad, la responsabilidad, el respeto, la participación en la comunidad y el cuidado del medioambiente.

Tabla 1.7

Matriz de evaluación teórica de la Prueba Simce Historia, Geografía y Ciencias Sociales 2015. Nivel 6° básico. Ejes temáticos

Ejes temáticos	6° básico
Historia	45%
Geografía	30%
Formación ciudadana	25%

En la prueba de Historia, Geografía y Ciencias Sociales de II medio, cuya matriz de evaluación teórica se muestra en la Tabla 1.8, se distinguen tres ejes temáticos: Sociedad en perspectiva histórica, Espacio geográfico y Democracia y desarrollo.

El eje Sociedad en perspectiva histórica incluye contenidos referidos a la historia mundial del siglo XX y a la historia de Chile, desde el periodo colonial hasta fines del siglo XIX. La aproximación a los diversos procesos históricos se orienta hacia la distinción de características propias de las distintas etapas y procesos y hacia su comprensión desde diversas dimensiones e interpretaciones, propiciando una visión de conjunto (no fragmentada) del desarrollo histórico. En particular, los conocimientos sobre Historia Mundial del siglo XX incluyen las guerras mundiales, el periodo entre guerras y el impacto de las guerras en el ordenamiento mundial y en la creciente valoración de la paz y los derechos humanos. En relación con la historia de Chile, se incluye una visión panorámica, distinguiendo elementos de continuidad y cambio y abordando temas como la permanencia de las culturas indígenas, el proceso de mestizaje, el legado colonial y la conformación del estado-nación y su evolución política, social, económica, cultural y territorial a lo largo del siglo XIX.

El eje Espacio geográfico abarca los contenidos referidos a los procesos de urbanización del siglo XX, considerando migración campo-ciudad (efecto en la composición urbano-rural en distintas regiones a nivel mundial), ventajas y desventajas de la vida en las ciudades, y geografía de la población mundial (diversidad cultural, dinámica y estructura, y problemas e impacto del desarrollo tecnológico, entre otros).

El eje Democracia y desarrollo incluye contenidos referidos al proceso de globalización e internacionalización de la economía en el mundo actual, considerando sus principales características, sus falencias (como las desigualdades o contrastes que este proceso genera), sus logros y su impacto en la valoración de los derechos humanos. También aborda las características de la llamada sociedad de la información, esto es: interconectividad global; nuevas tecnologías de información y comunicación); problemas y desafíos provocados por la tensión entre la difusión de patrones culturales a nivel planetario, y permanencia de las identidades locales.

Tabla 1.8

Matriz de evaluación teórica de la Prueba Simce Historia, Geografía y Ciencias Sociales 2015. Nivel II medio. Ejes temáticos

Ejes temáticos	II medio
Sociedad en perspectiva histórica	70%
Espacio geográfico	15%
Democracia y desarrollo	15%

1.1.5 Prueba de Ciencias Naturales

En la prueba de Ciencias Naturales aplicada en 2015 en 8° básico, se distinguen tres ejes de habilidades (Conocimiento, Aplicación y Razonamiento) cuya matriz de evaluación teórica se muestra en la Tabla 1.9.

El eje Conocimiento incluye habilidades referidas a la comprensión de hechos, relaciones, conceptos y procesos propios de las Ciencias Naturales, tales como recordar, reconocer y ejemplificar afirmaciones, hechos y conceptos, y dominar vocabulario científico (símbolos y unidades).

El eje Aplicación supone la aplicación del conocimiento científico en situaciones sencillas, lo que implica habilidades tales como comparar, contrastar, clasificar e interpretar información relacionada con algún concepto o principio de la ciencia, así como organizar información para comunicarla, y explicar y relacionar información obtenida de diagramas o modelos.

El eje Razonamiento refiere al uso del razonamiento científico para resolver problemas, extraer conclusiones, tomar decisiones fundamentadas y abordar situaciones nuevas.

Tabla 1.9

Matriz de evaluación teórica de la prueba Simce Ciencias Naturales 2015. Nivel 8° básico. Ejes de habilidades

Ejes de habilidades	8° básico
Conocimiento	42%
Aplicación	40%
Razonamiento	18%

Respecto de los ejes temáticos, cuya matriz de evaluación teórica se muestra en la Tabla 1.10, se distinguen cinco ejes: Estructura y función de los seres vivos; Organismos, ambiente y sus interacciones; Materia y sus transformaciones; Fuerza y movimiento; y Tierra y Universo.

El eje Estructura y función de los seres vivos incluye estructura y función celular (con foco en su rol portador del material genético); relación de la célula con los sistemas; función integrada de los sistemas circulatorio, respiratorio, digestivo y excretor; relaciones entre requerimientos nutricionales de las personas y aspectos fisiológicos tales como la edad, el sexo y la actividad física; aspectos biológicos, psicológicos, sociales y de salud involucrados en la sexualidad humana (incluyendo reproducción, control de la natalidad y prevención de enfermedades de transmisión sexual), y efectos, consecuencias y medidas de prevención del consumo de drogas.

El eje Organismos, ambiente y sus interacciones involucra las diversas interacciones que se producen entre los organismos de un ecosistema; los procesos básicos de los ciclos del carbono y el nitrógeno; la función que cumplen los organismos productores y descomponedores (incluyendo los efectos de la intervención humana en estos procesos); las principales teorías acerca del origen de la vida, y su impacto en la sociedad, y conceptos y principios básicos de la evolución.

El eje Materia y sus transformaciones incluye las transformaciones físico-químicas de la materia; la formación de moléculas y macromoléculas; fenómenos como la electrización, la conductividad

eléctrica y calórica y la emisión y absorción de luz; la relevancia de los elementos químicos como constituyentes de los seres vivos y su importancia industrial; las reacciones químicas comunes y la formación de compuestos; las principales teorías y modelos atómicos (Dalton, Thompson, Rutherford y Bohr), y las leyes que explican el comportamiento de los gases ideales para describir fenómenos atmosféricos y cotidianos.

El eje Fuerza y movimiento incorpora conocimientos sobre las fuerzas que actúan simultáneamente sobre un cuerpo y la dirección en que lo hacen, aplicados en situaciones concretas. Asimismo, incluye los efectos que producen las fuerzas gravitacionales sobre los cuerpos que se encuentran en las cercanías de la superficie de la Tierra; las nociones cuantitativas de período, amplitud y frecuencia, para describir algunos movimientos periódicos, y los fenómenos como la electrización y el movimiento de cargas, incluyendo las fuerzas eléctricas y el rol que ellas juegan en la estructura atómica y molecular.

El eje Tierra y universo abarca el conocimiento sobre las características de algunas estructuras cósmicas (cometas, asteroides, meteoritos, nebulosas, galaxias y cúmulos de galaxias), y sobre su ubicación en la Vía Láctea y el Sistema Solar, empleando unidades de tiempo-luz para analizar las distancias entre cuerpos celestes. Asimismo, incluye las estructuras y la dinámica del planeta Tierra; los principales tipos de rocas y su ciclo de formación; evidencias sobre las transformaciones que han experimentado la atmósfera, la litósfera y la hidrósfera a través del tiempo geológico, y conocimientos elementales sobre los fenómenos naturales que se producen en estas capas (temporales, mareas, sismos y erupciones volcánicas) y su impacto sobre la vida.

Tabla 1.10

Matriz de evaluación teórica de la prueba Simce Ciencias Naturales 2015. Nivel 8° básico. Ejes temáticos

Ejes temáticos	8° básico
Estructura y función de los seres vivos	24%
Organismos, ambiente y sus interacciones	10%
Materia y sus transformaciones	36%
Fuerza y movimiento	20%
Tierra y universo	10%

1.2 Elaboración y validación de las preguntas

Las preguntas de las pruebas Simce se construyen a partir de Objetivos de Evaluación (OE) que corresponden a una operacionalización de los ejes descritos en el apartado anterior.

En las pruebas Simce se incluyen dos formatos de preguntas: de opción múltiple ("preguntas cerradas") y de respuesta abierta ("preguntas abiertas")[2]. La elección de estos formatos está determinada por el contenido y el comportamiento cognitivo que se busca evaluar (Haladyna y Rodríguez, 2013). Por ello, las pruebas Simce se construyen mayoritariamente sobre la base de preguntas de opción múltiple, pues este formato permite reportar información de la mayoría de los constructos a evaluar en forma efectiva y eficiente, asegurando validez, confiabilidad y objetividad del instrumento en su totalidad (Downing y Haladyna, 2006). Las preguntas de respuesta abierta se utilizan fundamentalmente para

[2] Solo las pruebas de 2° básico contienen otros dos tipos de formatos de preguntas: de selección de respuesta, donde los estudiantes deben encerrar en un círculo la o las respuestas correctas; y de ordenación, donde los estudiantes deben enumerar acciones o acontecimientos.

evaluar aquellos aprendizajes que no son susceptibles de ser medidos con preguntas de respuesta cerrada y que, en términos generales, corresponden a habilidades cognitivas de orden superior, para las cuales el formato abierto permite mayor fidelidad al constructo medido (Haladyna y Rodríguez, 2013).

Las preguntas cerradas se construyen a partir de un enunciado para el cual existe una única opción correcta (clave) y dos o tres distractores (opciones incorrectas), dependiendo de la variedad de errores plausibles que sea posible recoger. La única excepción la constituyen las preguntas de las pruebas de 2° básico, que se construyen únicamente con tres opciones de respuesta (una clave y dos distractores), con el objeto de disminuir la carga verbal para estos estudiantes que se encuentran en etapas iniciales de lectura. Los distractores se elaboran sobre la base de errores comunes que cometen los estudiantes que no han logrado los aprendizajes (Haladyna y Rodríguez, 2013). En algunos casos, las preguntas se encuentran contextualizadas con información que sirve de base para la selección de la respuesta, tales como textos, gráficos, situaciones problemáticas, etc.

Las preguntas de respuesta abierta, por su parte, se construyen a partir de enunciados ("estímulos") que plantean una tarea o interrogante que debe ser desarrollada por el estudiante. Al igual que en las preguntas cerradas, algunas preguntas abiertas también se encuentran contextualizadas con información de base para el desarrollo de la tarea. Las pautas de este formato de preguntas son de dos tipos: algunas corresponden a "pautas de respuesta única", es decir, documentos que establecen la o las respuestas correctas posibles, y otras corresponden a rúbricas, esto es, documentos que establecen descripciones de desempeño, dispuestas en una escala, las que se elaboran sobre la base de los aspectos que se observarán en la ejecución de una tarea y/o producto[3].

El proceso de elaboración de las preguntas de las pruebas Simce 2015 se llevó a cabo mediante la modalidad de contratación directa de elaboradores externos por área y nivel escolar evaluado. Se conformaron tres grupos de elaboradores, en tres regiones del país: Coquimbo (La Serena), del Biobío (Concepción) y Metropolitana (Santiago)[4]. El periodo de elaboración se extendió por ocho meses, desde febrero hasta septiembre de 2014, en función de un calendario de producción y revisión semanal de preguntas. Una vez finalizada la primera mitad del periodo de elaboración, se realizaron paneles de revisión de expertos que determinaron, para cada uno de los ítems revisados, si estos debían ser aprobados, rechazados o modificados, y realizaron comentarios generales que permitieron nutrir los procesos de elaboración. Al finalizar el proceso de elaboración externa, se realizaron talleres de elaboración interna, conformados por los profesionales de los equipos de construcción de la Agencia, que tuvieron como propósito elaborar preguntas para cubrir aquellos contenidos y habilidades que no quedaron representados en el set de preguntas elaboradas por externos.

Una vez elaborados los conjuntos de preguntas para cada una de las pruebas, los ítems fueron distribuidos en cuadernillos para su aplicación experimental conducente a su validación.

[3] En el caso específico de la prueba de Escritura se utilizan dos tipos de rúbricas: holística y analítica. La rúbrica holística permite evaluar los textos desde una perspectiva global, estableciendo niveles de desarrollo de la respuesta desde un enfoque comunicativo. La rúbrica analítica permite medir aspectos específicos del texto de manera independiente. La rúbrica holística se utiliza para puntuar las respuestas de los estudiantes y establecer el puntaje promedio, dando cuenta de su competencia de expresión escrita, mientras que la rúbrica analítica complementa la información, entregándola desagregada en habilidades de escritura que conforman dicha competencia.

[4] La selección de los elaboradores se realizó mediante concurso público, contemplando análisis de antecedentes profesionales y aplicación de dos pruebas técnicas (una individual y otra grupal). Una vez seleccionados, los elaboradores fueron capacitados por miembros de los equipos disciplinarios de la Agencia. En esta capacitación se abordan los criterios de calidad de los ítems y las especificaciones técnicas, y se realizan talleres de elaboración con retroalimentación de los productos.

Las pruebas experimentales se aplican el año anterior a la aplicación censal definitiva, a una muestra representativa de la población evaluable. El análisis de estas pruebas permite obtener información cuantitativa y cualitativa de los ítems para determinar si poseen la suficiente calidad como para conformar una prueba Simce definitiva. En el caso de las preguntas abiertas, la información permite validar las pautas y rúbricas, y elaborar los manuales de corrección. En el caso de las preguntas de opción múltiple, el proceso permite obtener información sobre la calidad de los enunciados y las opciones (véase el detalle de las aplicaciones experimentales que sirvieron de insumo a las pruebas censales 2015, en el anexo A).

El resultado de los análisis cuantitativos y cualitativos da origen a un conjunto de preguntas que, sumadas a otras provenientes de aplicaciones previas, conformarán las pruebas definitivas. Todas estas preguntas deben cumplir con los requisitos de representatividad y cobertura curricular, cobertura del rango de habilidades de los estudiantes y alineación a los Estándares de Aprendizaje[5], en los niveles en que existan estándares.

1.3 Armado y diagramación de las pruebas

El diseño de las pruebas Simce censales busca minimizar la posibilidad de que los resultados se vean afectados por factores externos al constructo. Cada prueba está compuesta por varias formas (cuadernillos), las que en su conjunto cubren todos los ejes temáticos y de habilidades de las especificaciones técnicas[6], los que a su vez representan una amplia gama de los Objetivos de Aprendizaje del Currículo Nacional.

En el ensamblaje de las pruebas censales, junto con procurar el cumplimiento de los porcentajes de las matrices de las especificaciones técnicas, se incluyen preguntas que miden los diversos objetivos de evaluación para cada uno de los ejes; sin embargo, uno de los elementos centrales del armado consiste en establecer equiparaciones, tanto entre las formas de una misma prueba, como entre pruebas aplicadas en distintos años, pues de lo contrario no sería posible comparar los resultados.

En virtud de lo anterior, la primera acción del armado consiste en seleccionar el conjunto de ítems de equiparación entre años (*equating* entre años) para cada prueba. Estos conjuntos conforman subpruebas de las mediciones de origen, esto es, una representación del total de ítems de cada evaluación, en términos de constructo y características psicométricas. Estos ítems de equating corresponden a cerca de 20% de las preguntas de la prueba de la medición anterior.

En todas las pruebas Simce 2015, la medición anterior correspondió a las pruebas aplicadas en 2014, a excepción de la prueba de Ciencias Naturales de 8° básico, cuya base de comparación fue aplicada en 2013, y de las pruebas de Historia, Geografía y Ciencias Sociales, de 6° básico y II medio, pues estas

[5] Los Estándares de Aprendizaje están contenidos en los decretos supremos de educación del Ministerio de Educación de Chile, N.°.129/2013, N.°.81/2014 y N.°.178/2015 y se encuentran disponibles en: http://www.curriculumenlineamineduc.cl/605/w3-article-33859.html

[6] Para los cursos con menos de diez estudiantes por nivel escolar se elabora una prueba extendida, que contempla todas las formas regulares, siempre y cuando, en su conjunto, no excedan una determinada cantidad de ítems. Si se excede la cantidad límite de ítems, se analizan los datos psicométricos y se decide qué ítems se pueden excluir de la prueba extendida, sin que se altere la función de información de la prueba regular. Las pruebas Simce 2015 que requirieron contar con una versión extendida fueron las de 2°, 4° y 6° básico y el límite de ítems máximo se estableció en 50 ítems en 2° básico y en 80 ítems en 4° y 6° básico.

se aplicaron por primera vez en 2015 y por tanto constituyen línea base para las futuras mediciones. También constituyó línea base la prueba de Escritura 2015, pues pese a que hubo aplicaciones previas en 2013 y 2014, estas experimentaron diversos cambios requeridos para la mejora del constructo.

La selección de los ítems de *equating* entre años (que corresponden únicamente a preguntas de opción múltiple), se realiza en función de una serie de criterios que aseguren su calidad, tales como que cuenten con parámetros IRT adecuados, que no presenten problemas de construcción, que no estén publicados (ni asociados a un estímulo publicado) y que, en su conjunto abarquen el rango de dificultad de la prueba y sean representativos de la matriz teórica de la prueba respectiva del año y de la prueba definitiva del ciclo anterior.

Una vez validados los ítems de *equating*, la etapa siguiente consiste en posicionarlos. Para ello se consideran dos criterios: (a) mantener el mismo ordenamiento relativo y (b) mantener una posición relativa similar a la que tuvieron los ítems en la prueba definitiva previa, en las distintas formas, asegurando en este último caso, que no exista una distancia superior a ±5 posiciones respecto de la ubicación original en las distintas formas[7].

La tercera etapa del armado consiste en posicionar los ítems de equiparación entre formas, en función de los mismos criterios anteriores, es decir, asegurando que no exista inversión de posiciones entre las distintas formas y que no exista una distancia mayor a ±5 posiciones en la ubicación del ítem en las distintas formas de la prueba. Los ítems que permiten equiparar las formas representan aproximadamente 20% del total de preguntas de cada forma.

Una vez posicionadas las preguntas de equiparación entre formas y años, se distribuyen y posicionan los restantes ítems, atendiendo a los siguientes criterios[8]:

- El posicionamiento general debe atender a la lógica interna de cada disciplina. Así, por ejemplo, en el caso de las pruebas de Lectura, el ordenamiento está supeditado a los textos para la lectura, y en el caso de las pruebas de Ciencias Naturales y de Historia, Geografía y Ciencias Sociales, la secuenciación debe realizarse en función del contenido, según recomienda el informe ACER (2013).
- Cada forma debe tener similar carga cognitiva, en términos de presencia de estímulos y figuras que deban procesarse de manera simultánea.
- Los primeros ítems de cada forma deben ser sencillos y motivadores, y los de mayor dificultad o más complejos deben estar en posiciones intermedias.
- Los ítems de respuesta abierta no deben ubicarse en las posiciones finales.
- Las claves (letras de opciones correctas) deben estar distribuidas de manera homogénea al interior de cada prueba y se debe asegurar que no existan patrones.
- Si hubiera ítems dependientes o encadenados (esto es, que contengan información útil para responder correctamente otras preguntas), se deben ubicar en formas distintas, para garantizar la independencia local de las preguntas.

[7] En el caso de los ítems asociados a un estímulo común (testles), la distancia puede tener un leve desplazamiento adicional a las ±5 posiciones, dado que deben ser ubicados en torno al estímulo que les sirve de base.

[8] El ensamblaje de la prueba de Escritura se realiza con un procedimiento distinto al del resto de las pruebas. La prueba de Escritura 2015 está conformada por 8 formas de prueba de dos ítems cada una (uno de producción de textos narrativos y uno de textos informativos). En total la prueba consta de 8 preguntas (4 de cada tipo) que se distribuyen en 4 bloques de producción escrita, con el objetivo de rotar la posición de los ítems según el tipo de texto evaluado.

Una vez finalizado el posicionamiento de las preguntas en cada forma, se evalúa que las formas exhiban una función de información y una curva característica, similares entre sí y, en caso de detectar falencias en este ámbito, se realizan las modificaciones pertinentes.

La última acción de la construcción de los instrumentos, consiste en la diagramación de los cuadernillos. En efecto, las preguntas de las pruebas Simce se presentan en cuadernillos impresos, cuyas portadas contienen las instrucciones generales de la aplicación. Adicionalmente, se diseñan y elaboran cuadernillos de respuesta donde los estudiantes deben marcar la opción que consideren correcta (salvo en las pruebas de 2° básico, que se responden en el mismo cuadernillo de prueba, y en la prueba de Escritura, que no posee cuadernillo de respuestas[9]). Las preguntas abiertas se responden en el mismo cuadernillo de prueba, en 4° y 6° básico, y en cuadernillos de respuesta, en 8° básico y II medio.

La diagramación de las pruebas Simce se realiza en función de los criterios contenidos en un Manual de Diagramación, y las pruebas diagramadas son sometidas a múltiples revisiones al interior de la Agencia (desde las jefaturas de área hasta el Secretario Ejecutivo) y por parte de profesionales de la UCE del Ministerio de Educación. Los cuadernillos aprobados tras estas revisiones, son derivados a imprenta.

1.4 Prueba muestral de Educación Física

Como se mencionó al comienzo de este capítulo, junto con las pruebas Simce censales, el sistema de evaluación contempla pruebas muestrales que se aplican a una parte de la población objetivo previamente determinada. En el año 2015 se aplicó solo una prueba muestral: el Estudio Nacional de Educación Física.

En el ámbito educativo, la condición física se vincula fundamentalmente con la salud, y considera aquellos componentes que se ven afectados favorable o negativamente por el nivel habitual de actividad física (Lamela, 2009 y Nogueira, 2002). De manera específica, en el Currículo Nacional vigente al momento de la aplicación (Mineduc, 2002) se promueve el desarrollo de programas de ejercicios físicos, señalando entre sus Objetivos Fundamentales y Contenidos Mínimos Obligatorios, los siguientes: "realizar programas de ejercicios y actividades físicas, reconociendo los progresos personales en aquellas cualidades físicas asociadas a la salud y la calidad de vida" y "ejercitación de planes de trabajo físico individuales o colectivos para el progreso de la aptitud física: evaluación y comparación de los progresos personales obtenidos por los alumnos".

Considerando estos antecedentes, la prueba muestral de Educación Física 2015 evaluó la condición física de los estudiantes de 8° básico, a partir de la medición de los siguientes componentes:

Antropometría: referida a los aspectos relacionados con las dimensiones corporales de una persona. Las pruebas aplicadas para evaluar este componente permiten determinar el Índice de Masa Corporal (IMC) y la acumulación de grasa en la zona central del cuerpo.

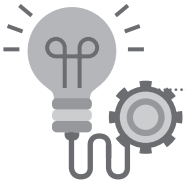
[9] En la prueba de Escritura el cuadernillo contiene una hoja de planificación del texto y una hoja de desarrollo del escrito.

Resistencia aeróbica y rendimiento cardiovascular: que se relaciona con la capacidad biológica que permite mantener un esfuerzo a intensidad media o baja con suficiente aporte de oxígeno. Para estimar el rendimiento cardiovascular, se usan como parámetros la frecuencia cardíaca y el consumo de oxígeno.

Rendimiento muscular: relacionado con la capacidad de trabajo de los músculos, incluye fuerza y resistencia muscular. La fuerza muscular, entendida como la capacidad de un músculo o grupo de músculos de generar tensión; y la resistencia muscular, como la capacidad de los músculos de aplicar una fuerza submáxima (expresión de fuerza que no alcanza su máxima capacidad) de forma repetida, o bien de mantener una contracción muscular durante un periodo de tiempo prolongado.

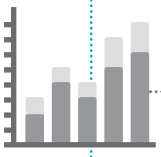
Flexibilidad: referida a la capacidad funcional de las articulaciones de moverse en todo su rango o bajo la influencia de fuerzas externas, sin dañar los músculos ni las articulaciones. Incluye la movilidad articular y la elasticidad muscular.

Potencia aeróbica máxima: entendida como la capacidad máxima de consumo de oxígeno del organismo, durante un esfuerzo físico máximo.



Capítulo I:
Diseño y construcción de las pruebas Simce

Capítulo II:
Logística y aplicación de las pruebas Simce



Capítulo III:
Procesamiento y análisis de datos

Capítulo IV:
Comunicación de resultados educativos

Capítulo II: Logística y aplicación de las pruebas Simce

La aplicación Simce involucra procesos operativos y logísticos que son tan relevantes para la calidad de los resultados obtenidos, como la fase de diseño y construcción descrita en el capítulo anterior.

Estos procesos de aplicación de los instrumentos (pruebas y cuestionarios[10]) se llevan a cabo en función de instrucciones, manuales y protocolos de actuación que se basan en los criterios de estandarización internacional definidos en el documento Standards for Educational and Psychological Testing (AERA, APA y NCME, 2014). Tales documentos tienen por objeto ofrecer condiciones equivalentes a todos los estudiantes al momento de rendir la prueba, a la vez que permiten asegurar confidencialidad de la información y resguardo de los materiales aplicados.

Considerando la extensión del territorio nacional y la cantidad de pruebas que se aplica cada año, la Agencia ha externalizado algunos servicios asociados a las operaciones de campo a través de licitaciones públicas, para optimizar la eficiencia y calidad de los procesos. Sin embargo, es un equipo perteneciente a la Agencia el que tiene la responsabilidad de velar porque estos servicios externos cumplan estrictamente los criterios de calidad que exige el proceso, para lo cual desarrolla un plan de supervisión.

En el año 2015, con el propósito de asegurar la calidad de los procesos logísticos y de aplicación, la Agencia puso a disposición de la comunidad escolar, por primera vez, una mesa de ayuda telefónica especial, para responder consultas, reclamos y/o sugerencias relacionadas con la aplicación. Adicionalmente, desarrolló un plan de supervisión para el cual se contrató a 16 supervisores, uno por región (dos en la Región Metropolitana). Finalmente, para reforzar la logística de los procesos, la Agencia envió profesionales de apoyo a la supervisión (al menos uno por región), para atender las contingencias de alto impacto y para proporcionar apoyo y orientación a las autoridades regionales, tanto a nivel de Departamento Provincial (DEPROV) como de Secretaría Regional Ministerial (SEREMI).

Todos estos apoyos fueron fundamentales en la aplicación de ese año, ya que se registraron al menos tres situaciones de alta complejidad que debieron ser resueltas desde la Agencia:

- Las movilizaciones estudiantiles y huelgas de funcionarios que afectaron a una importante cantidad de establecimientos a lo largo del país.
- El sismo que afectó a la Región de Coquimbo, alterando el normal funcionamiento de muchos de los establecimientos de la zona.
- Los fuertes temporales del centro-sur del país, que bloquearon muchas de las rutas de acceso a los establecimientos de zonas rurales de las regiones del Libertador Bernardo O´Higgins y del Maule, especialmente en la alta cordillera y la costa.

[10] Los cuestionarios son instrumentos que permiten obtener información sobre Indicadores de Desarrollo Personal y Social de los estudiantes y se aplican a los estudiantes y a docentes, directivos y padres y apoderados de los estudiantes evaluados.

Las movilizaciones estudiantiles y huelgas de funcionarios afectaron especialmente la aplicación de las pruebas en los cursos de 8° básico y II medio; el sismo de la Región de Coquimbo (ocurrido en septiembre de 2015, un mes antes de la aplicación) afectó a los establecimientos de la Provincia del Choapa, y los cortes de caminos causados por los temporales de la zona centro-sur alteraron específicamente la aplicación de las pruebas de 6° básico.

El plan de contingencia adoptado por la Agencia, en todos los casos antes descritos, consistió en instruir a las empresas aplicadoras para que efectuaran gestiones con los directores de los establecimientos afectados y lograran coordinar sesiones de aplicación especiales en fechas posteriores a las estipuladas en el calendario oficial. Esta solución permitió disminuir el número de cursos no evaluados, aunque en el caso del sismo, pese a las gestiones, no en todos los establecimientos fue posible conseguir que las instalaciones estuvieran en condiciones idóneas para realizar la aplicación en una fecha posterior.

Los procesos involucrados en la logística y aplicación de las pruebas, que se describen en este capítulo, son los siguientes:

- Empadronamiento de los establecimientos y los estudiantes
- Impresión, mecanizado y distribución del material de prueba
- Aplicación de pruebas Simce
- Entrega del material aplicado a los centros de captura

2.1 Empadronamiento de establecimientos y estudiantes

El empadronamiento corresponde al proceso de construcción de una base de datos consolidada de los estudiantes que rendirán las pruebas, y de los establecimientos y cursos a los que asisten. En 2015, el proceso de empadronamiento se llevó a cabo entre los meses de junio y julio de ese año.

Esta base de datos es construida y luego validada mediante un procedimiento que contempla al menos un contacto telefónico con cada establecimiento, a través del director, el subdirector o el jefe de la Unidad Técnica Pedagógica (UTP), para confirmar la información obtenida en primera instancia, esto es, la cantidad de matrícula por curso y los datos de los establecimientos (RBD o "rol base de datos", dirección, teléfono y nombre del director). Toda esta información resulta clave para la coordinación y ejecución de los procesos involucrados en la aplicación.

La construcción de la base de datos de pruebas censales se realiza en tres niveles: establecimientos, cursos y estudiantes.

Respecto de los establecimientos, la Ley General de Educación (Mineduc, 2009a) establece que las pruebas Simce deben aplicarse en "todos los establecimientos educacionales de enseñanza regular del país" que tengan al menos un estudiante matriculado en el nivel escolar en que se aplica una determinada prueba, por tanto, solo quedan excluidas del proceso las escuelas de enseñanza no regular: escuelas carcelarias, hospitalarias, nocturnas y de educación de adultos.

Para la definición de estos establecimientos se utilizan los datos que entrega el Sistema de Información General de Estudiantes (SIGE)[11] del Ministerio de Educación, los que se encuentran sistematizados con códigos según el tipo de enseñanza impartida. En la Tabla 2.1 se muestran las categorías de los establecimientos que fueron considerados para la aplicación de las pruebas Simce 2015.

Tabla 2.1

Códigos SIGE de establecimientos según tipo de enseñanza impartida

Código	Categoría
110	Educación básica
310	Educación media humanista-científica
410	Educación media técnico-profesional comercial
510	Educación media técnico-profesional industrial
610	Educación media técnico-profesional técnica
710	Educación media técnico-profesional agrícola
810	Educación media técnico-profesional marítima
910	Educación media artística

Respecto de los estudiantes, la población objetivo la constituyen todos los alumnos de los establecimientos antes señalados, a excepción de: (a) aquellos que se encuentren ausentes en las fechas de aplicación de las pruebas; (b) los estudiantes que presenten algún problema físico temporal que dificulte o haga imposible la ejecución de las pruebas; y/o (c) los alumnos que no hablen español o que, no siendo hablantes nativos de español, hayan recibido menos de un año de enseñanza en el sistema.

Todos los demás estudiantes deben rendir las pruebas, incluso aquellos que presentan Necesidades Educativas Especiales Permanentes (NEEP), aun cuando sus resultados no se consideren al momento de informar los puntajes. La única excepción entre los estudiantes con NEEP, son aquellos para los que existen pruebas Simce de Discapacidad Sensorial (pruebas Simce DS). En efecto, la Agencia, a instancias de lo establecido en la Ley General de Educación (Mineduc, 2009a), ha avanzado en realizar adecuaciones en sus instrumentos para que los estudiantes con NEEP puedan participar del proceso de evaluación en igualdad de condiciones. En concreto, desde el año 2012, se aplican pruebas Simce DS de Lectura y Matemática, que permiten evaluar a los estudiantes con discapacidad sensorial[12] que se encuentran matriculados en los establecimientos de enseñanza regular. Las pruebas Simce DS 2015 se aplicaron solo en 6º básico, en los establecimientos identificados según códigos SIGE N.º 211 (Educación especial discapacidad auditiva) y N.º 213 (Educación especial discapacidad visual), además de los establecimientos de educación básica (código N.º 110) a los que asiste parte de esta población de estudiantes. Para la identificación de los estudiantes que en 2015 rindieron la prueba Simce DS, la Agencia desarrolló una plataforma web para que cada director registrara (acreditando con los certificados correspondientes) a los estudiantes con discapacidad sensorial, y luego esta información fue validada con el registro de estudiantes de los Proyectos de Integración Escolar (PIE) que posee el Mineduc.

[11] El SIGE obtiene la información del reporte de matrícula informado y actualizado mensualmente por los directores de los establecimientos educacionales del país.

[12] Los estudiantes con discapacidad sensorial corresponden a aquellos que presentan discapacidad visual total o parcial y/o discapacidad auditiva, de acuerdo a lo establecido en el Decreto Supremo N.º 170 del Ministerio de Educación (Mineduc, 2009b).

Respecto de las **pruebas muestrales**, el proceso de empadronamiento supone configurar primeramente una muestra representativa de los estudiantes de los establecimientos de todo el país que cursen el nivel escolar evaluado. En 2015, como se señaló en el capítulo anterior, correspondió a los estudiantes de 8º básico, a quienes se les aplicó la prueba muestral de Educación Física.

La única excepción en la conformación de la muestra 2015 fueron aquellos establecimientos que presentaron uno o más de los siguientes criterios de exclusión, dadas las dificultades logísticas asociadas:

- Establecimientos ubicados en las localidades de Juan Fernández, Isla de Pascua y la Antártica.
- Establecimientos con menos de diez estudiantes en 8º básico.
- Establecimientos con más de cuatro cursos.

Adicionalmente, también se excluyeron los establecimientos que habían tenido accidentes en aplicaciones anteriores (por resguardo de la seguridad de los estudiantes) y aquellos que habían sido evaluados el año anterior (pues no se repiten los establecimientos para evitar una posible sobrecarga de evaluaciones).

Ahora bien, para la selección de la muestra de establecimientos, se realiza un muestreo estratificado por región y dependencia administrativa del establecimiento (municipal/ particular subvencionado/ particular pagado[13]). La asignación del tamaño de la muestra en 2015 se realizó en dos fases: en la primera se usó una afijación simple por región y en la segunda, una asignación proporcional según dependencia administrativa dentro de cada región. Esto implicó un cambio respecto del muestreo realizado el año previo (2014), que se consideró necesario incorporar con el fin de reportar resultados sobre toda la población objetivo, utilizando expansores de la muestra. El factor de expansión que se utilizó fue el inverso de la probabilidad de selección de un estudiante en cada estrato.

Como resultado del empadronamiento 2015, para la aplicación de las pruebas censales de Lectura, Matemática, Historia, Geografía y Ciencias Sociales y Ciencias Naturales, se estableció una población total de 1.183.980 estudiantes, distribuidos en 8.902 establecimientos. En el caso de la prueba de Escritura, se realizó un empadronamiento de 7.511 establecimientos. Finalmente, para la aplicación muestral de Educación Física, el tamaño de la muestra se definió en 374 establecimientos, reuniendo un total de 15.696 estudiantes de 8º básico, distribuidos equitativamente en cada una de las 15 regiones del país, y con representación aproximadamente equivalente por género (48% mujeres y 52% hombres).

2.2 Impresión, mecanizado y distribución del material de prueba

La impresión, mecanizado (empaque de los documentos correspondientes a un curso en cajas o sobres) y la distribución del material, son fases que, al igual que las anteriores, resultan clave para la obtención de resultados válidos de las evaluaciones. Un error en la impresión o mecanizado puede afectar gravemente el registro y la posterior captura de la información. Del mismo modo, una falla de impresión (ambigüedad en un color, pérdida de legibilidad de un texto o imagen, error en la compaginación, etc.), podría terminar invalidando las interpretaciones sobre los resultados de los estudiantes (Martin, Rust

[13] En 2015 el estrato particular pagado estuvo desierto en la Región de Aysén.

y Adams, eds., 1999). Igualmente, si se produce una mecanización con errores podría verse afectada la aplicación de uno o varios cursos, e incluso de un establecimiento completo, sin que fuera posible obtener resultados de esos estudiantes. Por todas estas razones, los procesos de impresión, mecanizado y distribución se realizan bajo una estricta supervisión de los proveedores de estos servicios.

En la aplicación 2015, la impresión, mecanizado y distribución del material fueron adjudicados mediante licitación pública[14] a dos proveedores del área de impresión, y el servicio se ejecutó entre julio y octubre de ese año.

El proceso de **impresión** incluye la impresión de los cuadernillos de prueba y cuadernillos de respuesta, de los cuestionarios, de los materiales de capacitación (pruebas de selección de examinadores y demás materiales requeridos para el proceso de selección y capacitación de los examinadores) y del Manual de Aplicación, que se describen más adelante.

El **mecanizado** corresponde al proceso de ordenamiento y embalaje del material de aplicación en cajas y sobres rotulados para su distribución posterior en los establecimientos y respectivos cursos. Cada sobre es identificado con una etiqueta con un código de barras que contiene los datos del establecimiento, el nombre y dirección del destinatario (centro de operaciones) y, en el caso de las cajas contenedoras de los sobres, el número de unidades contenidas por tipo de material.

El mecanizado se realiza atendiendo a dos momentos de la aplicación: la visita previa al establecimiento y el día de la aplicación. Para la visita previa al establecimiento se realiza mecanizado del material administrativo utilizado como conductor del proceso de aplicación de las pruebas Simce, y para el día de la aplicación, se mecaniza lo que se conoce como caja-curso, la que contiene todos los materiales requeridos para la aplicación de ambos días (incluyendo material complementario): pruebas, cuestionarios, lápices, gomas de borrar, sacapuntas y sello de seguridad plástico para el cierre de la caja en los centros de operaciones de los aplicadores. Adicionalmente, en los cursos donde es requerido se incluyen las herramientas de apoyo para la aplicación de las pruebas Simce DS (atril, regleta y punzón). Junto con las cajas-curso, se mecanizan las cajas contenedoras en las que se deben agrupar las cajas-curso de cada día de aplicación (que identifican el tipo de evaluaciones que contienen: censal regular, extendida, DS o experimental) y las cajas de contingencia, que corresponden a cajas de idéntico contenido al de las cajas-curso, pero que no se encuentran asociadas a un establecimiento ni curso particular, por lo que sirven para reemplazo de cajas-curso en caso de que alguna de estas presente fallas de impresión de material, surja un curso no empadronado, o cualquier otra situación emergente que requiera de su uso.

La **distribución** consiste en el traslado del material mecanizado, para el cual la Agencia instruye un esquema de tránsito del material que pone énfasis en el uso de códigos para identificar la trazabilidad durante todo el proceso de aplicación.

En una primera instancia, el material mecanizado se distribuye a los puntos de recepción, informados previamente por cada agente aplicador que, en 2015 fue uno por región[15]. En este proceso, el material de capacitación se traslada a los centros de capacitación de examinadores y supervisores, y el material complementario y las cajas-curso, a los centros de acopio de los aplicadores. Los centros de acopio

[14] Licitación "Servicio de impresión, mecanizado y distribución", ID 721703-2-LP15.

[15] En 2015 hubo una disminución de los puntos de entrega, en relación con los definidos en la aplicación anterior (2014), lo que permitió mantener un mejor control de la trazabilidad del material de aplicación, asegurando con esto una oportuna distribución a los centros de acopio.

corresponden a bodegas y casas habilitadas exclusivamente para este servicio. El servicio supone la recepción, revisión, almacenamiento, manipulación, custodia y despacho del material de aplicación de todos los centros de operaciones asociados al mismo. Para la aplicación 2015 se establecieron 16 centros de acopio distribuidos entre las regiones de Arica y Parinacota y Magallanes, a los cuales se despacharon, desde las dependencias de la imprenta en la Región Metropolitana, un promedio aproximado de 10.500 sobres complementarios y 11.500 cajas-curso por nivel escolar evaluado.

Para asegurar una adecuada trazabilidad durante el proceso de aplicación, la Agencia solicita a los aplicadores el desarrollo de una plataforma de visualización de la información asociada a cada actividad. Este sistema permite realizar el seguimiento del estado de avance de las actividades en cada etapa de la aplicación y recibir oportunamente la información relativa al cumplimiento de hitos específicos, como la definición de los centros de capacitación y de operaciones y el listado de examinadores seleccionados y sus resultados. En 2015, la Agencia solicitó que toda la información de avance de los procesos fuera actualizada con un desfase máximo de 24 horas. En el caso del material complementario y de aplicación, la actualización de los estados de avance debía tener un desfase máximo de 6 horas.

La plataforma permitió el monitoreo en línea del estado de avance de, al menos, las siguientes etapas:

- Convocatoria y formación de examinadores y supervisores.
- Evaluación y selección de examinadores y supervisores (personal aprobado y rechazado y puntajes obtenidos).
- Asignación de examinadores por curso y establecimiento.
- Recepción del material en los centros de acopio.
- Revisión y cuadratura del material de aplicación recepcionado (antes de aplicar).
- Realización de visitas previas a los establecimientos.
- Operativo de los días de aplicación (salida y devolución del material y uso o devolución del material de contingencia).
- Recuperación de los cuestionarios.
- Revisión y cuadratura del material de aplicación recepcionado (después de aplicar).
- Despacho de los materiales desde los centros de acopio (incluyendo confirmación de entregas al proveedor de captura).

2.3 Aplicación de las pruebas

El proceso de aplicación de las pruebas Simce, al igual que los anteriores, se realiza mediante externalización de servicios. En la aplicación 2015 de las pruebas censales, este servicio fue adjudicado mediante licitación pública para toda la extensión del territorio, desde la Región de Arica y Parinacota hasta Magallanes[16].

Para realizar la licitación de aplicación de pruebas censales, la Agencia divide el territorio nacional en ítems. En 2015 se conformaron cinco ítems:

[16] Licitación "Servicio de aplicación de pruebas Simce 2015", ID 721703-9-LP15. Para la aplicación de la prueba muestral de Educación Física, la Agencia contrató a través de licitación pública, a un proveedor distinto del adjudicado para la aplicación de las pruebas censales.

- Ítem 1: regiones Metropolitana Norte[17], Arica y Parinacota, Tarapacá, Antofagasta, Atacama y Coquimbo.
- Ítem 2: Región de Valparaíso.
- Ítem 3: regiones Metropolitana Sur[18], Libertador Bernardo O'Higgins y Maule.
- Ítem 4: Región del Biobío.
- Ítem 5: regiones de La Araucanía, Los Ríos, Los Lagos, Aysén y Magallanes.

En cuanto al calendario de aplicaciones 2015, este se estableció en las siguientes fechas para las pruebas censales:

- 2° básico: 8 de octubre
- 4° básico: 6 y 7 de octubre
- 6° básico: 14 y 15 de octubre
- 8° básico: 20 y 21 de octubre
- II medio: 27 y 28 de octubre

Por su parte, la prueba muestral de Educación Física se aplicó entre el 2 y el 20 de noviembre de 2015.

2.3.1 Población evaluada

En términos de la población evaluada, las pruebas Simce 2015 de Lectura, Matemática, Historia, Geografía y Ciencias Sociales y Ciencias Naturales, fueron aplicadas en promedio, en 99% de los establecimientos y 93% de los estudiantes empadronados (1.105.878 estudiantes), con diferencias de cobertura según nivel evaluado, como se detalla en la Tabla 2.2. Los establecimientos en los cuales no fue posible aplicar las pruebas, corresponden a aquellos que se encontraban cerrados (y esta información no estaba actualizada en la base SIGE) y a aquellos que no estaban en funcionamiento por motivos de fuerza mayor, como los que se indicó al comienzo de este capítulo.

En cuanto a la población evaluada en la prueba Simce de Escritura, esta corresponde a los 197.908 estudiantes de 6° básico del país, matriculados en establecimientos con diez o más estudiantes por curso (5.160 establecimientos)[19], lo que representa 82,4% de la matrícula total nacional de ese nivel escolar.

[17] Comunas de Cerro Navia, Colina, Conchalí, Estación Central, Huechuraba, Independencia, La Reina, Lampa, Las Condes, Lo Barnechea, Lo Prado, Macul, Ñuñoa, Pedro Aguirre Cerda, Peñalolén, Providencia, Pudahuel, Quilicura, Quinta Normal, Recoleta, Renca, San Joaquín, San Miguel, Santiago, Tiltil y Vitacura.

[18] Comunas de Alhué, Buin, Calera de Tango, Cerrillos, Curacaví, El Bosque, El Monte, Isla de Maipo, La Cisterna, La Florida, La Granja, La Pintana, Lo Espejo, Maipú, María Pinto, Melipilla, Padre Hurtado, Paine, Peñaflor, Pirque, Puente Alto, San Bernardo, San José de Maipo, San Pedro, San Ramón y Talagante.

[19] Esto se debe a que, como se explicó en el capítulo anterior, la prueba Simce de Escritura no posee una versión extendida que la haga aplicable a establecimientos con menos de 10 estudiantes.

Tabla 2.2

Establecimientos y estudiantes evaluados en la aplicación censal de las pruebas Simce 2015

Nivel escolar	N.° establecimientos evaluados	% cobertura establecimientos	N.° estudiantes matriculados	N.° estudiantes evaluados (*)	% estudiantes evaluados
2° básico	7.414	98,3%	244.545	222.180	90.9%
4° básico	7.431	98,5%	231.508	220.335	95,2%
6° básico	7.367	98,5%	239.460	224.559	93,8%
8° básico	5.946	99,2%	242.142	229.636	94,8%
II medio	2.864	99,1%	226.325	209.168	92,4%

* Corresponden a aquellos estudiantes que estuvieron presentes en la aplicación de al menos una de las pruebas Simce aplicadas en ese nivel escolar.

Por su parte, como se muestra en la Tabla 2.3, la prueba muestral de Educación Física fue rendida por 9.568 estudiantes de 8° básico, distribuidos en 309 establecimientos a lo largo del país. Esto implicó una cobertura del 82,6% de los establecimientos empadronados y 61% de los estudiantes definidos en la muestra teórica. Este porcentaje es suficientemente representativo para entregar resultados por diversas agregaciones, en función del uso de expansores de la muestra.

Tabla 2.3

Establecimientos y estudiantes evaluados en la aplicación muestral del Educación Física 2015. Nivel 8° básico

Unidad	N.° muestral	N.° evaluado	% evaluado
Estudiantes	15.696	9.568	61,0%
Establecimientos	374	309	82,6%

Finalmente, las pruebas Simce DS fueron rendidas por 84,9% de los estudiantes de 6° básico a nivel nacional que presentan discapacidad sensorial, lo que corresponde a 220 estudiantes, distribuidos en 162 establecimientos[20].

2.3.2 Fases del proceso de aplicación

La aplicación supone tres fases operativas: definición de centros de acopio y centros operativos a nivel regional; activación de la plataforma de información y seguimiento; y recepción de materiales y documentos, para su distribución posterior a las entidades capturadoras; además cuenta con tres fases que involucran a examinadores y supervisores, las cuales se describen brevemente a continuación:

[20] El detalle de la población evaluada en cada nivel escolar y asignatura, incluyendo la prueba DS y la prueba muestral de Educación Física, se encuentran en el archivo "Cobertura 2015" disponible, previa petición formal, en www.agenciaeducacion.cl, índice Simce/bases de datos nacionales 2015.

i. Convocatoria, selección y capacitación de examinadores y supervisores

En esta fase, la Agencia fija un perfil para cada uno de estos cargos y luego, el proveedor del servicio convoca y selecciona a los candidatos que cumplen con cada perfil y realiza las contrataciones y capacitaciones correspondientes.

La convocatoria de examinadores y supervisores la realiza el proveedor del servicio de aplicación, mediante llamado abierto, publicado en los medios de comunicación regionales.

La selección de los postulantes contempla la revisión de antecedentes en función del perfil de cada cargo y el nivel de logro alcanzado en dos pruebas de selección diseñadas por la Agencia para este propósito: la Prueba de Aplicación Regular (PAR) y la Prueba de Evaluación de Competencias del Examinador (TEC). La prueba PAR tiene el propósito de medir el nivel de conocimiento y dominio de los candidatos respecto de los conceptos y orientaciones que la Agencia define para las aplicaciones Simce, los que se explicitan en el Manual de Aplicación. La Prueba TEC, por su parte, evalúa competencias y habilidades conductuales de los postulantes en relación con el perfil. En particular, esta evaluación mide habilidades de razonamiento, orientación pro-social, estabilidad emocional, dominancia social y ajuste a las normas. En el proceso 2015 se agregó una sexta dimensión: manejo de impresión.

En 2015 se seleccionó a todos los postulantes que, cumpliendo con el perfil, respondieron correctamente 90% o más de las preguntas de las pruebas antes mencionadas. En cuanto a la cantidad de seleccionados, la Agencia exigió que en cada grupo de aplicadores se seleccionara al menos 10% adicional de los requeridos efectivamente, como contingencia, con el objeto de cubrir eventuales ausencias de los titulares.

Por su parte, para la selección de examinadores de las pruebas Simce DS, se considera como candidatos solo a quienes hayan aprobado la prueba PAR y hayan cursado satisfactoriamente la capacitación para la aplicación regular. A los postulantes a aplicadores de estas pruebas se les aplican dos pruebas adicionales: la Prueba Selección Aplicación DS, que mide el dominio de los procedimientos del Manual de Aplicación Simce DS, y la Evaluación de Competencias Técnicas en DS, que determina el grado de dominio que tienen los postulantes sobre la atención educativa requerida por los estudiantes con discapacidad sensorial, como las ayudas técnicas y los sistemas de comunicación.

Una vez seleccionados los examinadores y supervisores, se inicia la capacitación, que está a cargo de relatores (previamente formados por la Agencia para tal efecto) y se basa en materiales elaborados por los equipos de la Agencia (manuales, presentaciones y formularios). Estos materiales de capacitación permiten estandarizar los procesos de formación, pero igualmente la Agencia supervisa una muestra de las sesiones de capacitación para evitar variaciones en la entrega de información que posteriormente puedan derivar en fallas sistemáticas o aleatorias que pongan en riesgo la validez de los resultados de la aplicación. En específico, los supervisores de la Agencia cumplen las funciones de: validar que los centros de capacitación cumplan con el plan de capacitación definido, asegurar que las capacitaciones se desarrollen en los términos convenidos y evaluar a los postulantes a examinadores y supervisores.

La capacitación de examinadores y supervisores incluye exposición de la información relevante, revisión de los materiales a utilizar en la aplicación y talleres prácticos de ejercitación en el uso de dicho material.

Los documentos informativos que se abordan en la capacitación son los manuales de aplicación, existiendo uno distinto para cada tipo de aplicación. En 2015 se utilizaron los siguientes manuales:

- Manual de Aplicación pruebas censales y experimentales.
- Manual de Aplicación Discapacidad Sensorial[21].
- Manual de Aplicación Estudio Nacional de Educación Física 8° básico.

En estos manuales están consignadas las características de la aplicación, según la prueba a aplicar; las responsabilidades y funciones del examinador; los procedimientos que deben seguirse antes, durante y después de la aplicación; y las consideraciones especiales requeridas en algunas de las pruebas.

Los documentos que se utilizan en los talleres prácticos corresponden al "material complementario", esto es, los formularios que el examinador deberá completar en las distintas fases del proceso.

Para la aplicación regular, se utilizan dos formularios:

- Formulario Lista de Curso, que constituye el documento guía de la aplicación por curso. En él se incluye el detalle de los códigos y series de las pruebas asignadas a cada estudiante en cada curso y la identificación del establecimiento y de los estudiantes. Este formulario contempla el registro de hasta 48 estudiantes en cada curso y permite que los examinadores completen la información a mano, con lápiz pasta, en situaciones de reasignación de identificación o incorporación de estudiantes nuevos no registrados en la base previamente impresa.
- Formulario de Aplicación, en el que, identificando los datos del establecimiento y del curso, se registra toda la información relacionada con la gestión de la aplicación: cantidad de material recibido y entregado, pruebas anuladas, uso de material de contingencia, etc. Este formulario permite además, tanto al examinador como al director del establecimiento, registrar cualquier situación anómala que ocurra durante la aplicación, para lo cual cuenta con una sección denominada "Acta de Control de Eventos[22]".

Para la aplicación Simce DS, por su parte, se utiliza el Formulario DS, que corresponde a una fusión de los formularios utilizados en la aplicación regular, es decir, incluye las mismas secciones de ambos documentos, pero con adaptaciones ad hoc a los procesos de control y registro que se requieren en esta aplicación específica.

ii. Realización de la visita previa a los establecimientos

La visita previa a los establecimientos constituye el primer contacto entre el equipo de aplicación (examinadores y supervisores) y los establecimientos. En esta instancia se realizan los preparativos para la aplicación y se verifican y validan los datos de los estudiantes.

[21] La capacitación de examinadores DS, en cuanto a los materiales utilizados y contenidos tratados, se desarrolla según las pautas internacionales de pruebas estandarizadas, con el fin de ofrecer condiciones idénticas a los estudiantes evaluados de todos los cursos y establecimientos del país (AERA, APA y NCME, 2014 y OCDE, 2011).

[22] Para la aplicación experimental, el Formulario de Aplicación incluye una sección para el registro de las consultas de los estudiantes durante la aplicación (información que sirve de base para el análisis posterior de la calidad de los ítems).

En este hito, el equipo de aplicación debe reunirse tanto con la dirección del establecimiento, como con cada uno de los cursos en los que se aplicarán las pruebas.

En la reunión con la dirección del establecimiento, el equipo debe:

- Informar sobre los procedimientos de la aplicación y señalar el cronograma de aplicación correspondiente a cada día.
- Entregar el modelo de comunicación informativa de la aplicación que debe socializarse con los padres y apoderados.
- Solicitar la colaboración de los directivos el día de la aplicación, para que el proceso se desarrolle en condiciones óptimas. Sobre este punto, dada la importancia de la colaboración de los directores de los establecimientos, en 2015 la Agencia envió a los directores con antelación al proceso de aplicación, un documento en formato digital e impreso denominado Orientaciones para Directores (uno para directores de establecimientos de enseñanza básica y otro para los de enseñanza media). En este documento se incluyó información detallada del proceso y de las actividades para las cuales se requeriría su colaboración.
- Verificar que la información impresa en el Formulario Lista de Curso de cada estudiante coincida con la información contenida en los libros de clases, y completar y corregir la información, en caso de detectar discrepancias.

En la visita a los cursos, el equipo debe:

- Presentarse a los estudiantes y motivarlos para que asistan el día de la aplicación y rindan las evaluaciones seriamente.
- Informar sobre los materiales que recibirán el día de la aplicación (cuadernillos, lápiz grafito y goma de borrar).
- Indicar sobre la prohibición de uso de artículos electrónicos durante la aplicación, tales como celulares, calculadoras, etc., así como de cualquier otro material de apoyo, como diccionarios, libros, cuadernos, etc.
- Verificar que la sala de clases que se ocupará el día de la aplicación posea una adecuada luminosidad, que se encuentre en buen estado, que cuente con el número suficiente de mesas y sillas, que no esté próxima a ruidos externos molestos, y que no posea material didáctico a la vista. En caso de que la sala no cumpla con algunos de estos requisitos, el examinador debe solicitar al director del establecimiento el cambio de sala o su acondicionamiento, para el día de la aplicación.

iii. Aplicación de las pruebas en aula

En función del calendario de evaluaciones, en cada nivel escolar se evalúa una o más asignaturas. En caso de aplicarse más de una prueba, cada una se aplica por separado.

La aplicación de cada prueba regular tiene una duración mínima de 45 minutos y una máxima de 90 minutos, a excepción de las pruebas de 2º básico para las que existe un tiempo mínimo de 30 minutos y no se fija tiempo máximo. Adicionalmente, junto con las pruebas censales, se aplica el Cuestionario para Estudiantes (salvo en 2º básico), para el cual se dispone de un tiempo mínimo de 25 minutos y un máximo de 60 minutos.

En cada uno de los días de aplicación, el examinador debe cumplir con tres funciones claves: (a) retirar el material de aplicación (caja-curso del día) desde el centro de operaciones o centro de acopio

correspondiente al establecimiento; (b) aplicar las pruebas y cuestionarios; y (c) devolver el material aplicado al centro de operaciones o centro de acopio respectivo.

Respecto de la aplicación de pruebas y cuestionarios, antes de la aplicación, la primera tarea del examinador consiste en revisar el material recibido y reportar al supervisor eventuales inconsistencias. Luego debe preparar la sala de clases en función de los criterios de ordenación de puestos y asignación de cuadernillos que se establecen en el Manual de Aplicación. Una vez que está todo dispuesto, instruye a los estudiantes para que ingresen a la sala y, finalmente, verifica la asistencia y da lectura en voz alta de las instrucciones generales de cada prueba, contenidas en la portada de los cuadernillos. Esto último es clave para asegurar que todos los estudiantes conozcan previamente las condiciones generales para responder la prueba. Durante la aplicación, el examinador debe registrar cualquier situación o acontecimiento irregular en la sección correspondiente del Formulario de Aplicación y, una vez finalizada, debe retirar las pruebas, realizar cuadratura del material y sellar la caja que, con las pruebas en su interior, debe ser devuelta al centro de acopio correspondiente.

Al igual que en la etapa de capacitación, la aplicación en aula es supervisada con el objeto de resguardar que se cumplan los estándares del proceso y los procedimientos definidos. En 2015, la supervisión fue realizada por un equipo de profesionales de la Agencia, con apoyo de profesionales del Mineduc, a una muestra de 524 cursos a nivel nacional y se llevó a cabo en función de una serie de criterios establecidos en una "Pauta de observación de la aplicación en aula".

Como se señaló al comienzo de este apartado, junto con las fases antes descritas, que involucran directamente a examinadores, supervisores de aplicación y demás estamentos de supervisión, se encuentran los procesos que involucran a los centros de acopio y centros de operaciones, que son los encargados de la recepción, almacenamiento, custodia y distribución del material de aplicación Simce.

El servicio que prestan estos centros es de tal relevancia que los supervisores contratados por la Agencia deben realizar su labor en al menos las siguientes cuatro etapas:

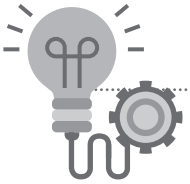
- Constatación de características y requerimientos mínimos de infraestructura y ubicación de los centros. En esta etapa, los supervisores deben constatar, por medio de visitas en terreno, las condiciones de infraestructura de cada uno de los centros de acopio y de operaciones establecidos en cada región, de acuerdo a las características exigidas por las bases de licitación a los proveedores del servicio de aplicación.
- Monitoreo del cumplimiento de los procedimientos por parte del proveedor del servicio de aplicación. Esta etapa supone la observación de los procesos vinculados a la salida del material complementario para las visitas previas a cada establecimiento y a la recepción, descarga y almacenamiento del material complementario y del material de aplicación, incluyendo gestión de personal y cumplimiento de planes de contingencia, entre otros aspectos.
- Observación del funcionamiento de los centros de operación durante los días de aplicación. En esta etapa se verifica, a través de visitas en terreno, el cumplimiento de tiempos del proveedor de aplicación (por ejemplo, constatar los horarios de llegada y salida de los examinadores y supervisores desde los centros de operación) y el cumplimiento de los protocolos de traslado del material (tales como los que regulan el orden de salida y traslado de cajas) y de recepción del mismo.
- Resguardo del mecanismo de devolución del material aplicado al operador logístico (empresa de distribución y retiro). En esta etapa se monitorea la forma, los tiempos y la conformidad o disconformidad con que se desarrollan los procesos de devolución de material al proveedor de distribución, por cada nivel.

En 2015, este proceso de supervisión se llevó a cabo en los 16 centros de acopio y en los 96 centros de operaciones dispuestos a nivel nacional, cubriendo así el 100% de estos puntos. Esta supervisión permitió corregir a tiempo problemas de habilitación de algunos de los lugares de operación de estos centros y atender a problemas que se originaron los días de aplicación; en particular, se registraron atrasos en la apertura de los centros y en la llegada de examinadores a retirar el material que debían trasladar a los establecimientos.

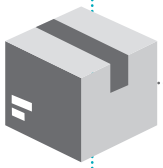
2.4 Entrega del material a los centros de captura

El retorno del material aplicado es una de las fases en las que el resguardo de la seguridad resulta crítico. En esta etapa las pruebas y cuestionarios respondidos por los estudiantes deben trasladarse desde los centros de acopio y de operaciones a los centros de captura.

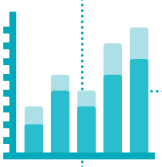
En 2015 este proceso estuvo a cargo de los mismos proveedores de los servicios de aplicación y se ejecutó dentro del plazo de 15 días hábiles establecido, contados a partir de la fecha de término de la aplicación en cuestión. Para asegurar la correcta recepción de todas las cajas y evitar congestiones en los centros de captura, se establecieron cuotas diarias de entrega del material aplicado.



Capítulo I:
Diseño y construcción de las pruebas Simce



Capítulo II:
Logística y aplicación de las pruebas Simce



Capítulo III:
Procesamiento y análisis de datos

Capítulo IV:
Comunicación de resultados educativos

Capítulo III: Procesamiento y análisis de datos

La fase de procesamiento y análisis de datos de las pruebas Simce, al igual que las anteriores, se realiza bajo estrictos controles de calidad, de manera de poder asegurar que los resultados obtenidos reflejen correctamente los constructos medidos. El procesamiento supone la creación de una base de datos que consolida la información de las respuestas de los estudiantes a las pruebas y la información de los cuestionarios para estudiantes, docentes, directivos y padres y apoderados. El análisis de los datos, por su parte, involucra tanto los análisis psicométricos de las pruebas, como los análisis agregados, tales como promedios nacionales, regionales, por género, etc.

En este capítulo se describen las acciones involucradas en la etapa de procesamiento y análisis de datos de las aplicaciones Simce, a saber:

- Elaboración y validación de la base de datos de pruebas y cuestionarios
- Análisis psicométrico de las pruebas
- Análisis de datos agregados

3.1 Elaboración y validación de la base de datos de pruebas y cuestionarios

La elaboración y validación de la base de datos de pruebas y cuestionarios de la aplicación 2015 se llevó a cabo entre noviembre de 2014 y enero de 2015.

Para obtener una base de datos consolidada y validada, la primera acción consiste en definir el "Plan de gestión de datos", documento que incluye los criterios y etapas que deberán desarrollar los proveedores externos del servicio. Esta acción se realiza antes de la aplicación de las pruebas y en 2015 involucró aspectos de mejora respecto del plan establecido en la medición anterior, tales como la inclusión de un "Informe de recepción de cajas en captura" que permitió rastrear las cajas que demoraban en llegar a los centros de captura.

Una vez establecido el Plan de gestión de datos, la siguiente acción consiste en generar el "diccionario de datos" que define los criterios a partir de los cuales se sistematizará la información de las respuestas y las normas, codificaciones y los rangos de valores permitidos para cada uno de los campos que serán utilizados en el procesamiento. La información para construir el diccionario de datos se obtiene de los distintos documentos utilizados durante la aplicación, como las listas de curso, los formularios de aplicación, los cuestionarios y las hojas de respuesta.

A partir de los documentos antes mencionados, se procede a construir las bases de datos, para lo cual se generan rutinas de identificación de estudiantes, cursos y establecimientos.

La identificación de los establecimientos se realiza con el número de Rol de Base de Datos (RBD) y para la identificación de los cursos se utiliza una variable de identificación del nivel escolar que consiste en un código de dos dígitos que identifica el nivel escolar y si dicho nivel corresponde a enseñanza básica (0) o enseñanza media (2). Así, en la aplicación 2015, se definieron cinco códigos para la variable de identificación, los que se muestran en la Tabla 3.1.

Tabla 3.1

Códigos de la variable de identificación por nivel escolar de la aplicación 2015

Nivel escolar	Código
2° básico	20
4° básico	40
6° básico	60
8° básico	80
II medio	22

A estos códigos se agrega la letra que designa el curso (por ejemplo, el 2° básico A se codifica 20A) y con ello se genera el Código de Localización Número Curso (CLN Curso).

La identificación de los estudiantes se realiza a través de un número de serie único (creado de manera exclusiva para este propósito) al cual se asocian todas las pruebas y cuestionarios respondidos por el estudiante, y luego se cataloga a los estudiantes en función de la variable de participación, cuya información se obtiene de las listas de curso. En la Tabla 3.2 se muestran las categorías de la variable de participación de estudiantes de la aplicación 2015.

A partir de la base de datos antes reseñada, se ajustan las rutinas de programación y los paquetes de carga (datos obtenidos de la captura de las respuestas a las preguntas cerradas y de la corrección de las preguntas abiertas[23]), para detectar posibles inconsistencias, y se realizan controles de calidad de la captura, en función de lo establecido en el Plan de gestión de datos. Para asegurar la calidad de estos procesos, cada año se actualiza el *software* de validación y corrección de los datos (Sistema de validación de datos, SVD), de acuerdo a las nuevas estructuras que presentan los formularios de control, las pruebas y los cuestionarios aplicados.

23 Véase el detalle del proceso de corrección de preguntas abiertas en el anexo B.

Tabla 3.2

Categorías de la variable de participación de estudiantes en la aplicación 2015

Categoría	Descripción
Presente	Estudiante asiste el día de la prueba.
Anulado	Estudiante atrasado, llega después de que se leen las instrucciones.
	Estudiante abandona por enfermedad (vómitos, náuseas, dolor de cabeza, dolor abdominal o diarrea, etc.).
	Error en la prueba (mala compaginación, impresión, legibilidad, etc.).
	Estudiante con mala conducta.
	Estudiante no habla español.
Anulado	Anulación Cuestionario Estudiantes.
	Otro que considere motivo de anulación.
No pertenece	Estudiante no pertenece al establecimiento.

Una vez que las rutinas y paquetes de carga se han implementado, se realizan pruebas de carga, utilizando un set de datos de 100 cursos por nivel escolar aplicado, el que es entregado por el proveedor de captura[24]. En esta etapa se respeta el formato establecido en el diccionario de datos y se verifica que las estructuras y contenidos cumplan con las especificaciones entregadas al inicio del contrato. La finalidad de este proceso es retroalimentar al proveedor en función de las inconsistencias detectadas, para que este realice las correcciones pertinentes antes de la entrega definitiva.

Posterior a estas pruebas de carga, se inicia el **proceso de captura**[25] que consiste en la lectura, a través de un *software*, de la información de los distintos documentos aplicados; su conversión a imágenes con caracteres digitales y, finalmente, a archivos de datos electrónicos. Para convertir la imagen con caracteres digitales a un archivo de datos electrónico, en 2015 se utilizaron tres procesos: Reconocimiento de Caracteres Inteligente (ICR), Reconocimiento Óptico de Caracteres (OCR) y Lectura de Marcas Ópticas (OMR). Para realizar el proceso de captura, el proveedor debía romper el sello de las cajas-curso, verificar que su contenido estuviera íntegro, ordenar el contenido por tipo de documento, guillotinar los documentos según los requerimientos del *software*, escanear la información para generar las imágenes[26] y traspasar los datos de imagen a formato CSV (valores separados por comas). En la Tabla 3.3 se muestra el detalle de los datos capturados en el proceso de aplicación 2015, en cada nivel escolar y prueba aplicada.

[24] En 2015 esta labor fue desarrollada por proveedores externos seleccionados en función de lo establecido en la licitación pública "Digitalización, captura, digitación y validación de datos y destrucción de documentos de la aplicación de pruebas Simce año 2015", ID 721703-13-LP15.

[25] Respecto del proceso de captura, cabe mencionar que el convenio con los proveedores establece que los materiales aplicados deben ser almacenados por un mínimo de 6 meses, hasta que la Agencia haya cerrado el procesamiento de los datos.

[26] Las respuestas a las preguntas abiertas también se escanean en esta fase y luego son entregadas a los proveedores del servicio de corrección, según se detalla en el anexo B.

Tabla 3.3

Datos capturados en 2015 por nivel escolar y prueba aplicada

Datos	2° básico	4° básico	6° básico	6° básico DS	8° básico	8° básico Ed. Física	II medio
Lista Curso (LC)	✓	✓	✓	✓	✓	✓	✓
Formulario Aplicación (FA)	✓	✓	✓	✓	✓		✓
Formulario Supervisión (FS)						✓	
Preguntas Cerradas (PC)	✓	✓	✓	✓	✓		✓
Preguntas Abiertas (PA)	✓	✓	✓	✓	✓		✓
Cuestionario Padres (CP)	✓	✓	✓	✓	✓		✓
Cuestionario Estudiantes (CE)		✓	✓	✓	✓		✓
Cuestionario Docentes Lenguaje (CDL)		✓	✓				✓
Cuestionario Docentes Matemática (CDM)			✓		✓		
Cuestionario Docentes Cs. Naturales (CDN)					✓		
Cuestionario Docentes Cs. Sociales (CDN)							✓

Con esta información la Agencia realiza los procesos de completitud y de validación de los datos.

El **proceso de completitud** consiste en verificar si todo el material impreso fue efectivamente capturado y entregado a la Agencia. El material extraviado se consigna en un campo de la Lista de Curso denominado "completitud", en el que se asigna el carácter cero (0) para identificar la existencia del material y uno (1) para identificar la falta de material. El proceso de completitud culmina con la generación de tres reportes: (a) reporte de completitud global (en el que se contrasta lo recibido con lo que debió llegar, detallando región, comuna y establecimiento); (b) reporte de completitud detallado (que contiene la información sobre el material faltante); y (c) reporte de completitud de gestión, en el que se comunica el estado de la base de datos a nivel macro y se establecen rangos de porcentaje de pérdidas que permiten prever posibles problemas en la entrega de resultados. En el proceso 2015 se obtuvo más de 90% de completitud.

La **validación de los datos** por su parte, se realiza mediante el SVD y distingue la validación de pruebas y la validación de cuestionarios.

El proceso de **validación de pruebas** consiste en la revisión de la consistencia e integridad de los datos de captura enviados por el proveedor y la corrección de la aplicación de acuerdo a las reglas de validación. En este proceso se marcan los registros en los que se detectan errores, se clasifican según formulario y prioridad, y se agrupan mediante un código. Luego se corrigen los errores en dos instancias consecutivas: corrección masiva y corrección registro a registro. La corrección masiva se realiza cuando un conjunto de registros marcados con error presenta un patrón idéntico. En estos casos se realiza una transacción que permite corregir masivamente el error en todos los registros afectados. La corrección registro a registro la realiza un equipo de profesionales (contratados especialmente para este propósito) al que se le asigna una serie de errores de manera aleatoria y, con la información de la prueba, la lista de curso, los cuestionarios y los formularios, cada validador resuelve las discrepancias encontradas. El proceso de validación de datos de pruebas 2015 permitió corregir 100% de los errores detectados modificables[27] y obtener una tasa aceptable de errores no modificables[28]: 5% en el caso de establecimientos con un curso por nivel evaluado y 3% en establecimientos con más de un curso por nivel evaluado.

La **validación de cuestionarios**, por su parte, consiste en verificar que existan respuestas para cada ítem y que la serie del cuestionario corresponda al estudiante, curso y establecimiento respectivos. Para validar los cuestionarios se compara la lista de empadronamiento (que contiene el total de cuestionarios de estudiantes, docentes y padres y apoderados impresos y capturados) con los datos de los cursos que efectivamente rindieron las pruebas y con los cuestionarios que retornaron de la aplicación. A partir de esta comparación de información se validan los encabezados, se analiza la cuadratura de los mismos y se corrobora que no existan duplicados de series o folios. Una vez validados los encabezados, se revisa caso a caso cada set de cuestionarios, para no incluir en el análisis posterior a aquellos que no correspondan. Por ejemplo, en el cuestionario de estudiantes, se requiere distinguir a aquellos que estuvieron presentes tanto al momento de rendir las pruebas como de contestar el cuestionario, de aquellos que únicamente rindieron las pruebas. También se debe velar porque cada estudiante tenga asociado un cuestionario docente[29].

Una vez ratificados todos los datos, se genera un reporte de frecuencias de análisis en el que se distingue cantidad y porcentaje de ocurrencia para cada campo capturado, lo que permite obtener una visión general de la omisión de respuestas y de las múltiples marcas, e identificar valores fuera de rango. En función de este reporte se definen los campos que es necesario corregir.

La última acción del proceso de elaboración y validación de la base de datos consiste en la **consolidación y cuadratura de la base**. La consolidación consiste en adjuntar a la base de datos la puntuación de la prueba de cada estudiante, en función de sus respuestas a las preguntas cerradas y a las preguntas abiertas, en caso que existan, y cargar los datos obtenidos de los cuestionarios, que permitirán realizar posteriormente los análisis agregados. Por su parte, la cuadratura consiste en un control de calidad de los datos consolidados, para lo cual se utiliza una muestra a partir de la cual se calculan las tasas de error por campo y material y luego se corrigen todos los errores detectados. En 2015 la muestra fue de 300 cursos por nivel evaluado.

[27] Se entiende por "error detectado modificable" aquellas inconsistencias producidas por el llenado de datos incorrectos en los formularios por parte del examinador o la incorrecta calibración de escáner que detecta respuestas donde hay borrones.

[28] Se entiende por "error detectado no modificable" las inconsistencias producidas por el estudiante, como, por ejemplo, que marque más de una alternativa ante una pregunta cerrada.

[29] Es decir, que el profesor del curso al que pertenece haya respondido el cuestionario Simce respectivo.

Los productos que se obtienen a partir de los procesos descritos en este apartado son los siguientes:

- Base de datos unificada, que incluye la información necesaria para elaborar los documentos comunicacionales de la Agencia, tales como el Informe de Resultados para Docentes y Directivos, el Informe de Resultados para Padres y Apoderados y la Ficha Web del Establecimiento. Esta base se encuentra disponible, previa petición formal, en www.agenciaeducacion.cl, índice Simce/bases de datos nacionales 2015.
- Base de datos de estudiantes, que contiene los resultados de cada estudiante y que es de uso restringido de la Agencia para propósitos de investigación interna.

3.2 Análisis psicométricos de las pruebas Simce

Los análisis psicométricos tienen por objetivo entregar indicadores de rendimiento individual de los estudiantes en cada una de las pruebas aplicadas, tanto a nivel de puntaje, como de clasificación en Estándares de Aprendizaje, de haberlos.

Estos análisis se realizan sobre la base de lineamientos internacionales (AERA, APA y NCME, 2014) que aseguran la estandarización de los procesos, y están sujetos a estrictos controles de calidad, incluyendo un procesamiento en paralelo por al menos dos profesionales, lo que permite detectar inconsistencias y de haberlas se verifican y ajustan de manera inmediata.

En las aplicaciones Simce los análisis psicométricos se realizan utilizando como insumo los archivos denominados "Planilla Maestra" y "Base de datos de estudiantes", los que contienen la información de los ítems aplicados (identificación, objetivo de evaluación, respuesta correcta (clave), posición, etc.) y de las respuestas de los estudiantes, respectivamente. Las respuestas a las preguntas cerradas tienen dos puntuaciones posibles: 0 (incorrecta) y 1 (correcta), y las respuestas a preguntas abiertas pueden tener una puntuación igual a las cerradas (respuesta dicotómica) o bien una de tres niveles o politómica: 0 (incorrecta), 1 (parcialmente correcta) y 2 (correcta), dependiendo de lo que establezca la pauta de corrección respectiva.

Los modelos que orientan los análisis psicométricos de las pruebas Simce son la Teoría Clásica y la Teoría de Respuesta al Ítem (IRT).

En 2015, la Teoría Clásica se utilizó para el análisis de las pruebas Simce DS, prueba de Escritura de 6° básico y prueba muestral de Educación Física (además del análisis de pruebas experimentales, detallado en el anexo A y para su ejecución se utilizaron los *softwares* SAS 9.1 (versión 2006) y TIAPLUS: build 307 (versión2010). El modelo IRT, por su parte, se utilizó para realizar los análisis de las pruebas censales de Lectura, Matemática, Historia, Geografía y Ciencias Sociales y Ciencias Naturales. En particular, se utilizaron: el Modelo Logístico de Tres Parámetros (3PL) para las preguntas cerradas; el Modelo Logístico de dos Parámetros (2PL), para los ítems de respuesta abierta dicotómica; y el Modelo de Crédito Parcial Generalizado (MCPG), para los ítems de respuesta abierta politómica.

Realizar los análisis según el modelo IRT, hace necesario revisar cada año el cumplimiento de los tres supuestos bajo los cuales este modelo resulta eficiente y robusto, a saber: (a) unidimensionalidad, esto es que las pruebas evalúen una dimensión de rendimiento predominante; (b) independencia local, es decir, que para los estudiantes con un mismo nivel de rendimiento no existan otros factores

que expliquen sus respuestas correctas a los ítems; y (c) monotonicidad creciente, es decir, que la probabilidad de responder correctamente a un ítem aumente progresivamente en aquellos estudiantes con mayor nivel de rendimiento.

En la aplicación 2015, para la revisión del supuesto de unidimensionalidad se aplicó un Análisis Factorial Exploratorio (AFE) (DeVellis, 2012) mediante el *software MPLUS*. Este *software* analiza la información a través de la magnitud de los valores propios resultantes y del porcentaje de varianza que estos entregan en el modelo, y luego permite visualizar los resultados en un *scree plot*[30]. Como resultado, en todas las pruebas 2015 se evidenció la existencia de un factor dominante, representado por un quiebre significativo entre la magnitud de los factores primero y segundo[31].

Para la revisión del supuesto de independencia local, en 2015 se realizó una revisión cualitativa tendiente a verificar que no existieran ítems encadenados, o que la respuesta correcta o incorrecta a un ítem no influyera en la respuesta correcta o incorrecta a otro. Adicionalmente, en aquellos conjuntos de ítems que compartían un mismo estímulo (imagen, texto, etc.), conocidos como teslet, se detectaron los ítems con valores inusualmente altos en el parámetro de discriminación, y se recalibró la prueba eliminando cada uno de manera separada. Si la estimación de los parámetros de discriminación restantes (sin el ítem eliminado) no poseía valores altos, se estimó que existía dependencia local y, por tanto, se eliminó el ítem.

Por último, para la revisión del supuesto de monotonicidad creciente, se revisaron las curvas características y las curvas de información y se verificó, en las curvas empíricas de cada ítem, que la correspondiente a la clave aumentara progresivamente su probabilidad en todos los ítems considerados para el análisis y su puntuación[32].

Ahora bien, respecto de los parámetros que se calculan en función de estos modelos, cabe señalar que estos dependen de si la pregunta es cerrada o abierta. En efecto, como se muestra en la Tabla 3.4, para las preguntas cerradas de la aplicación 2015 se calcularon seis indicadores según Teoría Clásica[33] y cinco según IRT. Para las abiertas, como se muestra en la Tabla 3.5, se calcularon dos indicadores de Teoría Clásica y cuatro de IRT.

[30] Método gráfico que muestra los valores propios de la matriz de correlación en orden descendente, esto es, se comienza con el valor propio de mayor magnitud, y se finaliza con el menor.

[31] Los *scree plot* de la aplicación 2015 se encuentran disponibles, previa petición formal, en www.agenciaeducacion.cl, índice Simce/bases de datos nacionales 2015.

[32] Los parámetros de los ítems y curvas características y de información de las pruebas se encuentran disponibles, previa petición formal, en www.agenciaeducacion.cl, índice Simce/bases de datos nacionales 2015.

[33] Adicionalmente se realizó una revisión de la correlación ítem-total de cada alternativa, con el objetivo de detectar posibles errores en la asignación de claves y remediar la situación antes de la etapa de calibración.

Tabla 3.4

Indicadores 2015 para preguntas de respuesta cerrada según Teoría Clásica e IRT

Teoría Clásica	IRT
• Porcentaje de respuesta para cada alternativa. • Correlación ítem-total de cada alternativa. • Índice de dificultad del ítem (porcentaje de aciertos). • Capacidad discriminativa del ítem (correlación ítem-total corregida). • Curvas empíricas de respuesta al ítem. • Porcentaje de omisión por ítems.	• Discriminación • Dificultad • Azar • RP67 • Indicador de comportamiento diferencial del ítem (por género y ruralidad).

Tabla 3.5

Indicadores 2015 para preguntas de respuesta abierta según Teoría Clásica e IRT

Teoría Clásica	IRT
• Distribución de respuestas por código de puntuación. • Capacidad discriminativa del ítem (correlación ítem-total, corregida).	• Discriminación • Dificultad • RP67 • Indicador de comportamiento diferencial del ítem (por género y ruralidad).

A continuación, se describen los análisis psicométricos realizados para puntuar a los estudiantes y clasificarlos según estándares en función de sus respuestas a las pruebas censales y también los análisis psicométricos realizados sobre las otras pruebas 2015 (Simce DS, prueba de Escritura y prueba muestral de Educación Física).

3.2.1 Análisis psicométricos de las pruebas Simce censales

Como se señaló anteriormente, para el análisis psicométrico de las pruebas Simce censales se utiliza como metodología base el modelo IRT. Para la puntuación de los estudiantes se emplea la estimación Expected A Posteriori (EAP), que, de acuerdo con lo planteado por Bock y Mislevy (1982), presenta diversas ventajas respecto de otros estimadores, como los basados en Máxima Verosimilitud (ML) o Máxima Verosimilitud a Posteriori (MAP). Entre estas ventajas destacan la facilidad de cálculo (dado que no es un proceso iterativo y no es necesario obtener la derivada de la función de verosimilitud) y la estabilidad para pruebas con extensiones diversas. Adicionalmente, según lo señalado por ACER (2013), la estimación EAP permite distinguir, con mayor precisión que otras, el promedio de agrupamiento de estudiantes (establecimientos).

Bajo el modelo 3PL, la probabilidad de responder correctamente está dada por:

$$P_i = c_i + (1 - c_i)\frac{exp[Da_i(\Theta - b_i)]}{1 + exp[Da_i(\Theta - b_i)]}$$

Donde:

- a_i, b_i y c_i = parámetros del ítem i
- D = constante de escalamiento

En función de lo anterior, la probabilidad de respuesta Xi a un ítem está dada por:

$$P(X_i|\Theta, a_i, b_i, c_i) = P(X_i = 1|\Theta, a_i, b_i, c_i) \bullet P(X_i = 0|\Theta, a_i, b_i, c_i)$$
$$= (P_i)^{X_i} \bullet (1 - P_i)^{(1-X_i)}$$
$$= (P_i)^{X_i} \bullet (Q_i)^{(1-X_i)}$$

Considerando las respuestas de N examinados a los K ítems, la función de verosimilitud se calcula en función de lo siguiente:

$$L(x_{11}, x_{12}, ..., x_{ji}, ..., x_{NK}|\Theta, a_i, b_i, c_i) = \prod_{j=1}^{N}\prod_{i=1}^{K}(P_{ji})^{X_{ji}} \bullet (Q_{ji})^{(1-X_{ji})}$$

De este modo, para obtener las estimaciones de las habilidades y de los parámetros es necesario resolver las siguientes ecuaciones:

$$\frac{\partial lnL}{\partial \Theta_j} = 0, j = 1, ..., N$$
$$\frac{\partial lnL}{\partial a_i} = 0, i = 1, ..., K$$
$$\frac{\partial lnL}{\partial b_i} = 0, i = 1, ..., K$$
$$\frac{\partial lnL}{\partial c_i} = 0, i = 1, ..., K$$

El análisis psicométrico de las pruebas Simce censales 2015 se realizó en cuatro etapas.

En la primera etapa se realizó una estimación de parámetros de los ítems que componen cada prueba sin considerar ningún anclaje con la medición anterior (en caso de que existiera). En términos de equiparación entre formas de la prueba, se optó por una calibración concurrente mediante ítems comunes entre ellas. Para las preguntas cerradas se utilizó una calibración bajo el modelo logístico 3PL, con constante de escalamiento 1,7, dado que para este valor el modelo se aproxima a la curva normal acumulada (Martínez, 2005). Los ítems de respuesta abierta dicotómica, por su parte, se analizaron

bajo el modelo 2PL, porque el efecto azar no necesita ser estimado. Por su parte, las preguntas abiertas de respuesta politómica, se analizaron bajo el Modelo de Crédito Parcial Generalizado (MCPG), que es similar al modelo 2PL, pero permite tener más categorías. Todo lo anterior se realizó con el programa PARSCALE 4.1 (2003). Desde el punto de vista técnico, la calibración se realizó con el método ML, a través del algoritmo *Expectation Maximization* (EM), especificando un máximo de 100 iteraciones, con 41 puntos de cuadratura y un criterio de convergencia de 0,005. Cada ítem de respuesta abierta se especificó en un bloque distinto (comando *BLOCK*). Los criterios que se debieron satisfacer para determinar la solución fueron los siguientes: (a) la función converge en menos de 100 iteraciones y (b) no existen errores estándar de dificultad mayores a 0,210.

En la segunda etapa los ítems de *equating* se compararon en términos de los parámetros estimados mediante la calibración libre (primera etapa) y de los parámetros obtenidos en el análisis de las pruebas anteriores. Este análisis de variación de los parámetros (conocido como *drift*, por su denominación en inglés) permitió determinar cómo los parámetros de los ítems se afectan con el paso del tiempo. Para detectar la posible presencia de un efecto *drift*, se utilizó la metodología de distancias perpendiculares (Crocker y Algina, 1996), según la cual un ítem es removido del conjunto de equiparación si la distancia estandarizada a la recta, en valor absoluto, es mayor a 2 para los parámetros de dificultad y/o de discriminación. Los ítems removidos o desanclados son incorporados al análisis de la prueba como ítems nuevos.

Una vez obtenido el conjunto de ítems de equiparación que no presentaba *drift* entre las dos mediciones comparadas, se inició la tercera etapa, que consistió en la equiparación de las pruebas definitivas con la última medición de la serie, con el propósito de conseguir estimaciones en una métrica comparable entre ambas mediciones. En esta etapa, se recalibró la prueba, importando como parámetros fijos los parámetros de los ítems de anclaje de las pruebas definitivas, y los restantes ítems se ajustaron a la métrica de estos. Para obtener la calibración óptima de las pruebas definitivas con parámetros anclados a la medición anterior, se requirió que el conjunto de preguntas satisficiera criterios aún más exigentes que los utilizados en la calibración libre. Estos criterios fueron: (a) convergencia de la función en menos de 100 iteraciones; (b) errores estándar de la dificultad, menores a 0,210; (c) azar menor o igual a 0,35, en los ítems de cuatro alternativas y menor o igual a 0,45, en los de tres alternativas; (d) discriminación mayor o igual a 0,6; y (e) parámetros de dificultad entre -2,4 y 2,4.

Por último, luego de obtener estimaciones estables en los parámetros de las pruebas anclados a la última medición de la serie, se procedió a ejecutar la cuarta y última etapa, consistente en el cálculo de la puntuación de cada estudiante. Esta se realizó a partir del patrón de respuestas individual, fijando los parámetros de todos los ítems ajustados y utilizando la estimación EAP. Para cada uno de los niveles y asignaturas evaluados, la habilidad estimada y su respectivo error de estimación fueron transformados a la escala Simce anclada[34], que posee una media de 250 puntos y una desviación estándar de 50 puntos. Adicionalmente, para la puntuación de los ejes de contenido de Matemática y Ciencias Naturales, se puntuó cada uno de los ejes utilizando la solución final de la calibración de la prueba completa respectiva. Una vez obtenidos los puntajes por eje se utilizó una escala con una media de 5 puntos y una desviación estándar de 2, la que fue truncada en los extremos dejando un mínimo de 0 puntos y un máximo de 10 puntos para los puntajes individuales. La elección de esta escala permitió obtener diferencias significativas de aproximadamente 1 punto, lo que simplificó las comparaciones.

[34] La escala Simce anclada para cada nivel escolar se definió en su año de origen, cuando se comenzó a analizar en función de IRT. Para II medio el año de origen fue 1998; para 4° básico, 1999; para 8° básico, 2000; para 2° básico, 2012, y para 6° básico, 2014.

Los análisis anteriores dejaron en evidencia resultados atípicos de la aplicación 2015. En efecto, tanto en la prueba de Lectura de II medio como en la de Ciencias Naturales de 8° básico, se observaron variaciones significativas en los puntajes promedio a nivel nacional. Este hecho derivó en una serie de nuevos análisis tendientes a determinar si dicha diferencia era real o atribuible a los instrumentos y/o a la metodología de análisis. Desde los análisis psicométricos, se realizaron ajustes en el tratamiento de los ítems de equiparación y se decidió, en acuerdo con la Secretaría Ejecutiva de la Agencia, presentar los resultados obtenidos por medio de la calibración concurrente[35] para los ítems de la serie actual y continuar utilizando el modelo 3PL.

3.2.2 Análisis psicométricos de las pruebas Simce censales

Como se señaló en el primer capítulo de este informe, en las asignaturas y niveles escolares en los que existen Estándares de Aprendizaje, parte de los resultados reportan en función de estos. Es decir, junto con la puntuación, se clasifica a los estudiantes según estos estándares que describen lo que los estudiantes saben y pueden hacer en las evaluaciones Simce, y constituyen un referente de distintos grados de cumplimiento de los objetivos de aprendizaje establecidos en el Currículo Nacional. Estos estándares comprenden tres niveles de aprendizaje (Adecuado, Elemental e Insuficiente) que son inclusivos, es decir, un estudiante que alcanza el nivel adecuado no solo demuestra que cumple con los requisitos establecidos para dicho nivel, sino que además cumple con los del nivel elemental.

La clasificación de los resultados 2015 en función de Estándares de Aprendizaje se realizó en todos los niveles escolares evaluados, a excepción de 6° básico, por cuanto, a la fecha de los análisis, estos aún no habían sido publicados.

Para la clasificación en estándares se utilizan puntajes de corte (véanse los utilizados en la aplicación 2015, en la Tabla 3.6).

Tabla 3.6

Puntajes de corte para la clasificación 2015 según Estándares de Aprendizaje

Nivel escolar	Prueba	Insuficiente	Elemental	Adecuado
2° básico	Lectura	X < 215 puntos	215 ≤ X < 265 puntos	X ≥ 265 puntos
4° básico	Lectura	X < 241 puntos	241 ≤ X < 284 puntos	X ≥ 284 puntos
	Matemática	X < 245 puntos	245 ≤ X < 295 puntos	X ≥ 295 puntos
8° básico	Lectura	X < 244 puntos	244 ≤ X < 292 puntos	X ≥ 292 puntos
	Matemática	X < 247 puntos	247 ≤ X < 297 puntos	X ≥ 297 puntos
	Historia, Geo. y Cs. Sociales[a]	X < 243 puntos	243 ≤ X < 296 puntos	X ≥ 296 puntos
II medio	Lectura	X < 250 puntos	250 ≤ X < 295 puntos	X ≥ 295 puntos
	Matemática	X < 252 puntos	252 ≤ X < 319 puntos	X ≥ 319 puntos

Nota: [a] X = puntaje obtenido por el estudiante.

[35] La calibración concurrente consiste en utilizar los datos de ambas mediciones comparadas, la actual y la última medición de la serie, para estimar los parámetros de los ítems de la serie actual. Luego, a través de transformaciones lineales, se obtienen los puntajes individuales de los estudiantes evaluados en la serie actual.

Como información adicional a la estimación de parámetros obtenidos en la calibración de ítems, se calcula el índice denominado RP67, que corresponde a la habilidad necesaria para obtener 67% de probabilidad de acertar la respuesta correcta de un ítem. Este indicador es utilizado para asociar los ítems a los Estándares de Aprendizaje y corresponde a una estimación de la dificultad reescalada a 0,67 (2/3) de probabilidad de acierto del ítem o alternativa, en el caso de ítems politómicos. El objetivo de calcular este valor es proveer información necesaria para establecer los futuros Estándares de Aprendizaje[36].

La estimación del valor RP67 para ítems dicotómicos se obtiene mediante la siguiente ecuación, la cual es derivada a partir de la probabilidad de responder correctamente un ítem en el modelo 3PL:

$$\Theta_{3PL} = \left(\frac{1}{Da}\right) \cdot ln\left(\frac{RP - c}{1 - RP}\right) + b$$

El procedimiento anterior es válido solo para preguntas cerradas y para preguntas abiertas con dos alternativas de puntuación. En el caso de los ítems de respuesta abierta puntuados de manera politómica, se calcula la dificultad reescalada utilizando diferentes ecuaciones, dependiendo de la cantidad de opciones, como lo establece el modelamiento de alternativas con el MCPG.

Finalmente, a modo de confirmación, se verifica la presencia de Funcionamiento Diferencial de los Ítems (DIF por sus siglas en inglés), según corresponda al grado evaluado (educación básica o educación media), aun cuando en el análisis experimental de las preguntas se haya descartado la presencia de sesgo en las variables género, ruralidad[37] y grupo socioeconómico (GSE). Para calcular DIF se considera como grupo de referencia aquel que, en teoría, está aventajado (en género, los hombres; en ruralidad, los establecimientos urbanos y en GSE, el estrato más alto) y como indicador de DIF se contempla una diferencia del parámetro, en valor absoluto, igual o mayor a 0,4[38].

3.2.3 Análisis psicométricos de las pruebas adicionales Simce

Para los efectos de los análisis psicométricos, la prueba de Discapacidad Sensorial (Simce DS), la prueba de Escritura y la prueba muestral de Educación Física aplicadas en 2015 requirieron algunos tratamientos especiales que se describen a continuación.

En las pruebas Simce DS, para transformar las respuestas de los estudiantes en puntuaciones de rendimiento individual, garantizando la calidad psicométrica de los resultados obtenidos, se debieron realizar inicialmente algunos análisis exploratorios que permitieran verificar tanto la frecuencia como la distribución de cada alternativa de respuesta y de las respuestas omitidas, además de obtener una

[36] En el caso de los ítems de respuesta cerrada, esta técnica supone una complejidad añadida por el uso del modelo 3PL, en el que, a diferencia de los modelos 1PL (o Modelo Rasch) y 2PL, el parámetro de dificultad no representa directamente el valor en la escala de habilidad para el cual un estudiante cualquiera tiene 50% de probabilidad de responder correctamente. En los modelos de 1PL y 2PL, el cálculo de la dificultad reescalada se obtiene mediante una interpolación directa, mientras en el caso del modelo logístico 3PL es preciso utilizar una ecuación.

[37] En II medio no se realiza análisis DIF por ruralidad, debido a que la población de establecimientos rurales en ese nivel escolar es casi nula.

[38] Dado que el programa Parscale realiza una estandarización del parámetro de dificultad del grupo de referencia, este valor representa la diferencia en porcentajes de desviaciones estándar, en este caso, 40% de la desviación estándar del grupo de referencia.

primera aproximación del comportamiento de los datos, en cuanto al ajuste al modelo que se pretende utilizar. Para el caso de las preguntas cerradas se calcularon los siguientes indicadores:

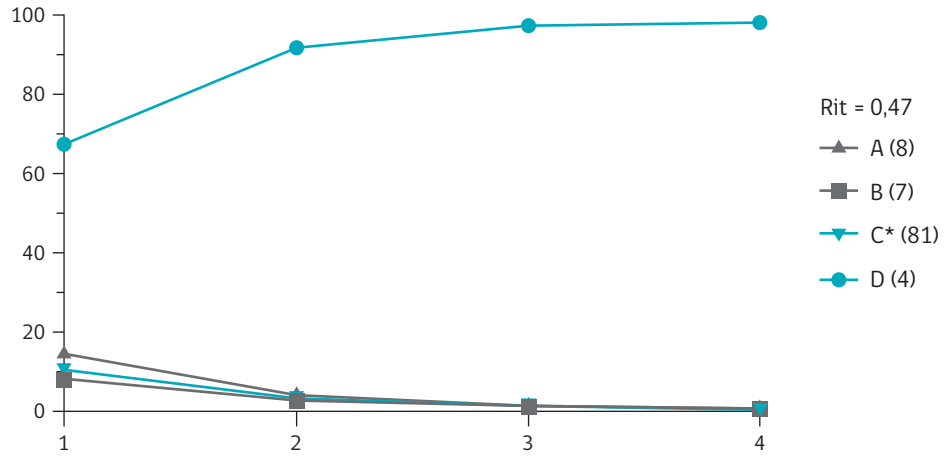
- Porcentaje de respuesta para cada alternativa.
- Correlación ítem-total de cada alternativa.
- Índice de dificultad del ítem (porcentaje de aciertos).
- Curvas empíricas de respuesta al ítem.
- Proporción de respuestas omitidas para cada ítem.

Luego, con el fin de optimizar el tiempo de procesamiento, se utilizaron las curvas empíricas generadas por el programa Tiaplus, para las subpoblaciones (estudiantes con discapacidad visual parcial y discapacidad auditiva[39]), cuya gráfica utiliza una curva para cada alternativa de respuesta señalando la que es correcta, como se muestra en la Figura 3.1.

Posteriormente, las puntuaciones de las pruebas Simce DS fueron calculadas en función del modelo de Teoría Clásica (por cuanto el reducido tamaño de las subpoblaciones de estudiantes que rindieron esta prueba no permite implementar IRT), en función del cual fue posible descomponer el puntaje observado del test como una combinación del puntaje verdadero más el componente de error. Esto implica que, a menor error, mayor es la relación entre el puntaje observado y el puntaje verdadero. Según este modelo, el puntaje verdadero de un sujeto está dado por V=X+e (Lord y Novick, 2008), es decir, el puntaje verdadero (V) se estima a través de una combinación lineal de los puntajes observados en la medición (X) más un componente de error (e).

Figura 3.1

Ejemplo de curva empírica del ítem



[39] Para los estudiantes con discapacidad visual total, no se generaron estas curvas porque en 2015 solo 7 estudiantes presentaron esa condición y esa cantidad no permite separar en subgrupos sus habilidades.

Una vez obtenida la estimación de habilidad individual, esta se transformó, utilizando una escala Simce de discapacidad sensorial[40], en la que se trabaja con los porcentajes de respuestas correctas (PRC) y se genera un rango de valores entre 0 y 100, donde cada puntuación se obtiene del siguiente modo:

$$PRC = \sum_{i=1}^{N} \frac{respuesta\ i \bullet 100}{N^o\ total\ preguntas}$$

Como se describió en el capítulo 1 de este informe, en los análisis de la prueba de Escritura 2015, aplicada a los estudiantes de 6° básico, se utilizaron únicamente preguntas de respuesta abierta, las que fueron corregidas con puntuaciones que van desde 0 a 6 (0 para respuestas con letra ilegible o para aquellos estudiantes que solo copiaron el estímulo; 1 a 5 para identificar distintos tipos de respuestas parcialmente correctas, y 6 para identificar las respuestas completamente correctas). Estas puntuaciones se aplicaron de manera independiente a cada una de las ocho habilidades evaluadas.

Para realizar el análisis psicométrico de esta prueba, también se utilizó la Teoría Clásica, debido a que, al igual que en Simce DS, no se cumplían todos los requisitos para el uso de IRT. El modelo clásico permitió construir un puntaje individual a partir de la suma de los puntajes obtenidos, y para el cálculo de ese puntaje, las respuestas omitidas fueron tratadas como perdidas, y con ello los estudiantes que omitieron toda la prueba no obtuvieron una estimación de puntaje.

Una vez obtenidas estas puntuaciones se realizó un proceso de equiparación de los puntajes de cada forma, a través de un método equipercentil (Kolen y Brennan, 2004), que permite obtener una escala única para las distintas formas de la prueba. La escala escogida para esta prueba fue de una media de 50 puntos y de una desviación estándar de 10 puntos.

Luego, para la estimación de los parámetros de la prueba, cálculo de confiabilidad y otros, se generaron los siguientes indicadores: (a) frecuencia en cada categoría (de los criterios de las rúbricas holísticas y analíticas) y correlación poliserial, y (b) confiabilidad por forma[41]. Junto con lo anterior, se calculó una matriz de correlación entre los siguientes puntajes: (i) puntaje obtenido en la pauta holística, (ii) puntaje obtenido en cada uno de los indicadores de la pauta analítica y (iii) puntaje obtenido en los ítems de cada uno de los objetivos de evaluación, por pauta.

Finalmente, en el caso de la prueba muestral de Educación Física, el objetivo del análisis fue construir indicadores que permitieran clasificar a los estudiantes según dos aspectos de la condición física, estructural y funcional, y de acuerdo a las categorías "satisfactorio" y "no satisfactorio". Los indicadores se evaluaron según edad y género.

[40] Esta escala es diferente a las usadas en las pruebas Simce de aplicación regular, por cuanto en estas pruebas DS no solo los ítems han sufrido acomodaciones específicas, sino que además la población está constituida por un reducido número de estudiantes.

[41] El detalle de la confiabilidad por forma se encuentra disponible, previa petición formal, en www.agenciaeducacion.cl, índice Simce/bases de datos nacionales 2015.

Para ello, a partir de los puntajes de corte obtenidos en cada dimensión evaluada[42], se clasificó la condición física de acuerdo a los aspectos funcionales y estructurales. Para establecer el indicador estructural se utilizaron las pruebas de abdominales cortos, salto largo a pies juntos, flexo-extensión de codos y flexión de tronco adelante. Para el indicador funcional se utilizaron los test de Cafra y Navette[43]. En cada uno de estos indicadores, los estudiantes fueron clasificados en la categoría "satisfactorio", cuando obtuvieron puntuaciones "aceptable" o "destacado" en todos los tests referidos al indicador.

3.3 Análisis de datos agregados

Este tipo de análisis permite obtener distintas agregaciones que son de utilidad tanto para informar la política pública, como para comunicar los resultados a los actores del sistema educativo (docentes, equipos directivos, estudiantes, etc.). Entre estas agregaciones se cuentan los promedios a nivel nacional, regional, comunal, etc.

Estos análisis tienen dos objetivos centrales: (a) conocer la realidad actual del desempeño de los estudiantes del país y (b) monitorear su evolución a través del tiempo, comparándolos con resultados de evaluaciones anteriores. Para lo primero se realizan comparaciones de rendimiento en un determinado año de aplicación, y para lo segundo, comparaciones entre años, como se describe a continuación.

- Diferencia significativa (en el año): compara el rendimiento promedio obtenido en una agregación con el rendimiento promedio de otra agregación en la misma medición. Por ejemplo, se comparan los puntajes promedio 4° básico Matemática Simce 2015 de una región con el puntaje promedio nacional.
- Variación significativa (entre años): compara el rendimiento promedio obtenido en una agregación con el rendimiento promedio de la misma agregación de la medición anterior. Por ejemplo, se comparan los puntajes promedio Simce 2014 y Simce 2015 en la evaluación de 4° básico Matemática.

En las pruebas que cuentan con Estándares de Aprendizaje, es posible, además, determinar la proporción de estudiantes de una agregación que se ubica en cada nivel: Adecuado, Elemental e Insuficiente.

Adicionalmente, los análisis agregados se realizan cruzando las variables de rendimiento (obtenidas de las pruebas) con otras variables, obtenidas de los cuestionarios, como la clasificación por GSE. Esta variable permite comparar el desempeño académico de una comunidad educativa respecto de otra, controlando las características socioeconómicas de cada una y obteniendo como resultado comparaciones más justas[44].

A continuación, se describen los análisis agregados realizados en las pruebas 2015 censales regulares y en las pruebas adicionales de Escritura, Educación Física y Simce DS.

[42] Los puntajes de corte permiten explicar el nivel de rendimiento que tienen los estudiantes según edad y género, y se definieron en 2010, en un trabajo conjunto entre profesionales del Mineduc y expertos del Instituto Nacional de Deportes (IND).

[43] La prueba de Navette solo la pueden rendir aquellos estudiantes que luego de haber rendido el test de Cafra tengan una frecuencia cardiaca inferior a 160 puls/min.

[44] Para la construcción de los grupos de establecimientos similares según GSE se utiliza una técnica estadística de análisis de conglomerados o clusters. La clasificación de las escuelas, según esta metodología, se hace por nivel evaluado, incluso cuando los establecimientos imparten todos los grados evaluados (véase en detalle el cálculo de GSE en Agencia de Calidad de la Educación, 2013).

3.3.1 Análisis agregados de las pruebas censales Simce

A partir de la información obtenida de los estudiantes que rindieron las pruebas regulares Simce 2015, se realizaron los siguientes análisis:

- Porcentaje de estudiantes evaluados, a nivel nacional y por región: número total de estudiantes que rindió la prueba, sobre el total de estudiantes perteneciente al nivel evaluado del sistema.
- Porcentaje de establecimientos evaluados, a nivel nacional y por región: número total de establecimientos donde se aplicó cada prueba, sobre el total de establecimientos del sistema que imparte el nivel escolar evaluado.
- Porcentaje de asistencia anual a nivel nacional: porcentaje promedio de asistencia que tuvieron todos los estudiantes del nivel escolar evaluado, durante el año escolar en curso.
- Porcentaje de asistencia Simce a nivel nacional: porcentaje de estudiantes que rinden la prueba sobre el total de estudiantes matriculados en el nivel escolar evaluado.
- Distribución de estudiantes evaluados: número de estudiantes evaluados por agregación. Se calcula para las agregaciones: nacional, regional, comunal, provincial, por departamento provincial, distrito, circunscripción, género, dependencia administrativa, GSE y GSE/dependencia.
- Distribución de establecimientos evaluados: número de establecimientos, por agregación, que contaron con estudiantes evaluados. Se calcula para las agregaciones: nacional, regional, comunal, provincial, departamento provincial, distrito, circunscripción, género, dependencia administrativa, GSE y GSE/dependencia.

Junto con lo anterior, se calculan el puntaje promedio y la distribución según Estándares de Aprendizaje, para cada asignatura en cada nivel escolar evaluado[45], según las agregaciones de interés.

Adicionalmente, tanto para la comparación de puntajes promedio como para la estimación de la proporción de estudiantes en cada nivel de los Estándares de Aprendizaje, se calcula la significancia estadística de las diferencias y variaciones entre agregaciones.

En la Tabla 3.7 se muestra una síntesis de los análisis agregados que se realizan, tanto a nivel de puntaje, como a nivel de estándares.

Tabla 3.7
Agrupación de análisis agregados de las pruebas regulares Simce 2015

Análisis	Descripción
Puntaje	Puntaje promedio de una agregación.
	Diferencia y significancia entre puntajes promedio de dos agregaciones en el mismo año.
	Variación y significancia entre puntajes promedio de una agregación obtenida en dos evaluaciones consecutivas.

[45] Es importante destacar que Simce reporta resultados por cada una de las asignaturas evaluadas en un determinado nivel escolar, porque cada asignatura tiene una escala propia, por lo que resulta técnicamente imposible generar análisis combinándolas.

Análisis	Descripción
Estándares de aprendizaje	Distribución en Estándares de Aprendizaje de una agregación.
	Diferencia y significancia de la distribución en Estándares de Aprendizaje de dos agregaciones en el mismo año.
	Variación y significancia entre la distribución en Estándares de Aprendizaje de una agregación obtenida en dos evaluaciones consecutivas.

Todos los análisis antes señalados se rigen por filtros y restricciones. Los filtros tienen por objeto definir la población estadísticamente comparable. Por ejemplo, al comparar los resultados Simce 2014 y 2015 en 4° básico Matemática, según GSE y dependencia administrativa, se considera solo a los establecimientos que en ambas evaluaciones mantuvieron su GSE y dependencia. Las restricciones, por su parte, determinan la publicación de datos, y por lo tanto son aplicadas de forma posterior al cálculo de los análisis agregados para resguardar la confidencialidad y robustez de los mismos. Por ejemplo, no se publican resultados según Estándares de Aprendizaje para aquellos establecimientos que hubieran tenido menos de diez estudiantes evaluados en una determinada asignatura y nivel escolar.

3.3.2 Análisis agregados de las pruebas adicionales Simce

Al igual que para los análisis psicométricos, existen diferencias entre los análisis agregados que se realizan para las pruebas regulares y los que se ejecutan sobre las pruebas adicionales.

Así, para la prueba Simce DS 2015, los análisis agregados reportaron el porcentaje promedio de respuestas correctas a nivel nacional para cada tipo de discapacidad: visual total, visual parcial y auditiva, en cada una de las pruebas: Lectura y Matemática. Este dato se entregó agregado también por tipo de establecimiento (regular/especial) y dependencia administrativa, por género de los estudiantes y por zona geográfica, distinguiendo tres zonas: Norte (regiones I, II, III, IV y XV), Centro (regiones V, RM, VI y VII) y Sur (regiones VIII, IX, X, XI, XII y XIV)[46].

En el caso de la prueba de Escritura, los análisis agregados fueron los puntajes promedio y los porcentajes de respuesta de los estudiantes en los diferentes niveles informados en las pautas de corrección.

Finalmente, en la prueba muestral de Educación Física, dado su carácter muestral, solo se reportaron resultados generales a nivel nacional. Estos fueron los siguientes:

- Distribución de estudiantes y establecimientos por región.
- Distribución de estudiantes, a nivel nacional, regional y por género, según las categorías de cada prueba.
- Distribución de estudiantes según género, región y dependencia administrativa, en la categoría de condición física "satisfactoria", desde el punto de vista estructural.
- Distribución de estudiantes según género, región y dependencia administrativa, en la categoría de condición física "satisfactoria", desde el punto de vista funcional.

[46] Para los análisis antes descritos se implementaron dos metodologías no paramétricas: la prueba U de Mann-Whitney, para las diferencias entre grupos independientes; y la prueba de Rangos Asignados de Wilcoxon, para la comparación de zona geográfica.

- Comparación de los porcentajes de condición física, desde el punto de vista estructural y funcional, con respecto a la evaluación anterior.
- Comparación de los porcentajes de condición física, desde el punto de vista estructural y funcional de cada agregación de interés, con respecto al resultado nacional de condición física en la actual evaluación.

3.3.3 Análisis de tendencias

Como se mencionó al comienzo de este apartado, junto con conocer la realidad actual del desempeño de los estudiantes del país, los análisis agregados buscan entregar información que permita monitorear su evolución a través del tiempo, comparándolos con resultados de evaluaciones anteriores.

En función de lo anterior, se realizan análisis que reportan las tendencias de resultados a nivel nacional cada año, para los distintos niveles escolares evaluados, en las diversas asignaturas medidas. Estos análisis permiten observar la evolución de puntajes, a partir de una medición de referencia, y son susceptibles de realizar desde la implementación del modelo IRT, en 1998.

Hacer este seguimiento implica controlar las características de los establecimientos (que pueden cambiar en el tiempo) con el objetivo de obtener poblaciones equivalentes que puedan ser comparadas. Al respecto, en el año 2015 la población fue distinta a la de la anterior medición (2014) y, además, aumentó el número de escuelas pequeñas (con menos de 10 estudiantes). Esto obligó a determinar una población de referencia, cuyas características se mantuvieran en el tiempo, de manera que fuera posible medir la evolución o tendencia de dicha población. Para obtener esta población comparable se construyó una muestra representativa, mediante muestreo aleatorio de conglomerados con afijación proporcional. Esto garantizó que la muestra seleccionada representara adecuadamente a la población en función de variables particulares denominadas estratificadoras. Estas variables fueron seleccionadas para reportar tendencias relativas a GSE (bajo, medio bajo, medio, medio alto y alto) y tamaño del establecimiento, medido en cantidad de cursos por nivel escolar evaluado, según las categorías que se muestran en la Tabla 3.8[47].

Tabla 3.8

Clasificación de establecimientos por tamaño en la aplicación 2015

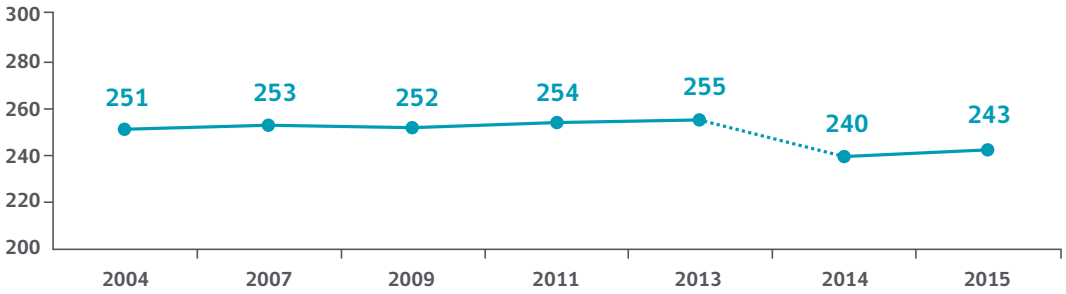
Tamaño del establecimiento	2º básico		4º básico		8º básico	
	N.º cursos	N.º estudiantes	N.º cursos	Nº estudiantes	N.º cursos	N.º estudiantes
1	1	n < 4	1	n < 4	1	n ≤ 20
2	1	4 ≤ n < 11	1	4 ≤ n < 11	1	n > 20
3	1	11 ≤ n	1	11 ≤ n	2	no aplica
4	2	no aplica	2	no aplica	3	no aplica
5	> 2	no aplica	> 2	no aplica	> 3	no aplica

[47] La unidad de muestra correspondió a los establecimientos educacionales, ya que estos son las unidades susceptibles de ser muestreadas. De este modo, se trabajó con todos los estudiantes evaluados de cada establecimiento, lo que permitió obtener un error muestral acotado.

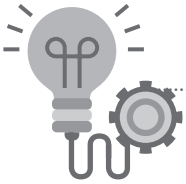
Una vez definida la muestra del año para cada nivel escolar evaluado, se calculó el puntaje promedio nacional por asignatura y la información se plasmó en gráficos que permitieran observar la tendencia a través del tiempo (véase un ejemplo de estos gráficos en la Figura 3.2).

Figura 3.2

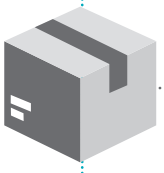
Ejemplo de gráfico de tendencia de puntaje a través del tiempo



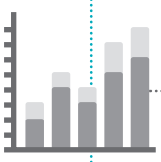
Los datos, gráficos y curvas que sustentan el presente capítulo, están como contenido descargable en la sección Informes Técnicos Simce del sitio web de la Agencia.



Capítulo I:
Diseño y construcción de las pruebas Simce



Capítulo II:
Logística y aplicación de las pruebas Simce



Capítulo III:
Procesamiento y análisis de datos

Capítulo IV:
Comunicación de resultados educativos

Capítulo IV: Comunicación de resultados educativos

En atención a que la Ley que establece el Sistema Nacional de Aseguramiento de la Calidad de la Educación señala que la Agencia de Calidad debe "proporcionar información en materias de su competencia a la comunidad en general y promover su correcto uso" (Ley N.° 20529, art. 10, letra e), la comunicación de los resultados Simce constituye uno de sus mandatos centrales.

Al respecto, la Agencia ha establecido los siguientes objetivos estratégicos[48]:

- Proveer información útil y orientaciones para el trabajo pedagógico e institucional en los establecimientos, con foco en el aprendizaje de los estudiantes.
- Entregar orientaciones técnicas, pedagógicas y de gestión a todos los establecimientos, para instalar capacidades en las comunidades escolares orientadas al mejoramiento de los procesos institucionales y los resultados educativos.
- Entregar información para la toma de decisiones y la articulación de los distintos actores del sistema escolar, así como impulsar el trabajo colaborativo y en red.
- Contribuir al sistema a partir de su carácter técnico, basado en evidencia, independiente, transfiriendo la información generada a partir de las evaluaciones y estudios que realiza la Agencia, para influir en la toma de decisiones en materia de política educativa, identificando fortalezas y debilidades de los procesos y resultados educacionales.

Para cumplir en parte con estos objetivos, la Agencia realiza una serie de publicaciones asociadas a las aplicaciones Simce que buscan informar y orientar sobre el uso de sus resultados a los distintos actores del sistema educativo, entre los que se cuentan: docentes, directivos, sostenedores de establecimientos, padres y apoderados, líderes en política educativa e investigadores.

Los docentes, directivos y sostenedores reciben información sobre los resultados de las pruebas y de los cuestionarios, así como orientaciones para el uso pedagógico de la información, de manera de impulsar la planificación de acciones concretas que aporten al proceso de mejora continua de la calidad de la educación impartida. A los padres y apoderados, por su parte, se les entrega información sintetizada y organizada de manera sencilla, para facilitar la comprensión por parte de todos y de ese modo, favorecer su interés por involucrarse de manera directa en el sistema. Finalmente, los líderes en política educativa y los investigadores, aunque tienen acceso a la información general que se publica a nivel nacional, también pueden acceder, si lo solicitan formalmente, a información específica (que incluye bases de datos) sobre los resultados de una o más pruebas Simce y sobre los Indicadores de Desarrollo Personal y Social. Lo anterior, con el objeto de que puedan desarrollar sus proyectos contando con información de la fuente directa.

En este capítulo se describen los principales documentos comunicacionales que publicó la Agencia en el contexto de las pruebas Simce 2015 regulares y la prueba muestral de Educación Física.

[48] Disponibles en www.agenciaeducacion.cl

4.1 Documentos informativos previos a la aplicación

Antes de la aplicación de pruebas Simce 2015, la Agencia publicó tres documentos generales que tuvieron como propósito entregar información relevante sobre la evaluación: Programas evaluaciones de aprendizaje para docentes, Programas aplicación evaluaciones educativas para directivos y Programa Estudio Nacional de Educación Física 8° básico para directivos[49].

El documento Programas evaluaciones de aprendizaje para docentes (que se entregó por separado para educación básica y educación media) tuvo como propósito que los profesores conocieran las características generales de las pruebas, el cronograma de aplicación y los conocimientos y habilidades que se evaluarían, los que se comunicaron en referencia al Currículo Nacional.

Los Programas de aplicación evaluaciones de aprendizaje para directivos (para educación básica, educación media, pruebas experimentales y escuelas especiales), tuvieron como propósito informar a los directores sobre el proceso y favorecer su colaboración en los días de aplicación, a lo que le correspondía a su función. Estos documentos incluyeron el cronograma de aplicación de las pruebas e información específica sobre los estudiantes que rendirían las evaluaciones (incluyendo a aquellos estudiantes que podrían no rendirlas por presentar algún impedimento) y sobre los procedimientos para los cuales se requeriría la colaboración de los directivos del establecimiento.

Finalmente, el Programa Estudio Nacional de Educación Física 8° básico para directivos se publicó con el objeto de informar sobre las características de los test y los procedimientos a realizar antes, durante y después de la aplicación de la prueba.

4.2 Informes de resultados

Después de la aplicación de pruebas Simce 2015, la Agencia publicó informes de resultados conteniendo tanto los resultados de aprendizaje Simce como los resultados referidos a los Indicadores de Desarrollo Personal y Social. Ello, con el propósito de propiciar la realización de diagnósticos más completos e integrales de los aprendizajes de los estudiantes.

Adicionalmente, en los informes de resultados de 4° y 6° básico se incluyeron por primera vez ejemplos de preguntas de las distintas asignaturas evaluadas, los que se presentaron con comentarios y ordenados en un mapa, según dificultad. En el caso de las pruebas de Matemática de estos niveles escolares, se presentaron resultados según ejes temáticos, con el propósito de entregar información pedagógica específica para los docentes.

Los informes de resultados contemplan tres versiones: una para docentes y directivos; una para padres y apoderados; y una para sostenedores[50].

[49] También se publicaron programas de aplicación evaluaciones de aprendizaje para directivos de escuelas especiales.

[50] Estos informes se evalúan cada año por las audiencias respectivas, con el objeto de mejorar la calidad de los productos comunicacionales y fomentar el uso adecuado de la información. En estos procesos de evaluación se determina por ejemplo la necesidad de incorporar nueva información, de cambiar el enfoque o de profundizar en ciertos contenidos.

El Informe de Resultados para Docentes y Directivos tiene por objetivo comunicar los resultados de aprendizaje del conjunto de estudiantes del establecimiento, en las distintas asignaturas y niveles escolares evaluados, y entregar orientaciones para una correcta interpretación y uso de los mismos.

En este documento se entrega la siguiente información específica:

- Puntajes promedio de todos los estudiantes, en cada una de las pruebas.
- Variaciones de puntaje promedio del establecimiento, respecto del año anterior.
- Tendencias de puntaje promedio del establecimiento, a través del tiempo.
- Variaciones de puntaje promedio del establecimiento, en relación con otros establecimientos de similares características socioeconómicas.
- Promedio de puntaje por curso.
- Porcentaje de estudiantes en cada uno de los niveles de desempeño de los Estándares de Aprendizaje y tendencia de distribución de estudiantes en los distintos niveles, en las últimas tres mediciones (solo para aquellas asignaturas y niveles escolares para los que existen estándares).
- Puntaje en cada Indicador de Desarrollo Personal y Social y sus dimensiones.
- Puntaje promedio en los ejes temáticos de Matemática (a nivel de establecimiento y curso).

Adicionalmente, en el año 2015 estos informes incluyeron observaciones sobre gestión escolar, para los directivos; y sobre gestión pedagógica, para los docentes. Para los directivos se presentó una propuesta de plan de trabajo para analizar los resultados Simce y relacionar estos resultados con datos internos de gestión, con el fin de contextualizarlos en la realidad del establecimiento. Con esto se buscó que la información entregada permitiera impulsar acciones conducentes a construir trayectorias de mejora de los aprendizajes y, en particular, que sirviera de base para evaluar los planes de mejoramiento educativo y los planes estratégicos de los establecimientos. Para los docentes, por su parte, se presentó una propuesta de desarrollo de un ciclo de evaluación que no termina con la aplicación del instrumento, sino que se inicia con ese hito, y que incorpora una fase de análisis e interpretación de resultados, el diseño de propuestas de mejora, y la implementación, monitoreo y evaluación de los logros de aprendizaje. Junto con lo anterior, se presentaron recomendaciones para perfeccionar el proceso de evaluación en el aula. Esta información tuvo como propósito que los docentes pudieran evaluar sus prácticas pedagógicas al interior del aula y analizar la diversidad de aprendizajes de su grupo de estudiantes.

El Informe de Resultados para Padres y Apoderados tiene por objeto informar de los resultados a las familias de los estudiantes para involucrar a los padres y apoderados en el proceso educativo y convocarlos a participar activamente de él. Este informe incluye resultados de cada Indicador de Desarrollo Personal y Social (escala de 0 a 100 puntos) comparados con otros establecimientos de similares características socioeconómicas; resultados de aprendizaje Simce comparados con establecimientos de similares características socioeconómicas; y distribución de los estudiantes en niveles de desempeño, en las asignaturas evaluadas que cuentan con Estándares de Aprendizaje. Junto con lo anterior, se entrega información general sobre las funciones de la Agencia y recomendaciones para participar activamente del proceso de aprendizaje de sus pupilos.

Finalmente, el Informe de Resultados para Sostenedores contiene los principales resultados educativos, tanto a nivel general de la administración como el detalle de cada uno de los establecimientos. Además, se incluyen orientaciones para el correcto análisis y uso de la información.

4.3 Otros documentos comunicacionales

Junto con los informes antes descritos, después de la aplicación 2015 la Agencia publicó una serie de documentos que son susceptibles de ser clasificados en tres grupos: resúmenes ejecutivos, material de apoyo pedagógico y bases de datos.

Los **resúmenes ejecutivos** tienen el propósito de entregar una síntesis de los principales resultados de las evaluaciones Simce a la audiencia general, a través de los medios de comunicación masivos, con el objeto de informar y transparentar los procesos, además de propiciar el análisis y debate en torno a la educación del país.

Para la aplicación 2015 se publicaron tres resúmenes ejecutivos que se describen a continuación:

- **Resumen ejecutivo para la conferencia de prensa nacional.** Este consistió en la presentación de los objetivos estratégicos de la Agencia, de los resultados de aprendizaje y de indicadores de desarrollo de la aplicación Simce 2015, además de algunas conclusiones generales obtenidas a partir de dichos resultados. Este resumen ejecutivo se utilizó en un total de 4 conferencias de prensa realizadas en los meses de abril, junio, septiembre y octubre de 2016. Los videos de estas conferencias fueron transmitidos vía *streaming* y publicados en el sitio web de la Agencia.
- **Resúmenes ejecutivos para las conferencias de prensa regionales.** Estos documentos, uno por región, contienen una síntesis de los principales resultados Simce obtenidos por los estudiantes de la región. El público objetivo de estos documentos informativos son los Secretarios Regionales Ministeriales y los Jefes de los Departamentos Provinciales de Educación, a quienes se les envía la información por correo directo y luego se publica en el sitio web de la Agencia.
- **Presentación de resultados educativos para directores.** Junto con la entrega del Informe de Resultados para Docentes y Directivos, desde 2014 se entregó adicionalmente este documento que resume los principales resultados de cada establecimiento (puntaje promedio, distribución de estudiantes en cada nivel de aprendizaje, tendencias de resultados, puntaje promedio por curso evaluado y comparación con establecimientos similares) con el objeto de facilitar a los directores la presentación de los resultados a su equipo directivo, a los docentes y a la comunidad en general. Esta síntesis se envió por correo electrónico personal a los directores de todos los establecimientos evaluados, y luego quedó disponible para descargar en el sitio web de la Agencia, ingresando la clave única de acceso digital a la plataforma SIGE.

En cuanto al **material de apoyo pedagógico**, para la aplicación 2015 se publicó el taller Uso de los resultados educativos para la mejora de los aprendizajes, consistente en un archivo (en formato ppt) que sirvió de orientación para la realización de talleres para docentes y directivos, y una sección web referida a la prueba de Escritura, en la que se publicaron ejemplos de preguntas y de respuestas de los estudiantes, las pautas de corrección utilizadas y recomendaciones pedagógicas para los docentes.

Finalmente, en relación con las **bases de datos**, cabe señalar que la Agencia provee de dos tipos de bases de datos: una pública y una privada, para investigadores. La base de datos pública contiene los resultados por establecimiento, comuna, provincia, región, GSE y dependencia administrativa de los establecimientos. La versión 2015 de esta base se encuentra a disposición de la comunidad desde noviembre de 2016, en el sitio web de la Agencia. La base de datos para investigadores, por su parte, contiene datos específicos relativos a los resultados, pero no contiene individualización de estudiantes, profesores ni apoderados. A esta solo se puede acceder mediante solicitud formal a la Agencia, previa acreditación de los fines del estudio.

La difusión de todos los documentos descritos en este capítulo, referidos a la aplicación 2015, cuyo detalle se muestra en la Tabla 4.1, se realizó a través de medios impresos y plataformas digitales. Entre estas

últimas, se cuentan el sitio web de la Agencia, las redes sociales oficiales (Twitter y Facebook de la Agencia), correos electrónicos masivos y personales, y plataforma *streaming*, en el caso de las videoconferencias. Esta socialización mediante diversos medios de comunicación y difusión, busca asegurar un acceso efectivo y expedito a la información, por parte de los distintos actores del sistema educativo.

Tabla 4.1

Documentos comunicacionales de la aplicación Simce 2015

Tipo de documento	Título de la publicación	Soporte(s)	Fecha de la publicación/difusión
Programas evaluaciones de aprendizaje	Programas Evaluaciones de Aprendizaje para Docentes de Enseñanza Básica y Media.	digital	marzo 2015
	Programas de Aplicación Evaluaciones de Aprendizaje para Directivos de Enseñanza Básica y Media.	digital	agosto2015
	Programa Estudio Nacional de Educación Física 8° Básico para Directivos.	impreso	agosto 2015
	Programas de Aplicación Evaluaciones de Aprendizaje para Directivos de Escuelas Especiales 6° Básico.	impreso	agosto 2015
Informes de resultados de aprendizaje	Informe de Resultados Educativos de 2°, 4° y 6° Básico para Docentes y Directivos.	impreso/digital	abril 2016
	Informe de Resultados Educativos de 8° Básico y II Medio para Docentes y Directivos.	impreso/digital	junio 2016
	Informe de Resultados de Aprendizaje Escritura 6° Básico.	impreso/digital	septiembre 2016
	Informe de Resultados de Aprendizaje 2° Básico para Padres y Apoderados.	impreso/digital	abril 2016
	Informe de Resultados Educativos 4° Básico para Padres y Apoderados.	impreso/digital	abril 2016
	Informe de Resultados Educativos 6° Básico para Padres y Apoderados.	impreso/digital	abril 2016
	Informe de Resultados Educativos 8° Básico para Padres y Apoderados.	impreso/digital	junio 2016
	Informe de Resultados Educativos II Medio para Padres y Apoderados.	impreso/digital	junio 2016
	Informe de Resultados de Aprendizaje 6° Básico Escritura para Padres y Apoderados.	impreso/digital	septiembre 2016
	Informe de Resultados Educativos para Sostenedores de 2°, 4° y 6 ° Básico.	impreso/digital	abril 2016
	Informe de Resultados Educativos para Sostenedores de 8° Básico y II Medio.	impreso/digital	junio 2016
	Informe de Resultados Estudio Nacional de Educación Física.	impreso/digital	octubre 2016
Otros documentos comunicacionales	PPT Directores (2°, 4° y 6 ° Básico).	impreso/digital	abril 2016
	PPT Directores (8° Básico y II Medio).	impreso/digital	junio 2016

Anexos

Referencias bibliográficas

Anexo A: Pruebas experimentales Simce 2015

Las pruebas experimentales tienen un doble objetivo: en el caso de las preguntas cerradas, tienen como fin la obtención de datos psicométricos (desde la Teoría Clásica e IRT) que sirvan de insumo para seleccionar los ítems que integrarán las pruebas censales del año siguiente. En el caso de las preguntas abiertas, su corrección permite adicionalmente obtener información para validar las rúbricas y elaborar los manuales de corrección, proceso que se describe de manera general al final de este anexo.

Las pruebas experimentales son de carácter muestral y su aplicación se realiza en conjunto con las pruebas censales y demás instrumentos aplicados cada año y en los mismos niveles escolares.

Las pruebas experimentales que sirvieron de base para las censales 2015 (PE2015) se aplicaron en el año 2014 y correspondieron a las pruebas de Lectura aplicadas en todos los niveles evaluados (2°, 4°, 6° y 8° básico y II medio); de Matemática, en todos los niveles a excepción de 2° básico; de Ciencias Naturales en 8° básico[51]; de Escritura en 6° básico, y de Historia, Geografía y Ciencias Sociales, en 6° básico y II medio.

En la aplicación de las PE2015 se consideró una **muestra de estudiantes** de los niveles evaluados, pertenecientes a establecimientos de las regiones de Valparaíso, Metropolitana y del Biobío[52]. Para la determinación de esta muestra se contemplaron los mismos métodos descritos para la aplicación muestral de Educación Física 2015 (contenidos en el capítulo 2 de este informe), con la salvedad de que se excluyeron todos los establecimientos en los que se estuviera llevando a cabo un estudio paralelo en el mismo nivel escolar, para evitar una posible sobrecarga de evaluaciones.

Para la definición de la muestra de las PE2015 se establecieron **estratos**[53]. En los niveles de enseñanza básica estos estratos se conformaron según las siguientes variables: dependencia (municipal/particular subvencionado/particular pagado), ruralidad (urbano/ rural) y rendimiento Simce. Esta última variable se dividió en tres categorías (cuyos puntos de corte están dados por el promedio ±0,5 desviaciones estándar, respecto del grado y área correspondiente), para asegurar la participación de establecimientos de estudiantes de distinto rendimiento. Las muestras fueron calculadas por prueba y nivel escolar, según se muestra en la Tabla A1.

[51] Las PE2015 también contemplaron la aplicación de una prueba de Ciencias Naturales en 4° básico para servir de base a la construcción de la prueba definitiva de ese nivel escolar que se aplicaría en años posteriores, en función del Plan de Evaluaciones.

[52] No se seleccionó una muestra representativa de las distintas regiones, por cuanto la aplicación de la metodología IRT no lo requiere para analizar los datos.

[53] En caso de que en la formación de estratos se detectara una cantidad considerable de establecimientos nuevos (alrededor de 100 casos) que no contaran con resultados Simce de la medición anterior, se procedió a crear un estrato adicional que los agrupara. Además, se procuró que todos los estratos tuvieran una representación de al menos 1%. En caso contrario, estos últimos fueron asociados al estrato más afín.

Tabla A1

Categorías de rendimiento Simce por prueba y nivel escolar en enseñanza básica

Nivel escolar	Prueba	Categoría 1	Categoría 2	Categoría 3
2º básico	Comprensión de Lectura	X ≤ 229 puntos	229 < X ≤ 279 puntos	X > 279 puntos
4º básico	Comprensión de Lectura	X ≤ 239 puntos	239 < X ≤ 290 puntos	X > 290 puntos
	Matemática	X ≤ 230 puntos	230 < X ≤ 282 puntos	X > 282 puntos
	Ciencias Naturales	X ≤ 233 puntos	233 < X ≤ 278 puntos	X > 278 puntos
6º básico	Comprensión de Lectura	X ≤ 225 puntos	225 < X ≤ 275 puntos	X > 275 puntos
	Matemática	X ≤ 225 puntos	225 < X ≤ 275 puntos	X > 275 puntos
	Historia, Geo. y Cs. Sociales[a]	X ≤ 234 puntos	234 < X ≤ 283 puntos	X > 283 puntos
	Escritura	X ≤ 145 puntos	145 < X ≤ 155 puntos	X > 155 puntos
8º básico	Comprensión de Lectura	X ≤ 230 puntos	230 < X ≤ 280 puntos	X > 280 puntos
	Matemática	X ≤ 238 puntos	238 < X ≤ 286 puntos	X > 286 puntos
	Ciencias Naturales	X ≤ 248 puntos	248 < X ≤ 296 puntos	X > 296 puntos

Nota: [a] Para esta prueba se consideró el rendimiento de los establecimientos de 4º básico 2012 en esta asignatura.

Para las pruebas de II medio los estratos se conformaron con las siguientes variables: dependencia administrativa (municipal/particular subvencionado/particular pagado), modalidad educacional (científico-humanista/técnico-profesional/polivalente) y rendimiento Simce, dividido en tres categorías, según se muestra en la Tabla A2.

Tabla A2

Categorías de rendimiento Simce por prueba en II medio

Nivel escolar	Prueba	Categoría 1	Categoría 2	Categoría 3
II medio	Comprensión de Lectura	X ≤ 226 puntos	226 < X ≤ 282 puntos	X > 282 puntos
	Matemática	X ≤ 235 puntos	235 < X ≤ 300 puntos	X > 300 puntos
	Historia, Geo. y Cs. Sociales[a]	X ≤ 235 puntos	235 < X ≤ 283 puntos	X > 283 puntos

Nota: [a] Para esta prueba se consideró el rendimiento de los establecimientos de 8º básico 2011 en esta asignatura.

Respecto del **tamaño de la muestra**, cabe señalar que para la obtención de una estimación robusta de acuerdo al modelo utilizado (IRT de tres parámetros[54]), se requiere contar con un mínimo de 1.500 respuestas efectivas para cada pregunta; sin embargo, para las PE2015 se estimó un mínimo de 2.500 estudiantes por pregunta, considerando la amplia gama de análisis requeridos y la tasa de omisión promedio en este tipo de evaluaciones (menor a 5%). A partir de este dato se determinó el número de establecimientos que compondrían la muestra, la cantidad de formas de prueba de cada evaluación y la tasa de estudiantes por establecimiento.

[54] Existe variada información con respecto al mínimo necesario para estimaciones robustas. El valor más alto encontrado para un modelo logístico de tres parámetros es de entre 1.000 y 1.500 casos por ítem (Murphy y Elliot, 1991 y Ricker, 2006).

La **cobertura efectiva** que tuvo la aplicación de las PE2015 fue de 189.496 estudiantes, distribuidos por nivel escolar, como se muestra en la Tabla A3.

Tabla A3

Estudiantes según nivel escolar en las PE2015 y porcentaje de la muestra teórica que representan

Nivel escolar	N.º de evaluados	% de la muestra teórica
2º básico	22.291	88,4%
4º básico	45.783	90,2%
6º básico	48.009	85,5%
8º básico	43.057	85,6%
II medio	30.356	75,4%

La información sobre las preguntas que conformaron cada una de las PE2015 por asignatura y nivel escolar, se encuentra en las planillas maestras disponibles, previa petición formal, en *www.agenciaeducacion.cl*, índice Simce/bases de datos nacionales 2015.

El tipo de información que se encuentra en estas planillas corresponde a la identificación y a los datos métricos de los ítems, según se detalla en las tablas A4 y A5.

Respecto de las **preguntas abiertas** de pruebas experimentales, como se señaló al comienzo, el proceso de corrección permite validar las rúbricas y elaborar los manuales de corrección. Este proceso se inicia con la revisión de las preguntas a corregir, que consiste en un análisis cualitativo de las preguntas, a partir del cual se eliminan aquellas que presentan problemas, basándose fundamentalmente en el registro de consultas realizadas por los estudiantes durante la aplicación.

Una vez definidas las preguntas a corregir, se selecciona una muestra de respuestas con el objeto de ajustar las rúbricas preliminares[55] y los respectivos manuales de corrección. En la PE2015 la muestra constó de 500 respuestas de cada una de las preguntas abiertas seleccionadas.

En el proceso de ajuste de rúbricas y manuales, participan profesionales expertos de las distintas asignaturas, tanto externos como pertenecientes a la Agencia.

El proceso se inicia con la corrección de 200 respuestas, utilizando la rúbrica preliminar. En función de este proceso, se ajustan las rúbricas y se corrigen las restantes 300 respuestas, con el objeto de validarla. El resultado de esta fase del proceso es un Manual de Corrección ajustado que incluye la rúbrica definitiva, que identifica las distintas categorías de respuesta en función de códigos y ejemplos reales de respuestas de los estudiantes comentados, que permitirán orientar de mejor forma los procesos de corrección de la prueba experimental. Adicionalmente, el manual incluye una serie de ejercicios prácticos para que los correctores puedan ejercitar con antelación.

[55] Durante el proceso de elaboración de las preguntas abiertas se genera una rúbrica preliminar basada en información teórica sobre las respuestas esperadas. Estas rúbricas suelen incluir desde las respuestas más completas y correctas posibles, hasta las más erróneas e incompletas. Luego se establece una escala descriptiva (asociada a una puntuación) y, a partir de esta, se delimitan dos categorías (correcto e incorrecto) o tres categorías (correcto, incorrecto y parcialmente correcto).

En una segunda etapa, se selecciona y capacita a los correctores y supervisores que participarán en el proceso, con el objeto de asegurar que posean un dominio acabado del protocolo de corrección, de la rúbrica, y de los demás procedimientos necesarios para resguardar la consistencia intercorrectores.

En una tercera etapa, se realiza una marcha blanca, durante la cual se corrige 10% del total de las respuestas obtenidas. En este periodo se incluyen, en caso de ser necesario, precisiones al Manual de Corrección.

Finalmente, se corrigen todas las respuestas obtenidas de la aplicación experimental, realizando doble corrección de 50% de las respuestas, con el objeto de calcular la consistencia intercorrectores. En función del grado de consistencia se evalúa si el Manual posee la suficiente calidad para ser utilizado en una aplicación censal[56]. En las PE2015, para la corrección de preguntas abiertas, se estableció un requerimiento de consistencia igual o superior a 80%. Este porcentaje no solo aseguró validez de la corrección, sino también un adecuado grado de productividad (rapidez con la que se corrigen las respuestas), pues a mayor consistencia, mayor productividad. Adicionalmente, los supervisores corrigieron 10% de las respuestas asignadas a su grupo de correctores y 25% de las respuestas doblemente corregidas por su grupo.

Tabla A4

Código y descripción de los campos de identificación de los ítems de las PE2015

	Código del campo	Descripción de campo
Identificación	ÁREA	Área a la que pertenece el ítem (todos los ítems).
	ID_COD	Código identificador del ítem (todos los ítems).
	ID_TIPO	Tipo de ítem (todos los ítems).
	POS_PD	Posición en la prueba definitiva de aplicación anterior.
	POS_EXP	Posición en la PE2015.
	EQ_PD_AÑOS_2015	Ítems de *equating* entre años de la prueba definitiva 2015.
	EQ_PD_FORMAS_2015	Ítems de *equating* entre formas de la prueba definitiva 2015.
	POS_PD_2015	Posición(es) en la prueba definitiva 2015.
	CLAVE	Opción de respuesta correcta.
	ALT_VAL	Cantidad de alternativas válidas.
	CE_MCNB	Nivel escolar al que corresponde el desempeño y/o tarea evaluada.
	CE_OA	Objetivo de aprendizaje (de acuerdo al currículo vigente).
	CE_OBJ_EV	Objetivo de evaluación de las especificaciones técnicas.
	CE_EJE_TEMÁTICO_PD_2015	Eje Temático de la prueba definitiva 2015 según currículo[a].
	CE_DOMINIO_COGNITIVO_PD_2015	Dominio Cognitivo de la prueba definitiva 2015, según currículo[b].
	CE_TÍTULO_ESTIMULO COMÚN	Estímulo al que se asocia el ítem.

Nota: [a] En la prueba de Lectura corresponde a tipo de texto.
[b] En la prueba de Lectura corresponde a habilidades.

[56] En caso de detectarse falencias, se convoca a profesionales que no hayan participado del proceso previo y se realiza una corrección adicional de 300 respuestas, con el objeto de consolidar el Manual definitivo.

Tabla A5

Código y descripción de los campos de datos métricos de los ítems de las PE2015

	Código del campo	Descripción de campo
DATOS MÉTRICOS PREGUNTAS OPCIÓN MÚLTIPLE	%_A_2015	Porcentaje de respuestas de la alternativa A.
	%_B_2015	Porcentaje de respuestas de la alternativa B.
	%_C_2015	Porcentaje de respuestas de la alternativa C.
	%_D_2015	Porcentaje de respuestas de la alternativa D.
	%_OMIT_2015	Porcentaje de respuestas omitidas.
	CORR_A_2015	Correlación opción-test considerando la alternativa A.
	CORR_B_2015	Correlación opción-test considerando la alternativa B.
	CORR_C_2015	Correlación opción-test considerando la alternativa C.
	CORR_D_2015	Correlación opción-test considerando la alternativa D.
	DISC_IRT_2015	Parámetro de discriminación IRT.
	DIFIC_IRT_2015	Parámetro de dificultad IRT.
	AZAR_IRT_2015	Parámetro de azar.
	RP67_2015	Parámetro de dificultad en escala Simce[a].
	GENERO_2015	Indicador de funcionamiento diferencial del ítem (género favorecido).
	RURALIDAD_2015	Indicador de funcionamiento diferencial del ítem (zona favorecida).
DATOS MÉTRICOS PREGUNTAS ABIERTAS	%_PARC	Porcentaje de respuesta para la categoría 'parcialmente correcta'.
	%_CORR	Porcentaje de respuesta para la categoría 'correcta'.
	%_OMITA	Porcentaje de respuestas omitidas.
	CORR_A	Correlación opción-test considerando categoría con puntuación máxima.
	DISC_A_IRT	Parámetro de discriminación IRT.
	DIFIC_PARC_IRT	Parámetro de dificultad IRT para la categoría 'parcialmente correcta'.
	DIFIC_CORR_IRT	Parámetro de dificultad IRT para la categoría 'correcta'.
	RP67_PARC	Parámetro de dificultad en escala Simce para la categoría 'parcialmente correcta[b]'.
	RP67_CORR	Parámetro de dificultad en escala Simce para la categoría 'correcta[c]'.
	GENERO_A	Indicador de funcionamiento diferencial del ítem (género favorecido).
	RURALIDAD_A	Indicador de funcionamiento diferencial del ítem (zona favorecida).

Nota: [a b c] Ajustado a una probabilidad de respuesta de 67%

Anexo B: Corrección de preguntas de respuesta abierta

Como se señaló en el primer capítulo de este informe, las preguntas de respuesta abierta (o preguntas abiertas) se utilizan fundamentalmente para medir aprendizajes que no son susceptibles de ser evaluados con formato cerrado o cuyo constructo se mide en forma más precisa en formato abierto. Sin embargo, este tipo de preguntas requiere un mayor tiempo de corrección que el de las de opción múltiple y un proceso que supone contar con equipos de correctores alineados con los criterios de evaluación y consistentes entre sí, para mantener la validez de la corrección (Brancato y otros, 2006).

En efecto, la corrección de preguntas abiertas se realiza mediante rúbricas aplicadas por correctores, por lo que se requiere establecer mecanismos que aseguren la confiabilidad de la corrección de las respuestas. Entre estos mecanismos se cuentan: una capacitación de correctores, previa al inicio del proceso, con el objeto de lograr alineación con los criterios de evaluación, y un monitoreo constante, durante la corrección, de los porcentajes de consistencia entre correctores[57] (consistencia intercorrectores), en función de estándares previamente establecidos.

El proceso Simce 2015 requirió la corrección de preguntas de respuesta abierta en todas las pruebas censales aplicadas y, al igual que en años anteriores, las respuestas de los estudiantes a las preguntas abiertas fueron dispuestas para su corrección en formato de imágenes digitalizadas (las cuales fueron obtenidas en el proceso de captura, según se reseñó en el capítulo 3 de este informe), y para su corrección se utilizó el *software* Corrector 1.0.

En términos de los grupos de trabajo, el proceso de corrección 2015 se realizó conformando equipos expertos de las distintas disciplinas evaluadas. Administrativamente, este proceso se dividió en dos fases: corrección de preguntas de las pruebas de 2°, 4° y 6° básico, cuyos correctores fueron contratados directamente por la Agencia para prestar este servicio, y corrección de preguntas de las pruebas censales de 8° básico y II medio, cuya ejecución fue licitada a un proveedor externo, con supervisión de profesionales de la Agencia.

En la Tabla B1 se muestra la cantidad total de correctores que participó del proceso 2015 y la cantidad de grupos de corrección conformados para cada asignatura y nivel escolar.

Tabla B1

Distribución de correctores de preguntas abiertas pruebas censales Simce 2015

Nivel escolar	Prueba	N.° correctores	N.° grupo de correctores
2° básico	Lectura	30	3
4° básico	Lectura	30	3
	Matemática	18	2

[57] La consistencia entre correctores se refiere a la proporción de respuestas idénticas que son juzgadas del mismo modo por dos o más correctores distintos.

Nivel escolar	Prueba	N.º correctores	N.º grupo de correctores
6º básico	Lectura	7	1
	Matemática	18	2
	Historia, Geografía y Cs. Sociales	6	1
	Escritura	80	10
8º básico	Lectura	48	4
	Matemática	12	1
	Ciencias Naturales	22	2
II medio	Lectura	24	2
	Matemática	11	1
	Historia, Geografía y Cs. Sociales	24	2

En términos operativos se distinguen dos procesos de corrección de preguntas abiertas: el de las pruebas censales en general y el de la prueba de Escritura.

Corrección de preguntas abiertas de pruebas censales

La corrección de preguntas abiertas de pruebas censales se realiza en función de los manuales de corrección validados en la aplicación experimental previa, que en el caso de las pruebas censales 2015 correspondió a las aplicadas experimentalmente en 2014, cuyo detalle se encuentra en el anexo A.

El proceso se inicia con la selección y capacitación de correctores y supervisores, a cargo de instituciones convocadas mediante licitación pública. Los equipos de trabajo conformados realizan una marcha blanca durante la cual se realiza una doble corrección del 10% de las respuestas obtenidas de cada pregunta. Si durante la marcha blanca no se alcanza el mínimo de 80% de consistencia, se repite la capacitación y de ser necesario, se reemplazan correctores y/o supervisores, y se realiza una nueva marcha blanca.

Una vez ajustados los equipos de trabajo, se corrige el 90% restante de las respuestas, con doble corrección de 10% y triple corrección del 10% de las respuestas doblemente corregidas.

Al finalizar el proceso, los códigos asignados durante la corrección se traducen a puntajes que son equiparables con la escala de puntajes utilizada en la corrección de las preguntas de opción múltiple.

En las pruebas censales 2015 se corrigió un total de 41 preguntas abiertas, distribuidas por asignatura y nivel escolar, según se muestra en la Tabla B2.

Tabla B2

Cantidad de preguntas abiertas censales 2015

Nivel escolar	Prueba	N.º preguntas
2º básico	Lectura	5
4º básico	Lectura	6
	Matemática	4
6º básico	Lectura	4
	Matemática	4
	Historia, Geografía y Cs. Sociales	1
8º básico	Lectura	3
	Matemática	3
	Ciencias Naturales	2
II medio	Lectura	5
	Matemática	2
	Historia, Geografía y Cs. Sociales	2

Corrección de preguntas abiertas de la prueba de Escritura

Como se señaló en el capítulo 1 de este Informe, en la prueba de Escritura 2015, aplicada a estudiantes de 6º básico, solo se corrigieron las preguntas abiertas.

A diferencia de las otras asignaturas, para las que se licitó este servicio, la prueba de Escritura se corrigió de manera interna. Esto implicó el arriendo de un centro de corrección con la infraestructura idónea para llevar a cabo el proceso y la contratación de aproximadamente 80 correctores, 10 supervisores y 2 coordinadores (uno para cubrir la jornada de la mañana y otro para la de la tarde). Asimismo, se asignó a dos profesionales permanentes de la Agencia como encargados de proyecto.

El proceso se inició con la selección y capacitación de los equipos de trabajo. Para la selección se estableció que los equipos debían quedar conformados por profesionales titulados, egresados o estudiantes de último año de algunas de las siguientes carreras: Pedagogía Básica con mención en Lenguaje y Comunicación, Pedagogía en Lenguaje y Comunicación o Licenciatura en Letras. Adicionalmente, en algunos casos específicos, también se aprobó la inclusión de egresados o estudiantes de último año de carreras afines, como Periodismo, Filosofía o Pedagogía en algún idioma extranjero.

Una vez seleccionados los profesionales, se conformaron 10 grupos de trabajo y se realizó una capacitación en función del Manual de Corrección que contiene las rúbricas respectivas, asignándose distintas rúbricas a los diversos grupos de correctores, de manera de lograr cierto grado de especialización.

Las rúbricas que se utilizaron para puntuar las respuestas de los estudiantes (rúbricas holísticas) poseen 6 niveles de desempeño (puntajes 1 a 6), siendo 1 el nivel inferior y 6 el superior. Con estas rúbricas se realizó doble corrección de todas las respuestas y, en caso de detectar discrepancias en la puntuación superiores a un punto (un nivel de diferencia) entre ambos correctores, se solicitó al supervisor del grupo dirimir con una tercera corrección. Con esos tres puntajes, se seleccionaron los dos que fueran iguales (o con una diferencia de máximo 1 nivel) y esos fueron los que se asignaron al estudiante[58].

En cuanto a la metodología de corrección, en la prueba de Escritura no existió marcha blanca, sin embargo, todos los días se realizó un proceso de calibración previo que consistió en iniciar el día con la corrección, por parte del supervisor, de 20 respuestas de la pregunta asignada. Luego el corrector debía puntuar esas respuestas y si la consistencia era igual o superior a 70%, el corrector podía continuar con la corrección de las restantes respuestas del día. Si el porcentaje era inferior a 70%, se le asignaban otras 20 respuestas corregidas previamente por el supervisor. Si la consistencia era nuevamente inferior a 70%, el corrector no podía participar en la jornada de ese día.

Una vez calibrada la corrección del día, se corregían las respuestas asignadas para esa jornada. Para cada una de estas se realizaba doble corrección y se supervisaba la mantención de niveles de consistencia de 70% o más, exacta de puntaje, y de 90% o más, entre puntajes adyacentes.

[58] Como cada estudiante respondió dos preguntas de la prueba de Escritura 2015: texto narrativo y texto informativo, cada estudiante obtuvo cuatro puntuaciones de escala 1 a 6 en total: 2 por cada tipo de texto. Para calcular la puntuación total, se sumaron las 4 puntuaciones y se puntuó al estudiante en una escala cuyo rango fue de 2 a 24 puntos.

Referencias bibliográficas

Australian Council for Educational Research (ACER). (2013). *Evaluación de los procesos y los productos asociados a la elaboración de instrumentos, las operaciones de campo y la gestión de datos de las pruebas nacionales de Simce. Santiago, Chile: [Documento interno de trabajo].*

AERA, APA y NCME. (2014). Standards for Educational and Psychological Testing. Washington: autores.

Agencia de Calidad de la Educación. (2013a). *Metodología de Construcción de Grupos Socioeconómicos Simce 2013. Santiago,Chile:autor. [En: http://archivos.agenciaeducacion.cl/biblioteca_digital_historica/metodologia/2013/gse_2013.pdf].*

Bock, R. y Mislevy, R. (1982). Adaptative EAP estimation of ability in a microcomputer environment, *Applied Psychological Measurement,* 6. Illinois, Chicago: University of Illinois.

Brancato, G. y otros (2006). *Handbook of Recommended Practices for Questionnaire Development and Testing in the European Statistical System. European Commission Grant Agreement 2004103000002,* European Statistical System.

Case, B., Jorgensen M. y Sucker S. (2004). *Alignment in Educational Assessment: Assessment Report.* San Antonio, Texas: Pearson Education Inc.

DeVellis, F. (2012). *Scale development: theory and applications. Thousand Oaks,* California: Sage Publications.

Downing, S. y T. Haladyna (2006). *Handbook of Test Development.* Mahwah, New Jersey: Lawrence Erlbaum Associates.

Haladyna, T.M y Rodríguez, M. C (2013). *Developing and Validating Items.* Nueva York: Routledge.

Kolen, M. y Brennan, R. (2004). *Test equating, Scaling and Linking: Methods and Practices.* Nueva York: Springer.

Lamela, M. (2009). Valoración de la condición física relacionada con la salud en el ámbito educativo. *[En: http://centros.edu.xunta.es/cfr/lugo/files/valoracióncondiciónfísica.pdf]*

Lord, F. y Novick, M. (2008). *Statistical theories of mental test scores.* Reading, Massachusetts: Addison-Wesley.

Martin, M., Rust, K., Adams, R. (eds.) (1999). *Technical Standards for IEA Studies.* Amsterdam, Países Bajos: International Association for the Evaluation of Educational Achievement (IEA).

Ministerio de Educación de Chile (2002). *Objetivos Fundamentales y Contenidos Mínimos Obligatorios de la Educación Básica. [Actualización 2002].* Santiago de Chile: autor.

(2009a). Ley N.° 20.370: *Establece la Ley General de Educación*. Santiago de Chile.

(2009b). Decreto N.° 170: *Fija normas para determinar los alumnos con necesidades educativas especiales que serán beneficiarios de las subvenciones para educación especial.* Santiago de Chile.

(2009c). *Objetivos Fundamentales y Contenidos Mínimos Obligatorios de la Educación Básica y Media. Actualización 2009*. Santiago de Chile.

Ministerio de Educación. (2011a). *Decreto N.° 439/2012: Bases Curriculares Educación Básica: Lenguaje y Comunicación, Matemática, Ciencias Naturales, Historia, Geografía y Ciencias Sociales, Idioma Extranjero Inglés.* Santiago de Chile.

(2011b): Ley N.° 20.529: *Sistema Nacional de Aseguramiento de la Calidad de la Educación Parvularia, Básica y Media y su Fiscalización,* Santiago de Chile.

Murphy, E. J. y Elliot, S. (1991). A Sustained Strategy for Preventing Potential Bias in theTMTE, *Teacher Certification Testing: Recent Perspectives*. Amherst, Massachusetts: NESInc.

Nogueira, J. (2002). Valoración de la condición física en niños de 11-12 años con distinto nivel socio-económico. *Revista Internacional de Medicina y Ciencias de la Actividad Física y el Deporte, 2,6. [En http://cdeporte.rediris.es/revista/revista6/valoracion.html].*

Organización para la Cooperación y el Desarrollo Económicos (OCDE) (2011). *Resultados del Informe PISA 2009: Lo que los estudiantes saben y pueden hacer. Rendimiento de los estudiantes en Lectura, Matemáticas y Ciencias*. Santiago de Chile: OCDE/Santillana Educación.

Ricker, K. (2006). Setting Cut-Scores: A Critical Review of the Angoff and Modified Angoff Methods. *Alberta Journal of Educational Research,* 52,1. Alberta, Canadá: University of Alberta.

Contáctanos
@agenciaeduca
facebook.com/Agenciaeducacion
instagram.com/agenciaeducacion
contacto@agenciaeducacion.cl
www.agenciaeducacion.cl