

DEMRE



# Piloto PSU

## Informe 2016

# Índice

Presentación .....	1
Características del Piloto PSU 2016 .....	2
Planificación .....	4
1. Aplicación de forma ancla .....	5
2. Plan de difusión .....	7
3. Muestreo.....	9
3.1 Marco muestral .....	9
3.2 Tipo de muestreo .....	14
3.3 Cierre de muestra .....	16
Operativa.....	19
4. Trabajo de campo.....	20
4.1 Equipo de aplicación.....	21
4.2 Traslado de material .....	23
4.3 Condiciones para rendir las pruebas piloto .....	23
4.4 Casos especiales.....	24
5. Resumen de la aplicación.....	26
5.1 Asistencia por prueba.....	26
5.2 Asistencia por sexo y región .....	27
5.3 Asistencia por tipo de establecimiento.....	28
5.4 Muestra efectiva .....	30
Análisis .....	32
6. Informes a establecimientos educacionales y estudiantes .....	33
7. Análisis de resultados.....	34
7.1 Teoría Clásica del Test (CTT) .....	35
7.2 Teoría de Respuesta al Ítem (IRT) .....	37
7.3 Procedimiento de análisis.....	38
7.4 Funcionamiento Diferencial del Ítem (DIF).....	40
7.5 Revisión cualitativa de ítems .....	47
7.6 Finalización del proceso.....	48
7.7 Análisis de los resultados en la forma ancla .....	48



Bibliografía .....	50
Anexos .....	52





# Presentación

Cada año el DEMRE realiza procesos de pilotaje de nuevos ítems para la PSU, cuyo fin es validar estadísticamente el comportamiento de las preguntas que se construyen para las distintas pruebas. Esto permite, contar con un banco de preguntas disponibles para las pruebas oficiales que se utilizan en cada Proceso de Admisión Universitaria.

El presente informe describe la metodología y las distintas etapas asociadas a la aplicación piloto 2016. La aplicación piloto del año 2016 se realizó en el mes de septiembre y tuvo algunas particularidades que la diferencian de procesos anteriores. Los cambios realizados en esta aplicación tuvieron por objetivo mejorar la información disponible para los análisis de las preguntas y aumentar la tasa de participación de los estudiantes. Por lo tanto, estos cambios tuvieron relación con la metodología de muestreo, la difusión del proceso y los análisis realizados.

La muestra de este proceso se realizó en comunas de las regiones de Coquimbo, Valparaíso, Región Metropolitana y Bío-Bío, representando a las zonas norte, centro y sur del país. Estas reflejan lo mejor posible las características de su zona en cuanto a la distribución de las pruebas rendidas, composición de establecimientos educacionales por rama educativa y dependencia administrativa, condición de ruralidad, rendimiento en PSU y a nivel escolar (NEM y Ranking). En cada zona se seleccionaron los establecimientos y los cursos de 4° medio que participarían en la aplicación piloto. Los cambios realizados en el muestreo se complementaron con una difusión del piloto 2016, donde se realizó una estrategia comunicacional para motivar la asistencia de los estudiantes y su compromiso con la aplicación, llegando a una tasa de asistencia de aproximadamente el 65%-75% en cada prueba.

Para la evaluación de calidad de los ítems se utilizó la Teoría Clásica del Test (CTT, *Classic Theory of Tests*) y se complementó este análisis con el modelo de Teoría de Respuesta al Ítem (IRT, *Item Response Theory*). De esta forma, es posible establecer *modelos* capaces de evaluar las *propiedades psicométricas* de los instrumentos de medición. Específicamente, se estudian aquellos factores que influyen sobre las puntuaciones obtenidas en los test y sus ítems, proponiendo modelos que permitan controlar y minimizar los factores de error. Estos factores de error inciden en las estimaciones realizadas de los resultados a partir de la prueba.

El número de ítems aprobados en el proceso de pilotaje 2016 mejoró considerablemente respecto de años anteriores. Una vez realizados los análisis, se contó con un 53% a 84% de ítems aprobados por prueba disponibles para su uso en el ensamblaje de la prueba Oficial.

El informe se organiza en seis capítulos. El primero de ellos describe las principales características de la aplicación piloto. El segundo capítulo presenta las características del muestreo utilizado y el tercero presenta los detalles del trabajo de campo. El cuarto capítulo analiza la asistencia y las diferencias entre la muestra convocada a rendir y la que efectivamente asistió a las distintas pruebas. Finalmente, los dos últimos capítulos se refieren al informe de resultados para estudiantes y establecimientos educacionales, y a los análisis psicométricos de los ítems piloteados.

# Características del Piloto PSU 2016

La aplicación piloto de la PSU es una tarea crucial en el proceso de construcción de instrumentos de evaluación dentro del Sistema Único de Admisión, ya que en esta se realiza la validación estadística de los ítems que luego formarán parte de las pruebas definitivas. Esta validación se realiza en una muestra representativa de la población que rinde la PSU oficial, bajo estrictos parámetros que aseguran la calidad del producto final.

La aplicación piloto del año 2016 estuvo orientada a mejorar una dimensión que tradicionalmente presentaba dificultades en las aplicaciones de años anteriores: la baja asistencia de los estudiantes convocados. Así, este año se incorporó una serie de cambios en relación con la planificación e implementación del proceso. Desde el punto de vista del DEMRE, estos cambios permitirían aumentar la tasa de asistencia, y con ello, asegurar la recolección de información suficiente para analizar de manera confiable el comportamiento de los ítems piloteados. Esto permite velar por la calidad de los ítems a utilizar posteriormente en el instrumento oficial.

Para este proceso, el DEMRE difundió el piloto como una oportunidad para que los estudiantes pudiesen poner a prueba su futuro desempeño en la PSU en las mismas condiciones de aplicación que en la prueba oficial. Sumado a lo anterior, y para motivar aún más la participación, se entregó un reporte de rendimiento, tanto a nivel individual como a cada establecimiento educacional.

Para la aplicación, se seleccionaron regiones y comunas que permitiesen caracterizar a la población del país que rinde PSU en términos de tipo de establecimientos, características escolares (como rendimiento y puntajes PSU) y geográficas. Las regiones seleccionadas fueron Coquimbo, Valparaíso, Metropolitana y Bío-Bío.

Tras considerar una serie de variables —como dependencia, rama, tasa de participación y nivel de rendimiento en prueba oficial, entre otros—, se seleccionaron comunas al interior de cada región<sup>1</sup> y se aplicó un muestreo bietápico para la selección de estudiantes. Primero se seleccionaron los establecimientos educacionales participantes y, dentro de cada establecimiento, se seleccionaron cursos completos.

La aplicación piloto se realizó los días martes 6 y miércoles 7 de Septiembre, con la excepción de Chillán. En esta última comuna, la aplicación fue los días jueves 8 y viernes 9 del mismo mes.

Durante la aplicación piloto se contó con la estrecha colaboración de los Secretarios de Admisión que el DEMRE tiene a lo largo del país para cada una de las sedes que participaron en este proceso. Además, participaron directivos, profesores, profesionales y auxiliares de los establecimientos que facilitaron sus dependencias como locales de aplicación. Para la organización de la aplicación, el DEMRE también trabajó estrechamente con los establecimientos que fueron parte de la muestra, lo cuales presentaron alto interés y motivación por participar en el proceso.

Cada uno de los estudiantes convocados para el piloto rindió tres pruebas: las dos obligatorias —Lenguaje y Comunicación y Matemática— y una electiva (Historia o Ciencias). La asignación de la prueba electiva se realizó primordialmente en base a la inscripción a la PSU Oficial Admisión 2017. Para aquellos estudiantes que hubiesen seleccionado ambas pruebas electivas en su inscripción PSU

---

<sup>1</sup> Las comunas se detallan en la sección 3.1.



o que a la fecha de la selección de la muestra no estuviesen inscritos aun, los establecimientos educacionales informaron al DEMRE sobre la prueba electiva que sus estudiantes deseaban rendir.

Las pruebas piloto de Lenguaje, Matemática e Historia, Geografía y Ciencias Sociales tuvieron 75 preguntas, mientras que las de Ciencias estuvieron compuestas por 80 preguntas. En el Piloto se probaron 75 formas, distribuidas en las distintas pruebas que componen la batería PSU. Tal como se muestra en la Tabla 1.

Tabla 1. Pruebas Piloto PSU, cantidad de formas probadas

	Prueba	Numero de formas
	Lenguaje y Comunicación	12
	Matemática	12
	Historia, Geografía y Ciencias Sociales	15
Ciencias	Biología	9
	Física	9
	Química	9
	Técnico Profesional	9
	<b>TOTAL</b>	<b>75</b>

# Planificación





## 1. Aplicación de forma ancla

Realizar el pilotaje de preguntas en una muestra que caracteriza a la población que finalmente participará de la aplicación oficial permite suponer que los resultados obtenidos en la aplicación piloto serán similares a los de la aplicación oficial. Este supuesto permite confeccionar las pruebas oficiales con mayor información y anticipar el comportamiento de las preguntas utilizadas.

Uno de los desafíos al realizar el pilotaje de preguntas de forma independiente a la aplicación oficial es asegurar dicha similitud de las condiciones de rendición. Una de las condiciones de rendición que aparecen como inigualables entre la aplicación piloto y la oficial tiene relación con la motivación de los estudiantes para responder correctamente la prueba.

Para estudiar las diferencias entre los resultados de la aplicación piloto y la oficial, en el año 2016 se utilizó para cada disciplina una forma "ancla". La forma ancla fue una de las formas aplicadas en la prueba oficial aplicada en diciembre de 2015 para el proceso de Admisión 2016. Así, a la cantidad de formas detallada en la Tabla 1 se suma una forma ancla para cada disciplina.

En el caso de la prueba de Ciencias, la forma ancla utilizada correspondió a una de las pruebas oficiales Ciencias-TP. Esto permite que la forma ancla pueda aplicarse aleatoriamente a los estudiantes que participaron en la aplicación piloto de las pruebas de Ciencias. Luego, para verificar la estabilidad de los parámetros entre la prueba oficial y la aplicación piloto se comparan los resultados obtenidos en la forma ancla, pues fue utilizada en ambas instancias.



## 2. Plan de difusión

Para mejorar la participación de los establecimientos educacionales en el Piloto, el DEMRE llevó a cabo una serie de estrategias para involucrarlos intensamente en el proceso. Para ello, se realizaron reuniones informativas y visitas a los establecimientos educacionales seleccionados, donde se entregó material de difusión como afiches y dípticos (ver Anexo 1). La difusión también se realizó mediante el envío de correos electrónicos a establecimientos educacionales y estudiantes, que contenían toda la información relevante de la aplicación piloto.

Las reuniones informativas tuvieron el objetivo de comunicar los aspectos operativos y la relevancia de la participación de los establecimientos en el pilotaje. A estas reuniones se convocó a directivos u orientadores de educación media. Principalmente, se buscó la asistencia de representantes que tuviesen cercanía con los estudiantes para motivar la inscripción temprana en la PSU Oficial y la participación en la aplicación piloto.

En la difusión de la aplicación piloto, el DEMRE enfatizó a los establecimientos la necesidad de contar con información sobre la participación y la prueba electiva que los estudiantes deseaban rendir en la aplicación piloto. Esta última información era crucial para asegurar la asignación de una prueba electiva que respondiera a los intereses de los estudiantes y así evitar el ausentismo. El DEMRE definió dos vías para obtener esta información:

- Inscripción a la PSU
- Información proporcionada por los establecimientos educacionales

Cuando los establecimientos educacionales confirmaron su participación en el piloto, el DEMRE les envió una planilla con la nómina de estudiantes de 4<sup>to</sup> medio matriculados en el establecimiento, según los datos reportados al Ministerio de Educación. Esta planilla indicaba, para cada estudiante, su estado en la inscripción PSU Oficial (inscrito / no inscrito) y, en caso de estar inscrito, las pruebas electivas seleccionadas.

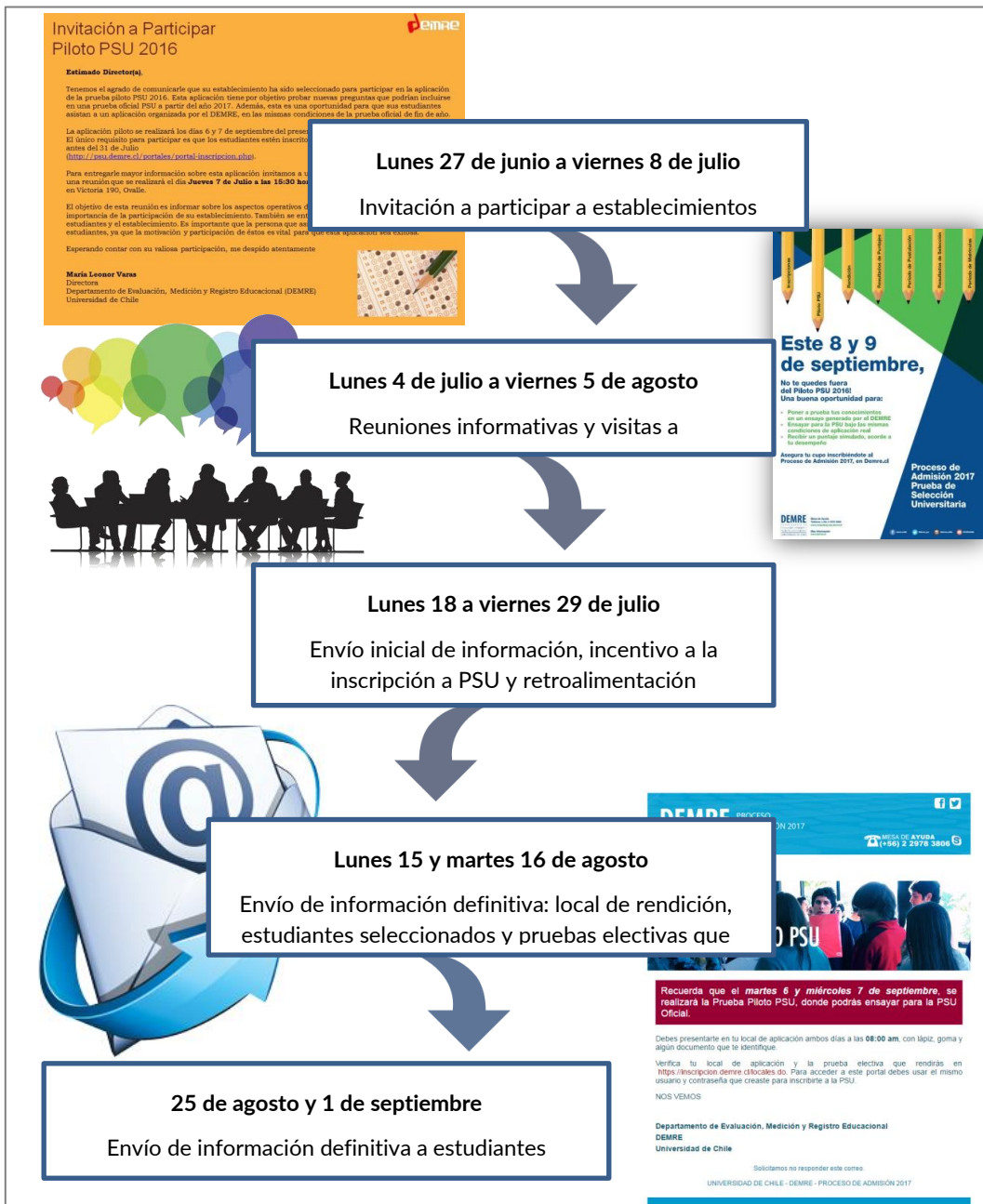
El propósito fue que, utilizando esta planilla, cada establecimiento consultara a sus estudiantes y retroalimentara al DEMRE sobre:

- Participación: que los establecimientos notificaran sobre estudiantes que no estuviesen en la nómina, estudiantes retirados del establecimiento, y aquellos que, estando en la nómina, no deseaban participar del Piloto.
- Pruebas electivas: que los establecimientos informaran la prueba electiva que deseaban rendir aquellos estudiantes aun no inscritos a la PSU Oficial y aquellos que en su inscripción habían seleccionado ambas pruebas electivas.
- Situaciones especiales: que los establecimientos informaran sobre los estudiantes que participan de programa de integración, o que se encuentran en situación de discapacidad, para que DEMRE tomara las medidas pertinentes en la aplicación piloto.

Esta información fue recibida hasta el día 8 de agosto, a las 7 am. Ese mismo día se inició la asignación de pruebas electivas y de locales de aplicación. Para aquellos establecimientos que no enviaron información, el DEMRE convocó a rendir solo a aquellos estudiantes que estaban inscritos en la PSU Oficial y asignó las pruebas electivas de acuerdo a dicha inscripción. En el caso de doble selección de prueba electiva, se privilegió la asignación a las pruebas de Ciencias.

A partir de la tercera semana de agosto, el DEMRE envió la información definitiva de la aplicación piloto a los establecimientos educacionales: ubicación del local de rendición asignado, protocolo sugerido por el Ministerio de Educación para informar la participación, nómina de estudiantes seleccionados y pruebas a rendir. A su vez, y en conjunto con el Servicio de Tecnología e Información de la Universidad de Chile, el DEMRE envió correos informativos a los estudiantes, para que conocieran los detalles del proceso, el local de rendición al que debían acudir y la prueba electiva que iban a rendir. El plan de difusión se resume en la Figura 1.

Figura 1. Plan de difusión, Piloto PSU 2016





## 3. Muestreo

### 3.1 Marco muestral

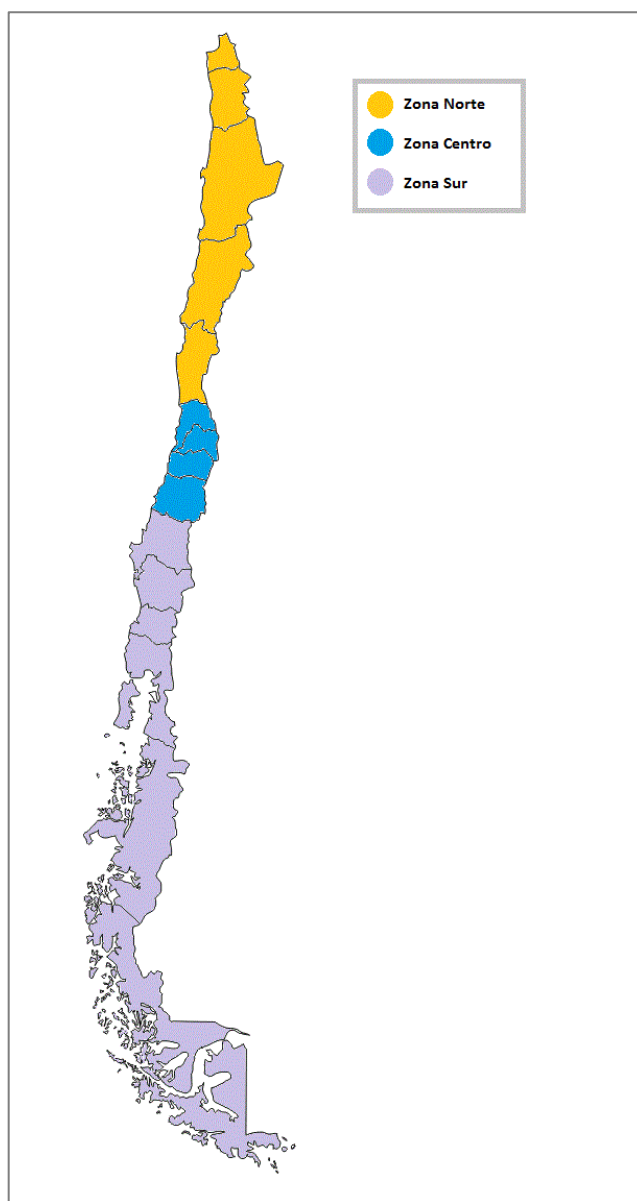
Previo a determinar los criterios de selección de la muestra y con el objetivo de tener información de referencia, se analizaron los datos de rendición PSU del año anterior (Admisión 2016) y otros antecedentes educativos. Los análisis se realizaron en distintos niveles (regional, comunal y por establecimiento) y se señalan en la Tabla 2:

Tabla 2. Análisis previos a la selección muestral

Análisis	Nivel		
	Regional	Comunal	Establecimiento
Asistencia a PSU Admisión 2016 por prueba	✓	✓	✓
Tasa de asistencia a aplicación piloto realizada en 2015 por prueba	✓	✓	✓
Distribución de establecimientos educacionales por rama de enseñanza	✓	✓	
Distribución de establecimientos educacionales por dependencia administrativa	✓	✓	
Promedio de rendimiento NEM y Ranking en PSU Admisión 2016, según rama de enseñanza	✓		
Promedio de rendimiento NEM y Ranking en PSU Admisión 2016, según dependencia administrativa.	✓		
Promedio de puntajes PSU Admisión 2016 (Lenguaje, Matemática, Historia, Biología, Física, Química y Técnico Profesional), según rama de enseñanza.	✓		
Promedio de puntajes PSU Admisión 2016 (Lenguaje, Matemática, Historia, Biología, Física, Química y Técnico Profesional), según dependencia administrativa.	✓		

Por razones prácticas y operativas, se tomó la decisión de dividir al país en tres zonas geográficas: norte, centro y sur. La zona norte consideró a las regiones I a IV y XV. La zona centro agrupó las regiones V a VII y Metropolitana. Por último, la zona sur correspondió a las regiones VIII a XII y XIV, tal como se expone en la Figura 2.

Figura 2. Zonas geográficas, Piloto PSU 2016



Tras una serie de análisis, se determinó que estas zonas serían representadas por las regiones de Coquimbo, de Valparaíso, Metropolitana y de Bío-Bío. La región de Coquimbo caracterizó a la zona norte, las regiones Valparaíso y Metropolitana, caracterizaron al centro y la región del Bío-Bío caracterizó al sur. Se buscó que las regiones seleccionadas reflejaran lo mejor posible las características de su zona en cuanto a la distribución de las pruebas rendidas, composición de establecimientos educacionales por rama educativa y dependencia administrativa, condición de ruralidad, rendimiento en PSU y a nivel escolar (NEM y Ranking), tal como lo señalan las Tablas Tabla 3 y Tabla 4. Otro criterio relevante para seleccionar estas regiones fue que éstas agrupan al 62,8% de los establecimientos educacionales del país.





Tabla 3. Composición de establecimientos por rama de enseñanza y dependencia administrativa, según zona del país y regiones seleccionadas

	Rama de enseñanza							Dependencia administrativa		
	HC diurno	HC nocturno	Técnico comercial	Técnico industrial	Técnico servicios	Técnico agrícola	Técnico marítimo	Particular pagado	Particular subvencionado	Municipal
<b>Zona norte</b>	<b>45,2%</b>	<b>14,6%</b>	<b>11,9%</b>	<b>13,9%</b>	<b>10,9%</b>	<b>2,3%</b>	<b>1,2%</b>	<b>7,6%</b>	<b>47,3%</b>	<b>45,0%</b>
<i>IV Región</i>	51,5%	18,6%	9,5%	9,1%	6,9%	3,9%	0,4%	4,3%	58,4%	37,2%
<b>Zona Centro</b>	<b>48,0%</b>	<b>17,1%</b>	<b>10,6%</b>	<b>10,1%</b>	<b>9,6%</b>	<b>4,2%</b>	<b>0,4%</b>	<b>8,5%</b>	<b>49,3%</b>	<b>42,2%</b>
<i>V Región</i>	54,4%	17,7%	7,9%	9,2%	7,4%	2,0%	1,5%	10,5%	52,2%	37,3%
<i>R.M.</i>	50,9%	16,0%	13,7%	9,7%	9,0%	0,7%	0,0%	14,4%	58,0%	27,7%
<i>Promedio V y R.M.</i>	52,6%	16,9%	10,8%	9,4%	8,2%	1,3%	0,7%	12,4%	55,1%	32,5%
<b>Zona sur</b>	<b>40,2%</b>	<b>20,2%</b>	<b>10,2%</b>	<b>11,5%</b>	<b>11,6%</b>	<b>4,1%</b>	<b>2,2%</b>	<b>4,9%</b>	<b>48,2%</b>	<b>47,0%</b>
<i>VIII Región</i>	43,6%	18,6%	10,4%	12,8%	8,9%	4,7%	0,9%	5,7%	40,7%	53,6%

HC: Humanista-científico

Tabla 4. Promedio de puntajes por factor de selección, según zona y regiones seleccionadas, Proceso de Admisión 2015

	NEM	RANK	LENG	MATE	HIST	CBIO	CFIS	CQUI	CTEC
<b>Zona norte</b>	<b>528,8</b>	<b>555,4</b>	<b>460,0</b>	<b>451,9</b>	<b>469,6</b>	<b>453,6</b>	<b>494,4</b>	<b>484,3</b>	<b>415,5</b>
<i>IV Región</i>	526,1	554,0	466,7	459,1	474,7	466,1	502,1	500,8	422,2
<b>Zona Centro</b>	<b>528,1</b>	<b>554,8</b>	<b>471,5</b>	<b>463,7</b>	<b>480,6</b>	<b>462,4</b>	<b>506,7</b>	<b>506,1</b>	<b>425,9</b>
<i>V Región</i>	527,2	552,1	477,6	466,9	484,5	465,8	509,6	498,1	421,3
<i>R.M.</i>	513,7	538,4	486,4	474,3	497,5	476,3	515,3	517,9	431,8
<i>Promedio V y R.M.</i>	520,4	545,2	482,0	470,6	491,0	471,0	512,5	508,0	426,5
<b>Zona sur</b>	<b>534,9</b>	<b>563,8</b>	<b>464,8</b>	<b>452,7</b>	<b>475,6</b>	<b>455,7</b>	<b>493,2</b>	<b>495,2</b>	<b>414,2</b>
<i>VIII Región</i>	538,5	565,2	460,2	454,9	467,6	455,3	497,0	499,7	414,8

**NEM:** Notas de Educación Media, **RANK:** Puntaje Ranking, **LENG:** Lenguaje y Comunicación, **MATE:** Matemática, **HIST:** Historia, Geografía y Ciencias Sociales, **CBIO:** Ciencias-Biología, **CFIS:** Ciencias-Física, **CQUI:** Ciencias-Química, **CTEC:** Ciencias-Técnico Profesional



Además de lo anterior, la inclusión de estas regiones consideró elementos como la tasa de asistencia a la aplicación Piloto del año 2015 y la PSU admisión 2016 y la disponibilidad de establecimientos educacionales en la región. Estos datos se presentan en la Tabla 5.

Tabla 5. Asistencia a Piloto 2015 y PSU Admisión 2016 y disponibilidad de establecimientos, según regiones seleccionadas

Región	Tasa de asistencia a Piloto 2015	Asistencia a PSU Admisión 2016	Cantidad de establecimientos educacionales
Coquimbo	75,7%	10.371 (4% del total de asistentes a nivel nacional)	231 (4,7% del total de establecimientos a nivel nacional)
Valparaíso	64,7%	27.463 (11% del total)	609 (12,3% del total)
Metropolitana	66,1%	100.817 (40% del total)	1.714 (34,7% del total)
Bío-Bío	67%	32.030 (13% del total)	548 (11,1% del total)

Dentro de cada región, las comunas a seleccionar también se escogieron analizando estos factores (asistencia a pruebas, disponibilidad de establecimientos según rama y dependencia, rendimiento). Dentro de todos los factores considerados para la selección de comunas, el más importante fue la asistencia al Piloto del año 2015, como se muestra en la Tabla 6.





Tabla 6. Tasa de asistencia a Piloto 2015 y disponibilidad de establecimientos, en comunas seleccionadas

Región	Comuna	Tasa de asistencia al Piloto 2015	Disponibilidad de establecimientos educacionales*
Coquimbo	Coquimbo	71,9%	76%
	La Serena	75,4%	
	Ovalle	82,9%	
Valparaíso	Quillota	69,5%	52%
	Quilpué	62,6%	
	San Felipe	78,6%	
	Valparaíso	62,5%	
	Viña del Mar	63,3%	
Metropolitana**	La Cisterna	54,1%	54%
	La Florida	64,3%	
	Las Condes	78,9%	
	Maipú	63,9%	
	Melipilla	76,5%	
	Nuñoa	62,9%	
	Providencia	72,7%	
	Puente Alto	65,3%	
	Quilicura	75,5%	
	San Bernardo	64,1%	
	San Miguel	69,2%	
	Santiago	53,4%	
Bío-Bío	Chillán	63,9%	40%
	Concepción	66,2%	
	Los Ángeles	62,3%	
	Talcahuano	68,3%	

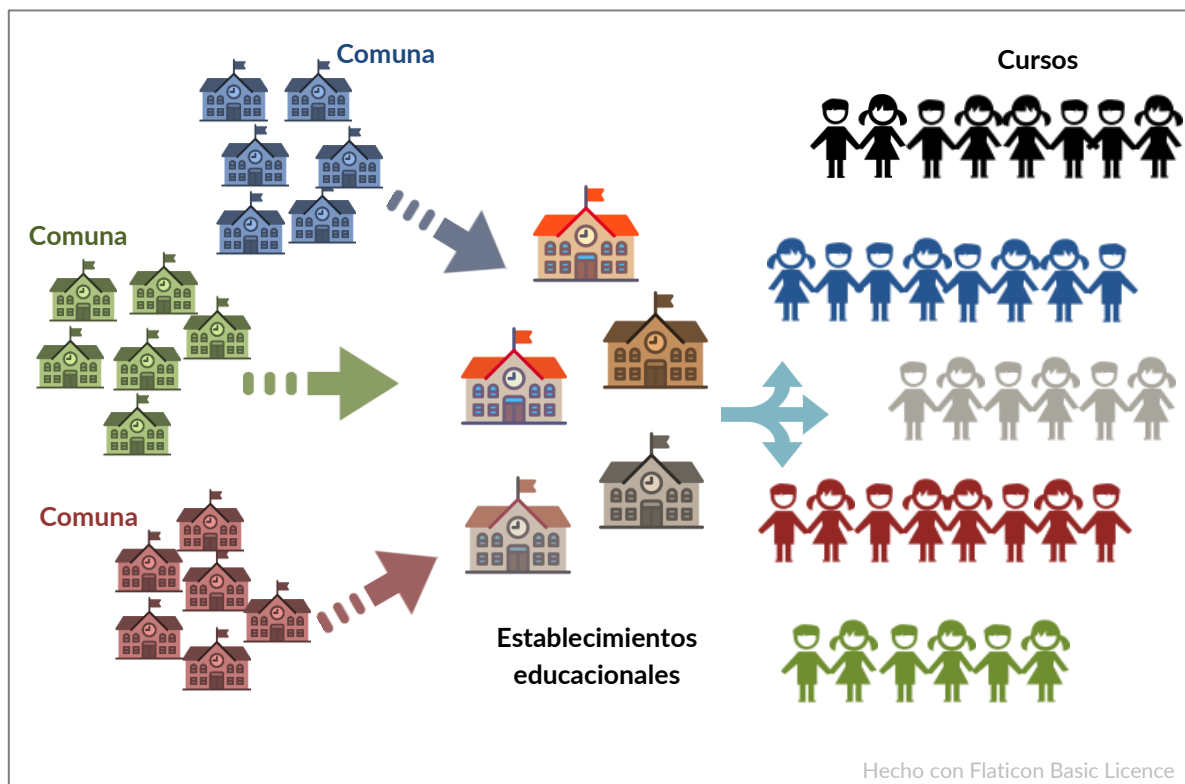
\* Porcentaje de establecimientos educacionales respecto del total de la región que agrupan las comunas seleccionadas.

\*\* Se agregó, además, la comuna de El Bosque, que no participó del pilotaje 2015 y, por lo tanto, no presenta datos de asistencia al piloto 2015.

## 3.2 Tipo de muestreo

Una vez determinadas las regiones y comunas que representarían a cada zona del país, se seleccionaron los establecimientos y los cursos de 4° medio que participarían en la aplicación piloto. Para ello, se utilizó un muestreo bietápico. En la primera etapa, se escogió una muestra probabilística cuya unidad de muestreo fue el establecimiento educacional. Posteriormente, en la segunda etapa, se seleccionaron los cursos al interior de cada establecimiento. Esta segunda etapa combinó criterios de selección aleatorios y requerimientos técnicos. El detalle del procedimiento se ilustra en la Figura 3 y se explica en el siguiente apartado.

Figura 3. Muestreo Piloto PSU 2016



### 3.2.1. Primera etapa

La primera etapa fue la selección de unidades educativas<sup>2</sup> al interior de las comunas elegidas para el pilotaje. Para la aplicación piloto, se consideró un solo criterio de exclusión dentro de selección de las unidades educativas. Las ramas Técnico Agrícola y Marítimo, y las modalidades nocturnas<sup>3</sup> (tanto HC como TP) no fueron incluidas en el marco muestral, ya que representan una proporción muy menor de la población que rinde la prueba Oficial.

<sup>2</sup> Cada "Unidad Educativa" es un elemento único dentro de la población, y está compuesta por un Rol Base de Datos ("RBD") y un código de enseñanza. Por ejemplo, si un establecimiento educacional polivalente (con un RBD "9999"), tiene dos modalidades de enseñanza (Científico Humanista y Técnico Industrial), entonces éste se consideraba como dos unidades distintas dentro de la población: "9999\_310" y "9999\_510".

<sup>3</sup> Respecto de las modalidades nocturnas, estas no fueron seleccionadas además por presuntos problemas de disponibilidad y compatibilidad horaria, lo que podría impactar en la asistencia al pilotaje.



Dentro de cada comuna, la población muestral consideró establecimientos educacionales de todas las dependencias administrativas, de las ramas HC Diurno, y Técnico Comercial, Industrial y Servicios. En base a la inscripción y el rendimiento en PSU para el Proceso de Admisión 2016, se seleccionaron unidades educativas en función de la cantidad de estudiantes requerida por cada comuna. La selección fue probabilística, ya que dentro de cada comuna cada establecimiento tenía una probabilidad conocida y distinta de cero de ser parte de la muestra (Vivanco, 2005).

Para llevar a cabo la selección, se utilizó una aplicación computacional que permitió escoger unidades educativas considerando los factores asociados a la rendición de la PSU. La aplicación utilizó la rama educativa de las unidades para la generación de estratos. Dentro de cada unidad educativa se calculó el número y proporción de inscritos a la PSU oficial y su rendimiento, para asignar una probabilidad de selección en base a estos criterios.

Debido a que a la fecha de la selección de la primera etapa de la muestra (unidades educativas), no se contaba con los datos de inscripción a la PSU Admisión 2017, se utilizaron como *proxy* los datos de la prueba oficial del año anterior. Así, la aplicación computacional desarrollada seleccionó las unidades educativas, mediante la estimación de un factor aleatorio compuesto por los elementos de la PSU Admisión 2016 antes descritos.

Tras la selección de los establecimientos que participarían en el piloto, la aplicación computacional analizaba la localización geográfica de estos establecimientos respecto de los locales de rendición. De esa forma, se calculó la distancia entre cada establecimiento y los locales de rendición dentro de su comuna, asignándosele el de menor distancia.

Una vez seleccionados los establecimientos, el DEMRE se comunicó con ellos para informar los detalles de la aplicación piloto (ver plan de difusión detallado en la sección 2). En caso de que algún establecimiento declinara la invitación, se estableció una estrategia de reemplazo que buscó una nueva unidad educativa de similares características dentro de la misma comuna.

### **3.2.2. Segunda etapa**

Para la selección de estudiantes se generó una base de datos con todos los cursos que componían cada unidad educativa seleccionada. La base contenía la matrícula por curso y la cantidad de estudiantes que deseaban rendir cada una de las pruebas electivas. Esta base, se actualizó constantemente a partir de los datos de inscripción a PSU 2017 y la información proporcionada por cada establecimiento educacional. Además, la información fue utilizada para monitorear la selección de pruebas electivas, que debía ajustarse a los requerimientos técnicos del muestreo.

En esta fase, las unidades muestrales se definieron como los cursos pertenecientes a las unidades educativas seleccionadas en la etapa anterior. La selección de cursos se llevó a cabo de tres formas; según un criterio de inclusión, de modo aleatorio y de forma intencionada.

Respecto del criterio de inclusión, éste consideró el ingreso de una cierta cantidad de cursos dependiendo del caso de cada unidad educativa. Primeramente, en las unidades educativas que actuaron como local de aplicación, se seleccionó la totalidad de sus cursos de 4° medio. En los establecimientos que no fueron locales de aplicación, se aseguró la participación completa de

aquellos que tenían hasta tres cursos<sup>4</sup>. En los establecimientos con cuatro o más cursos, se aplicó un muestreo aleatorio simple, en base a los criterios que se señalan en la Tabla 7.

Tabla 7. Criterios de selección en establecimientos de 4 o más cursos de 4° medio

Cursos de 4° medio en el establecimiento	Promedio de estudiantes por curso	Cantidad de cursos a seleccionar
Entre 4 y 5	< 25 estudiantes	Todos
	≥ 25 estudiantes	3
Entre 6 y 8	< 20 estudiantes	Todos
	≥ 20 estudiantes	5
9 o más	< 20 estudiantes	Todos
	≥ 20 estudiantes	7

Por último, se aplicó un criterio de ‘selección intencionada’ que incorporó a la muestra ciertos cursos que habían quedado fuera del muestreo aleatorio. El criterio de selección fue que los cursos tuviesen una alta proporción de inscritos en las pruebas de Ciencias, sobre todo en Física y Química. De ese modo, fue posible satisfacer los requerimientos técnicos para los análisis psicométricos posteriores.

Una vez obtenida la muestra final de establecimientos y cursos seleccionados para pilotaje (8 agosto), el DEMRE informó de esto a los involucrados. Con el fin de estimular la asistencia y hacer partícipes a los establecimientos del proceso, se realizaron algunos cambios de prueba electiva a los estudiantes que lo solicitaron y se agregó a la muestra aquellos alumnos que aún no estuviesen inscritos en PSU (sin importar el motivo). Estas acciones se realizaron oficialmente hasta el día 23 de agosto.

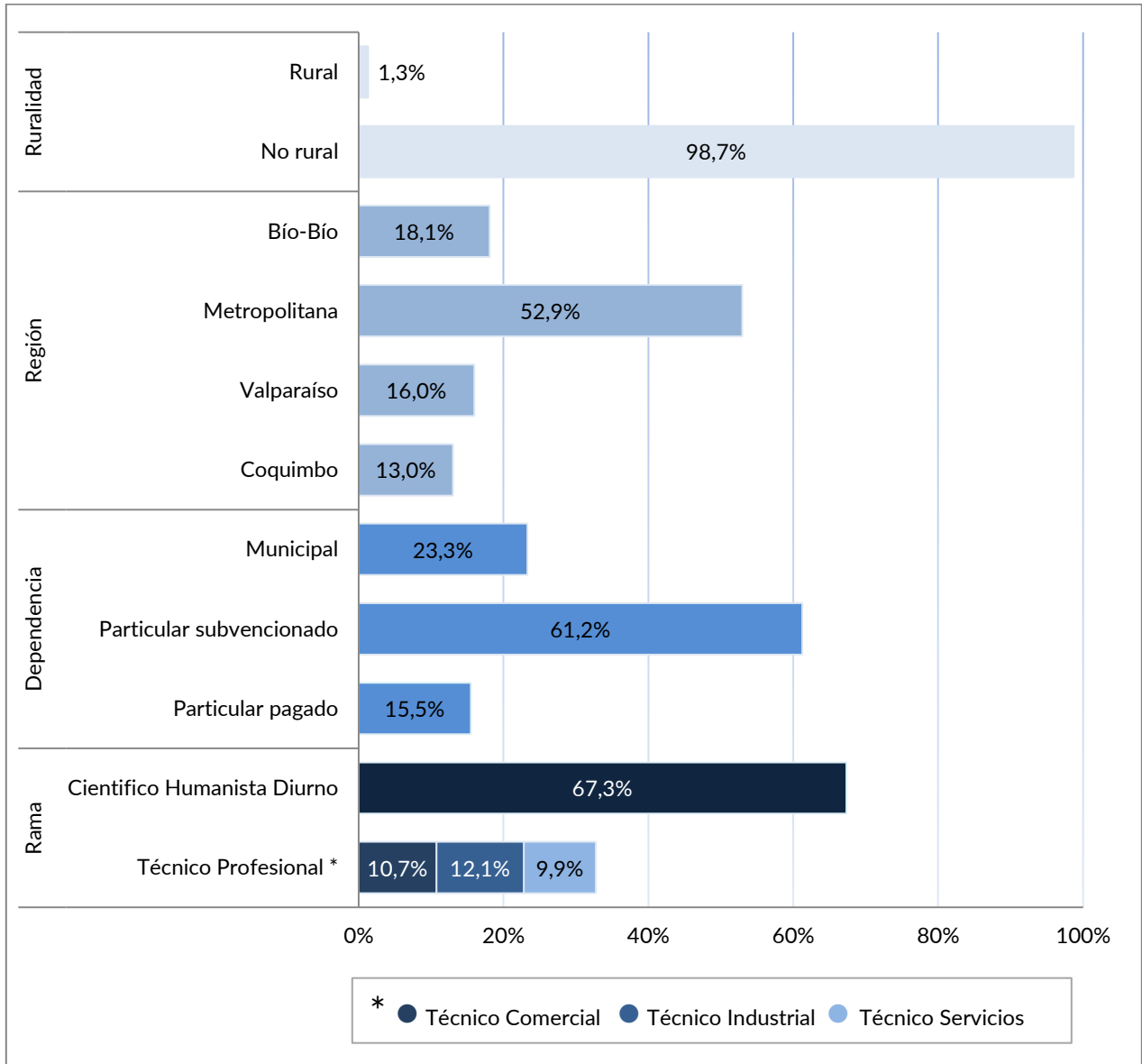
### 3.3 Cierre de muestra

En base a todos los procedimientos anteriormente descritos, para la aplicación piloto se seleccionó un total de 1153 cursos, distribuidos en 614 unidades educativas. En el Gráfico 1, se presentan las características principales de estos establecimientos. La desagregación por regiones puede verse en el Anexo 2. En términos generales, las características de los establecimientos que componen la muestra se ajustan al marco muestral a nivel nacional (ver datos de distribución de establecimientos para la Prueba Oficial en Anexo 3).

<sup>4</sup> Esta inclusión fue asegurada siempre y cuando los estudiantes de esos cursos estuviesen inscritos a la PSU Admisión 2017, o el establecimiento hubiese confirmado la participación de los estudiantes en el piloto.



Gráfico 1. Características de los establecimientos seleccionados para pilotaje



En el gráfico se observa que, del total de establecimientos convocados, la mayor cantidad corresponde a establecimientos Humanista Científico de modalidad diurna (67,3%). En tanto, establecimientos de rama Técnico Profesional corresponden a un tercio del total, en las tres modalidades señaladas.

La mayoría de establecimientos seleccionados son de dependencia particular subvencionada (61%), lo cual es consistente con la distribución de estudiantes en el sistema escolar. A su vez, se seleccionaron 143 establecimientos municipales y 95 particulares pagados, que representan el 23,3% y 15,5% de la muestra, respectivamente. Por razones relacionadas con la distribución de la población en el país, más del cincuenta por ciento de los establecimientos que se seleccionaron fueron de la Región Metropolitana.

Con estos establecimientos y cursos seleccionados, la muestra final convocada a rendir el piloto estuvo compuesta por 29.149 estudiantes. Estos se distribuyeron por región, sexo y prueba electiva de acuerdo a lo que se presenta en la Tabla 8.

Tabla 8. Muestra teórica Piloto PSU 2016, según región, sexo, y prueba electiva

Indicador	Categoría	Frecuencia	Porcentaje
Región	Coquimbo	3.418	11,7%
	Valparaíso	3.813	13,1%
	Metropolitana	14.243	48,9%
	Bío-Bío	7.675	26,3%
Sexo	Mujer	15.830	54,3%
	Hombre	13.319	45,6%
Prueba electiva	Historia	9.456	32,4%
	Ciencias-Biología	7.473	25,6%
	Ciencias-Física	4.104	14,1%
	Ciencias-Química	3.696	12,7%
	Ciencias-Técnico Profesional	4.420	15,2%



# Operativa

## 4. Trabajo de campo

Para llevar a cabo la aplicación, tanto del Piloto como de la PSU Oficial, se divide el país en unidades geográficas regionales y comunales denominadas Sedes de Rendición. A su vez, las sedes de rendición se constituyen por uno o más locales de aplicación de pruebas. La aplicación del Piloto PSU se llevó a cabo en 62 locales de aplicación, los días martes 6 y miércoles 7 de septiembre. La única excepción fue la comuna de Chillán, donde la prueba se realizó los días jueves 8 y viernes 9 del mismo mes.



Para la realización del Piloto se contó con la autorización oficial de Ministerio de Educación, considerándose esta instancia como un “Cambio de Actividad Lectiva”. Para justificar la asistencia de los estudiantes a la aplicación piloto y responder al control de subvenciones, el DEMRE hizo entrega de la asistencia oficial a los establecimientos educacionales. Un ejemplo de listado de asistencia puede verse en el Anexo 5.

Para evaluar el posible efecto en el orden de rendición de las pruebas obligatorias (Lenguaje y Matemática), algunas comunas iniciaron la aplicación con la prueba de Matemática, mientras otras comenzaron con Lenguaje y Comunicación (ver Anexo 4 para mayor detalle). Así, el primer día se rindió una prueba obligatoria y posteriormente se rindieron simultáneamente las dos pruebas electivas. En el segundo día de aplicación solo se rindió una prueba obligatoria. Lo anterior, se ilustra con dos ejemplos en la Figura 4.





Figura 4. Programación de días de aplicación, según comuna

 <b>La Serena</b>	<b>Día 1</b>	<b>Día 2</b>
	<b>Desde 8:00 am.</b> Prueba de Lenguaje y Comunicación (2 hrs. 30 min.) <b>Descanso</b> <b>Aprox. 11:00 am.</b> Pruebas electivas Ciencias (2 hrs. 55 min.) o Historia, Geografía y Ciencias Sociales (2 hrs. 30 min.)	<b>Desde 8:00 am.</b> Prueba de Matemática (2 hrs. 45 min.)
 <b>Talcahuano</b>	<b>Día 1</b>	<b>Día 2</b>
	<b>Desde 8:00 am.</b> Prueba de Matemática (2 hrs. 45 min.) <b>Descanso</b> <b>Aprox. 11:00 am.</b> Pruebas electivas Ciencias (2 hrs. 55 min.) o Historia, Geografía y Ciencias Sociales (2 hrs. 30 min.)	<b>Desde 8:00 am.</b> Prueba de Lenguaje y Comunicación (2 hrs. 30 min.)

#### 4.1 Equipo de aplicación

El equipo de aplicación de pruebas estuvo compuesto, preferentemente, por los docentes de los establecimientos que funcionaron como locales de aplicación. Solo quienes trabajan en preuniversitarios, o hayan sido objetados por su desempeño en procesos anteriores, están inhabilitados para participar. La composición del equipo, en orden jerárquico, se expone en la Figura 5.

Figura 5. Composición del equipo de aplicación



Una de las responsabilidades principales durante la aplicación es resguardar del material y asegurar un correcto desarrollo del proceso. Por lo tanto, el equipo aplicador no está autorizado para abrir, leer o revisar los folletos, usar celulares, cámaras u otro dispositivo electrónico, ni desarrollar otra tarea durante el proceso de aplicación.

Para asegurar el control de calidad de la aplicación, existe una serie de documentos que ayuda a supervisar y monitorear la adecuada realización del Pilotaje. Previo a la aplicación piloto se realizó



una reunión en cada local de aplicación para informar al equipo las reglas del proceso y para establecer los protocolos y responsabilidades de cada uno. Luego de esa reunión se identificaron los lugares relevantes dentro del local (e.g. para procedimientos de evacuación en caso de emergencia y salas de acopio de material), se prepararon las salas para la aplicación, y se informó sobre los procedimientos de emergencia –por ejemplo, en caso de sismo o incendio.

## 4.2 Traslado de material

En la Región Metropolitana, el traslado del material fue responsabilidad de los Coordinadores Técnicos. Los coordinadores se presentaron en el DEMRE para retirar las cajas con el material correspondiente a cada día de aplicación y trasladarlo al local. Una vez finalizado el día de aplicación, el coordinador retornó todo el material aplicado a las dependencias del DEMRE.

En regiones, el material fue despachado con algunos días de anticipación y quedó bajo la custodia de Carabineros de Chile, en una comisaría cercana al(los) local(es) de aplicación. En estos casos, fue responsabilidad del Delegado reunirse con el representante de carabineros a cargo del proceso y coordinar el resguardo y traslado del material a cada local de aplicación. El Delegado también fue el responsable de gestionar el traslado de material con los encargados de cada local. Al interior del local de aplicación, el Coordinador Técnico fue responsable de resguardar correctamente el material.

Los Coordinadores Técnicos recibieron los folletos de las pruebas correspondientes a su local ordenados en lotes agrupados por sala y en cajas selladas. Cada folleto y su respectiva hoja de respuesta fueron entregados en un envase sellado.

## 4.3 Condiciones para rendir las pruebas piloto

El horario de citación de los estudiantes fue a las 8:00 am para ambos días de aplicación. Además del local de aplicación al que debía concurrir cada estudiante, el DEMRE informó a colegios y a estudiantes la sala en la que rendirían las pruebas. Esto facilitó el ingreso e identificación de los estudiantes. Se aceptó el ingreso de estudiantes a las salas hasta media hora después de comenzada la aplicación.

Para ingresar a rendir cada prueba, los estudiantes debieron identificarse con su cédula de identidad u otro documento con fotografía que permitiera establecer su identidad (e.g. comprobante de solicitud de cédula de identidad, pase escolar). Los estudiantes no citados al piloto que se presentaron a rendir las pruebas fueron autorizados a rendir solo si el local contaba con material de aplicación disponible y el estudiante cumplía con tres condiciones:

- Cursar 4° medio,
- Presentar un documento de identificación apropiado
- Pertenecer a alguno de los establecimientos seleccionados para la muestra<sup>5</sup>.

Los estudiantes firmaron un documento donde expresaron su autorización o rechazo a que el DEMRE entregara sus resultados de la prueba piloto a su establecimiento educacional.

---

<sup>5</sup> En el caso de no pertenecer a un establecimiento en la muestra, el Jefe de Local debía comunicarse con el DEMRE para autorizar el ingreso del estudiante.

## 4.4 Casos especiales

Los “casos especiales” corresponden a las siguientes tres situaciones: estudiantes agregados antes de la aplicación, estudiantes con tratamiento especial para la aplicación y estudiantes con cambio de prueba electiva durante la aplicación.

### 4.4.1. *Estudiantes agregados antes de la aplicación*

En algunos casos, los establecimientos solicitaron incluir en la aplicación piloto a estudiantes que no estaban inscritos para rendir la PSU Admisión 2017 y que, por ello, no habían sido convocados a participar del piloto. DEMRE informó estos casos a cada Jefe de local antes de la aplicación. Solo se agregaron estudiantes a locales que contaban con material de reserva para estos estudiantes.

### 4.4.2. *Estudiantes con necesidades especiales*

Por primera vez, el piloto PSU incluyó estudiantes con necesidades especiales para la rendición. En esta categoría se consideraron estudiantes pertenecientes a Programas de Integración Educativa (PIE), en situación de discapacidad u otras situaciones que requerían atención especial.

El DEMRE implementó ajustes que permitieran a estos estudiantes participar de la aplicación piloto, bajo una lógica de inclusión. Esto se realizó considerando los recursos disponibles y la colaboración de los establecimientos para proporcionar condiciones especiales a estos estudiantes.

A diferencia de los trámites y protocolos que deben seguir los casos especiales para la rendición de la PSU oficial<sup>6</sup>, para el piloto solo se solicitó a los establecimientos informar de la situación al DEMRE. Para ello, cada establecimiento envió la nómina de los estudiantes en esta condición y las sugerencias para hacer más cómoda su participación. Dentro de las facilidades otorgadas se cuentan las siguientes:

- Realización de la prueba en sala especial, única o adaptada.
- Modificación del tiempo de rendición (se agregó un 20% por sobre el tiempo de cada prueba) y permitir pausas.
- Ayuda de un tutor o especialista que colaboró en el llenado de datos personales, lectura de la prueba, o marcaje de respuestas. La ayuda recibida consistió en la asistencia directa de dos examinadores externos o de personal especializado proveniente del establecimiento del estudiante. En ese último caso, a los trabajadores de los establecimientos de proveniencia de los estudiantes, se les canceló el honorario de un examinador.
- Modificación de las hojas de respuestas o folletos. A los estudiantes en situación de discapacidad que informaron un déficit visual, se les imprimieron folletos y hojas de respuesta con letra ampliada.
- Uso de lámpara, atril, lupa o lente especial y audífonos.

Todos los estudiantes que recibieron algún tipo de ajuste para la rendición del piloto firmaron una declaración jurada donde declaraban estar en conocimiento de que el ajuste proporcionado para la rendición de las pruebas piloto no era vinculante para la rendición de la PSU en el Proceso Oficial de Admisión 2017.

---

<sup>6</sup> Ver: DEMRE (s.f). “Casos especiales”. En línea, disponible en: <http://psu.demre.cl/inscripcion/casos-especiales> [Consultado el 26 de septiembre de 2016]



En total, 40 estudiantes recibieron tratamiento especial durante la aplicación piloto, tal como se expone en la Figura 6. De ellos, 38 fueron ubicados en sala especial, recibiendo tiempo extra y el apoyo de los examinadores. De los siete estudiantes que estaban en situación de discapacidad visual, tres recibieron folletos y pruebas ampliadas, y a cuatro se les leyó la prueba en voz alta. Por último, 21 estudiantes fueron acompañados de un tutor que colaboró en la lectura, marcaje de respuestas o llenado de datos, si correspondía.

Figura 6. Estudiantes con tratamiento especial, Piloto PSU 2016



#### 4.4.3. Cambios de prueba electiva

En algunos casos los estudiantes solicitaron cambiar la prueba electiva que se les había asignado. Esta solicitud fue atendible siempre y cuando existiese material disponible en el local (por estudiantes ausentes o material de reserva). Los estudiantes pudieron solicitar cambio de prueba

electiva entre Ciencias e Historia, Geografía y Ciencias Sociales –o viceversa–, o dentro de los módulos electivos de Ciencias<sup>7</sup>.

## 5. Resumen de la aplicación

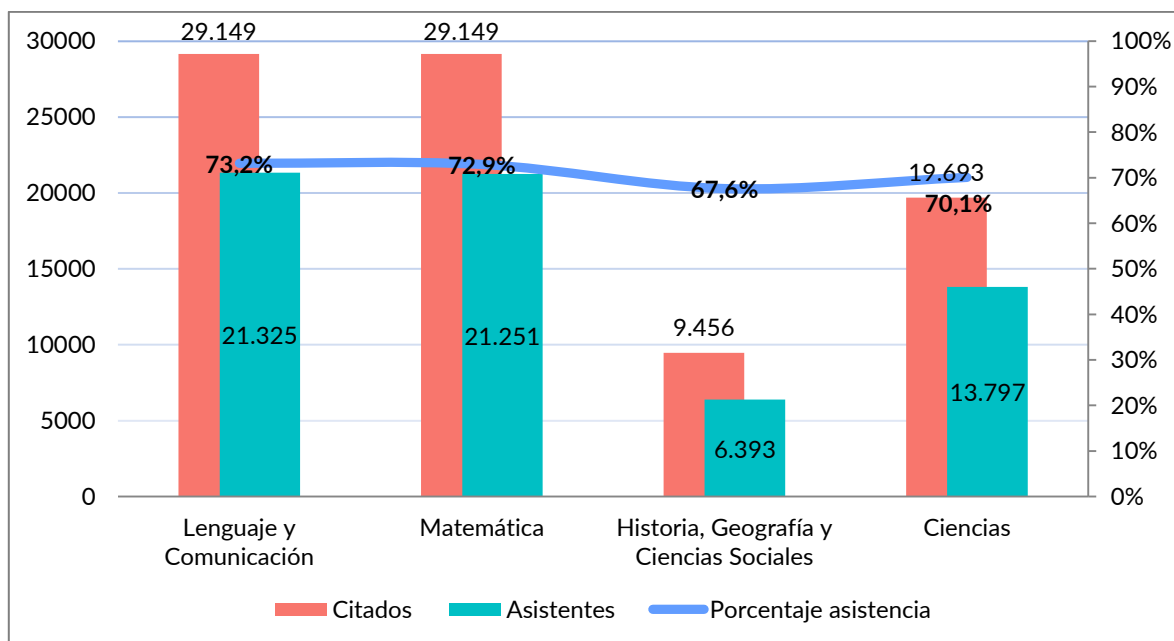
A continuación se presentan datos sobre la asistencia al pilotaje de forma resumida, según variables de contexto y de tipo de establecimiento. En primer lugar, se presenta la asistencia y su porcentaje por prueba. En segundo lugar, se muestra la asistencia en variables relativas a las características de los estudiantes; sexo y región de origen. En tercer lugar, se muestra la asistencia por variables relativas a los establecimientos educacionales de los estudiantes: dependencia administrativa y rama de enseñanza. Por último, se describe la muestra efectiva, es decir la distribución empírica de la muestra definida en la sección 3.3.

### 5.1 Asistencia por prueba

De un total de 29.149 estudiantes citados, 23.511 se presentaron a rendir al menos una prueba. Es decir hubo una tasa de asistencia al Piloto PSU del 80,7%. Este total considera además a 324 estudiantes que no fueron citados, pero que igualmente asistieron a la aplicación y fueron autorizados por el DEMRE para rendir las pruebas.

El Gráfico 2 muestra la cantidad de estudiantes asistentes al piloto respecto de los citados y el porcentaje de asistencia por disciplina. Como se observa, la tasa de asistencia bordeó el 70% en todas las pruebas.

Gráfico 2. Asistencia a Piloto PSU, por prueba

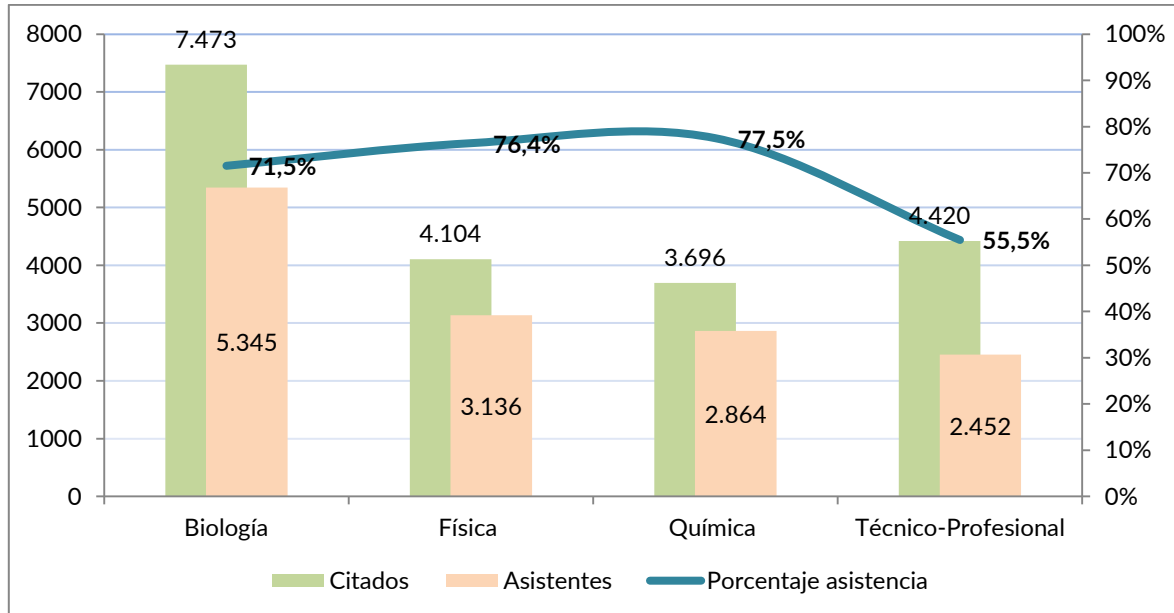


<sup>7</sup> En Ciencias, los estudiantes podían realizar cambios entre los módulos de Biología, Física o Química. Únicamente los estudiantes egresados de establecimientos Técnico-Profesionales podían cambiar entre uno de los módulos antes mencionados y el módulo Técnico-Profesional.



El Gráfico 3 presenta la asistencia a cada módulo electivo que compone la prueba de Ciencias. Tal como en el gráfico anterior, se observa una tasa de asistencia por sobre el 70% en las tres pruebas enfocadas a estudiantes de la modalidad Humanista-Científica. La prueba Técnico-Profesional tuvo una asistencia levemente superior al 50%.

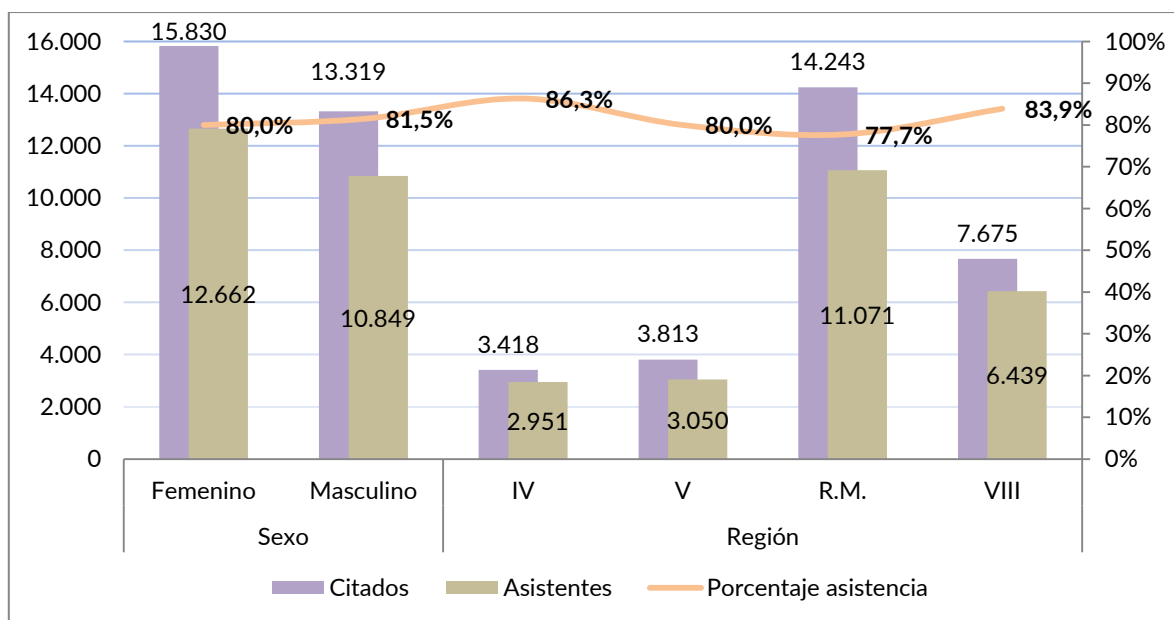
Gráfico 3. Asistencia a Piloto PSU, pruebas de Ciencias



## 5.2 Asistencia por sexo y región

A continuación, el Gráfico 4 presenta las características de los asistentes según sexo y región de origen de los estudiantes. En todas las categorías analizadas la tasa de asistencia fluctúa alrededor del 80%.

Gráfico 4. Asistencia a Piloto PSU, según sexo y región



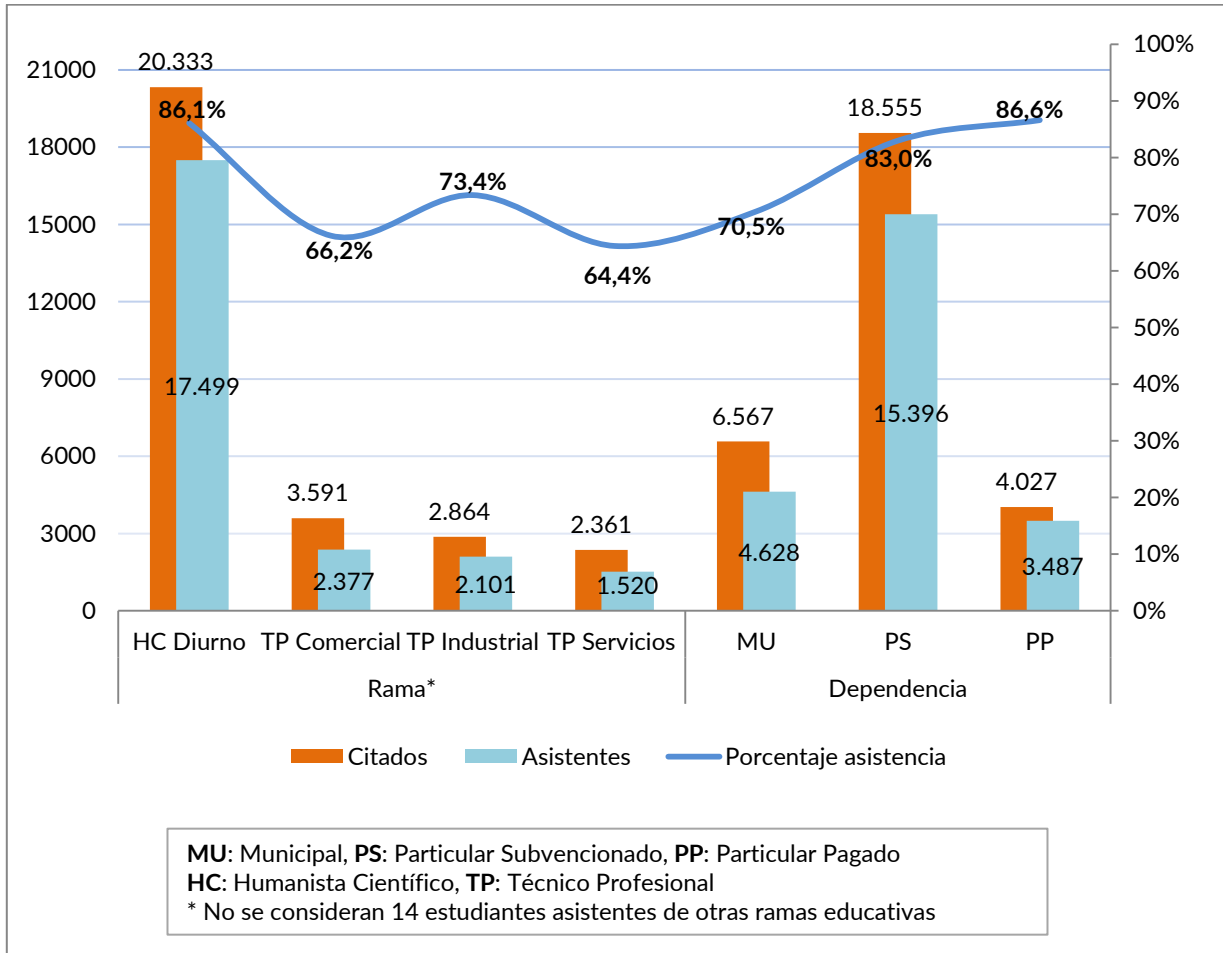
### 5.3 Asistencia por tipo de establecimiento

A continuación, se presentan las características de los establecimientos educacionales de los estudiantes que participaron en el piloto. Tal como lo muestra el Gráfico 5, la asistencia se distribuyó de distinta forma dependiendo del establecimiento educacional de origen.





Gráfico 5. Asistencia a Piloto PSU, según modalidad de enseñanza y dependencia administrativa



Al realizar el análisis por modalidad educativa, se observa que la mayor participación ocurre entre los estudiantes provenientes de establecimientos científico humanista, seguido por aquellos provenientes de establecimientos técnicos de rama industrial. Este resultado podría vincularse a que son los estudiantes HC quienes mayoritariamente se preparan para la PSU, con la expectativa de acceder a la educación superior por esta vía.

Además, otro de los factores que podría incidir en la participación según modalidad es el hecho de que la PSU está basada en los Contenidos Mínimos Obligatorios de la modalidad científico humanista. Estos contenidos no son totalmente cubiertos en la población técnica, la que destina mayor énfasis a la formación diferenciada, conforme a su especialidad (MINEDUC, 2013). No obstante, la participación de la población TP no es tan baja, sino solo comparativamente menor a la HC. Esta situación habla sobre el interés de estos estudiantes de participar en esta instancia y los procesos de admisión en general. Para el Proceso de Admisión 2016, el 27% de los estudiantes inscritos (71.205 estudiantes), pertenecía a la modalidad técnica profesional.

Respecto de la dependencia administrativa de los establecimientos, se observa que son los estudiantes de establecimientos particulares —subvencionado y no subvencionado— quienes tuvieron una mayor tasa de asistencia en el piloto PSU, con sobre el 80%. La Tabla 9 ilustra cómo se comportan las variables de modalidad de enseñanza y dependencia del establecimiento respecto al

ausentismo. La tabla muestra el porcentaje de ausentes por dependencia y rama educativa respecto de los citados en esa misma rama y dependencia.

Tabla 9. Porcentaje de ausentes, según combinaciones de modalidad de enseñanza con dependencia administrativa

			Dependencia		
			Municipal	Particular Subvencionado	Particular Pagado
Modalidad de enseñanza	Científico Humanista Diurno	Citados	3.844	12.462	4.027
		Ausentes	1.015	1.467	540
		Tasa de inasistencia	<b>26,4%</b>	<b>11,8%</b>	<b>13,4%</b>
	Técnico Comercial	Citados	1.025	2.566	0
		Ausentes	402	832	
		Tasa de inasistencia	<b>39,2%</b>	<b>32,4%</b>	
	Técnico Industrial	Citados	802	2.062	0
		Ausentes	226	593	
		Tasa de inasistencia	<b>28,2%</b>	<b>28,8%</b>	
	Técnico Servicios	Citados	896	1.465	0
		Ausentes	424	431	
		Tasa de inasistencia	<b>47,3%</b>	<b>29,4%</b>	

En la tabla se observa que la mayor cantidad de ausentes se concentra en la dependencia municipal, de modalidad técnica en servicios. No obstante, en la misma modalidad, pero de dependencia particular subvencionada, la inasistencia de los estudiantes bordea el promedio. Se observa también, que la menor proporción de estudiantes ausentes se encuentra en la modalidad científico humanista de dependencia privada, tanto particular como subvencionada.

## 5.4 Muestra efectiva

A continuación, la Tabla 10 presenta la composición de la muestra efectiva de estudiantes asistentes al piloto. En la tabla se muestra el total de citados, por cada variable y categoría, y el porcentaje de cada categoría respecto del total de los citados. Lo mismo para los asistentes, se muestra el total y la proporción de cada categoría respecto del total de asistentes. La columna variación muestra la diferencia porcentual, por cada categoría, entre los asistentes y los citados.

Tabla 10. Muestra efectiva según variables de caracterización

Variable	Categoría	Citados		Asistentes		Variación
		Cantidad	% respecto del total de citados	Cantidad	% respecto del total de asistentes	
Rama	HC Diurno	20.333	69,8%	17.499	74,5%	4,7%



	TP Comercial	3.591	12,3%	2.377	10,1%	-2,2%
	TP Industrial	2.864	9,8%	2.101	8,9%	-0,9%
	TP Servicios	2.361	8,1%	1.520	6,5%	-1,6%
<b>Total</b>		<b>29.149</b>	<b>100%</b>	<b>23.497*</b>	<b>100%</b>	
<b>Dependencia</b>	Municipal	6.567	22,5%	4.628	19,7%	-2,8%
	Particular Subvencionado	18.555	63,7%	15.396	65,5%	1,8%
	Particular Pagado	4.027	13,8%	3.487	14,8%	1%
<b>Total</b>		<b>29.149</b>	<b>100%</b>	<b>23.511</b>	<b>100%</b>	
<b>Sexo</b>	Femenino	15.830	54,3%	12.662	53,9%	-0,5%
	Masculino	13.319	45,7%	10.849	46,1%	0,5%
<b>Total</b>		<b>29.149</b>	<b>100%</b>	<b>23.511</b>	<b>100%</b>	
<b>Región</b>	IV	3.418	11,7%	2.951	12,6%	0,8%
	V	3.813	13,1%	3.050	13%	-0,1%
	R.M.	14.243	48,9%	11.071	47,1%	-1,8%
	VIII	7.675	26,3%	6.439	27,4%	1,1%
<b>Total</b>		<b>29.149</b>	<b>100%</b>	<b>23.511</b>	<b>100%</b>	

\* No se consideran 14 estudiantes asistentes de otras ramas de enseñanza

Como se observa en la tabla precedente, la mayor variación se encuentra en los establecimientos HC diurnos, los cuales se encuentran levemente sobre representados, en detrimento de los establecimientos técnicos, los que se encuentran levemente sub representados. Otra diferencia se observa en la dependencia municipal con una leve sub representación de estudiantes en la muestra efectiva. La variación en el resto de las categorías es bastante baja, por lo tanto, podemos concluir que la distribución de la muestra efectiva es muy cercana a la muestra teórica.



# Análisis



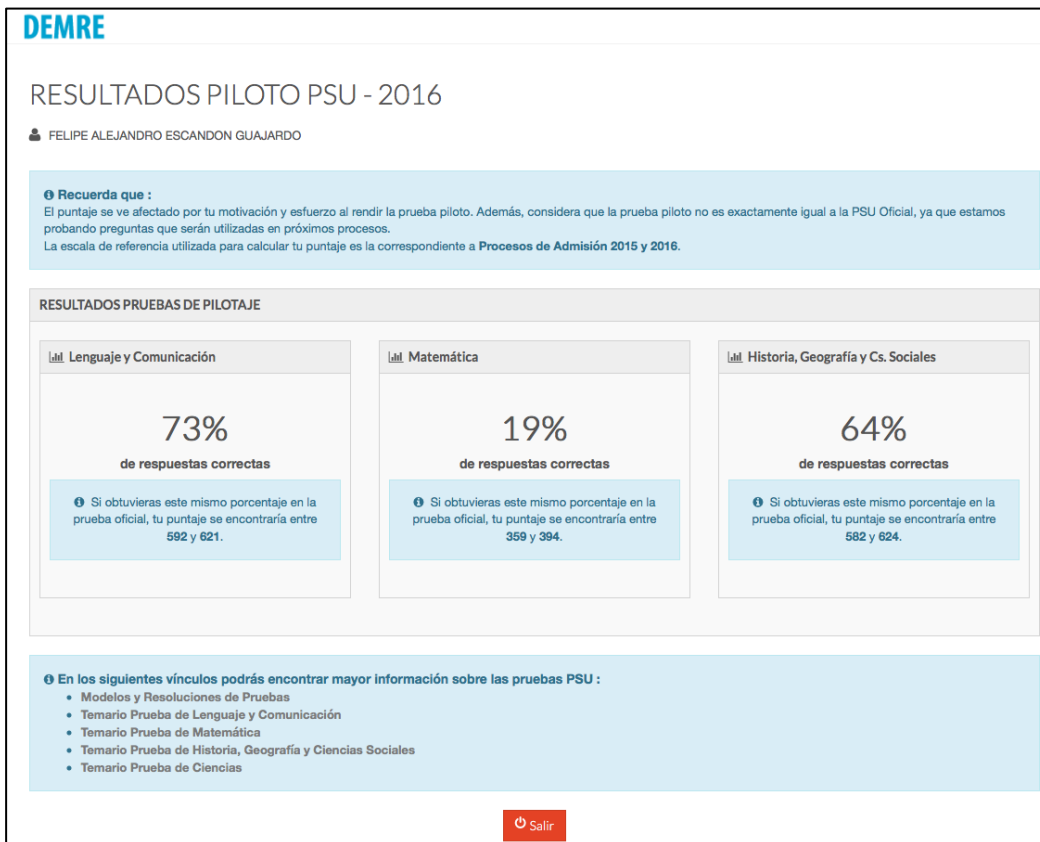
## 6. Informes a establecimientos educacionales y estudiantes

Tras obtener los resultados del piloto para cada estudiante (porcentaje de respuestas correctas), el DEMRE generó reportes individuales y agregados de desempeño. Los primeros fueron enviados a los estudiantes y los reportes agregados fueron enviados a los establecimientos educacionales.

El reporte individual presentó el porcentaje de respuestas correctas obtenido en cada prueba rendida por el estudiante. Dado que en la aplicación piloto se testean preguntas para utilizar en aplicaciones futuras, no es posible reportar puntajes con la precisión de una aplicación oficial. Sin embargo, el DEMRE entiende la necesidad de los estudiantes por contar con este tipo de información.

Para entregar información técnicamente correcta y de utilidad a los estudiantes, se entregó un rango de puntaje PSU de referencia asociado al porcentaje de respuestas correctas obtenido. Estos rangos de puntaje fueron calculados en referencia a la escala de puntajes de los años 2014 y 2015 para ese porcentaje de respuestas correctas. Para visualizar los resultados, se envió un correo electrónico a cada estudiante que contenía un link a la plataforma de resultados. Un ejemplo de la pantalla visualizada por los estudiantes se encuentra en la Figura 7.

Figura 7. Plataforma entrega de resultados Aplicación Piloto a estudiantes



El reporte a nivel agregado presentó los resultados por prueba, a nivel nacional, regional, comunal y de cada establecimiento. Además, se entregó como referencia, el resultado global obtenido por establecimientos educacionales con características similares. Por último, se entregaron los resultados del establecimiento en cada prueba, según áreas/ejes temáticos y habilidades cognitivas. Un ejemplo de Informe se encuentra en el Anexo 6. Cada establecimiento recibió, además, una nómina con los resultados individuales de aquellos estudiantes que accedieron a entregar sus datos mediante la firma del consentimiento informado explicado en la sección 4.3.

Como las preguntas del pilotaje estaban probándose, los resultados debían tratarse e interpretarse con sumo cuidado. Con este objetivo, el DEMRE especificó que los en el piloto responden a la motivación y esfuerzo que el/la estudiante dedicó al responder cada prueba. Además, se enfatizó que dado que la prueba piloto no es igual a la PSU oficial, los puntajes debían usarse de modo referencial.

Sobre los resultados entregados a los establecimientos —y por los mismos argumentos ya mencionados—, estos no debían usarse como indicador de desempeño de los estudiantes. Tal como con los resultados de la PSU oficial, este tipo de evaluaciones no pueden dar cuenta de todo lo que se hace en la escuela, ni menos deberían utilizarse para clasificar establecimientos o generar rankings (Sánchez, 2016). Además, debido a los objetivos de la PSU y su pilotaje, estos resultados no hablan de calidad, puesto que no se evalúan todas las asignaturas ni contenidos cubiertos en la educación media.

## 7. Análisis de resultados

Teniendo los datos y los resultados de las pruebas, la Unidad de Desarrollo y Análisis del DEMRE realizó los análisis psicométricos para determinar la calidad de los ítems y por tanto, la factibilidad de estos para ser utilizados en pruebas oficiales. A continuación, se describen los distintos análisis psicométricos realizados a las Pruebas Piloto. Se exponen los análisis, las formulas y procedimientos para llevarlos a cabo. A su vez, se muestran los criterios definidos por el DEMRE que aseguran que los ítems de la PSU cumplen con los estándares de calidad para una prueba estandarizada de altas consecuencias.

En la evaluación de calidad de los ítems se utiliza la Teoría Clásica del Test (CTT, *Classic Theory of Tests*) y se complementa este análisis con el modelo de Teoría de Respuesta al Ítem (IRT, *Item Response Theory*). De esta forma, es posible establecer *modelos* capaces de evaluar las *propiedades psicométricas* de los instrumentos de medición. Específicamente, se estudian aquellos factores que influyen sobre las puntuaciones obtenidas en los test y sus ítems, proponiendo modelos que permitan controlar y minimizar los factores de error. Estos factores de error inciden en las estimaciones realizadas a partir del instrumento de medición.

El concepto *propiedades psicométricas* refiere al análisis de las características métricas del ítem, que dan cuenta de la idoneidad del instrumento para medir lo que se desea medir, minimizando el error. En un sentido amplio, lo anterior puede ser definido como un *proceso* que se centra en el análisis del instrumento, en los siguientes tres niveles (Muñiz, Teoría Clásica de los Test, 2003):



1. Respecto de su comportamiento en tanto escala: refiere al estudio de la confiabilidad y validez del instrumento<sup>8</sup>.
2. Respecto de las características de sus ítems: se orienta a los análisis de las características propias de cada ítem, tales como su dificultad, discriminación, omisión, comportamiento de los distractores y funcionamiento diferencial del ítem.
3. Una combinación de ambas.

La evaluación de los ítems, tiene por objetivo apoyar el proceso de ensamblaje de las pruebas PSU, que se realiza con ítems previamente validados en una muestra representativa de la población. Este proceso de pilotaje asegura que los ítems utilizados cumplen los estándares mínimos y suficientes para asegurar la calidad técnica del instrumento.

## 7.1 Teoría Clásica del Test (CTT)

En cuanto a su formulación general, la CTT propone un modelo lineal en el que se asume que la puntuación obtenida por el sujeto  $i$  en un test ( $X$  o puntuación empírica) se compone de dos elementos aditivos: la puntuación verdadera obtenida por el sujeto ( $V$ ) y el error de medida presente en las puntuaciones observadas ( $e$ ). Formalmente, lo anterior queda definido en la Ecuación 1.

Ecuación 1.

$$X_i = V_i + e_i$$

La CTT ha formulado distintos criterios de valoración de la calidad de los ítems, entre los que destacan por su utilidad los siguientes: índice de dificultad, índice de discriminación y análisis de los distractores.

### 7.1.1. Índice de Dificultad

El *índice de dificultad* de un ítem ( $p$ ) se define como la proporción de sujetos que responde correctamente al mismo, en función del total de individuos que abordaron el ítem. Definido en la Ecuación 2.

Ecuación 2.

$$p_j = \frac{\sum_{i=1}^N A_{ij}}{N}$$

- $p_j$ : dificultad del ítem  $j$ .
- $A_{ij}$ : respuesta de persona  $i$  para el ítem  $j$ .  $A$  toma valor 1 si la respuesta es correcta y 0 si es incorrecta.
- $N$ : número de individuos que respondieron el ítem  $j$ .

El índice de dificultad de un ítem ( $p$ ) admite valores dentro de un intervalo que va de 0 a 1. Cuando  $p$  se acerca al valor 1, indica que muchos individuos han contestado correctamente el ítem, por lo que

---

<sup>8</sup> Los análisis de confiabilidad y validez se realizan a los resultados de la Prueba Oficial.

este resulta fácil. Por el contrario, a medida que  $p$  se aproxima o alcanza el valor 0, indica que el ítem en cuestión resulta difícil para los sujetos de la muestra o población en que fue aplicado. La Tabla 11 resume los puntos de corte utilizados por el DEMRE para valorar un ítem como fácil, mediano o difícil.

Tabla 11. Parámetros de dificultad

$p$	Clasificación
0,00 - 0,39	Difícil
0,40 - 0,59	Mediano
0,60 - 1,00	Fácil

Para efectos de los ítems que componen la batería de pruebas PSU, el índice de dificultad utilizado por el DEMRE se encuentra en los rangos  $0,10 \leq p \leq 0,80$ .

### 7.1.2. Índice de Discriminación

De manera amplia, el *índice de discriminación* puede ser definido como la correlación que se establece entre las puntuaciones que obtienen los sujetos en un ítem particular y la puntuación total en el test.

Según Muñiz *et. al.* (2005), una pregunta tiene poder de discriminación si es capaz de distinguir entre los sujetos que puntúan alto en una prueba de aquellos que puntúan bajo. Por lo tanto, “es condición de calidad de un ítem el que sea contestado correctamente por los estudiantes con mayor puntuación” (pág. 61).

El DEMRE establece los índices de discriminación de los ítems que componen los instrumentos de la batería de pruebas PSU, por medio de correlaciones. Específicamente, dadas las características de los ítems de selección múltiple y los requerimientos de la CTT, se utiliza el índice de correlación biserial ( $r_b$ ). Este permite relacionar respuestas de tipo dicotómicas y discretas (acierto versus no acierto), con una escala de tipo continua (puntuación total sobre la escala o prueba), evaluando así el grado de asociación y, por extensión, de discriminación de los ítems (Ecuación 3).

Ecuación 3.

$$rb_j = \frac{\bar{x}_{cj} - \bar{x}_t}{s_t} * \frac{p_j}{y}$$

- $\bar{x}_c$ : promedio de puntuación en la prueba del grupo que contesta correctamente el ítem  $j$ .
- $\bar{x}_t$ : promedio de puntuación en la prueba del grupo total.
- $s_t$ : desviación estándar del grupo total.
- $p_j$ : proporción de sujetos que contesta correctamente el ítem  $j$ .
- $y$ : ordenada correspondiente al valor de la puntuación típica ( $z$ ) que deja por debajo un área igual a  $p_j$ .





Los criterios internacionales para clasificar un índice de correlación biserial, son expuestos por Muñiz, *et al.* (2005). En la Tabla 12 se muestran los puntos de corte utilizados para clasificar el índice de correlación biserial ( $r_b$ ).

Tabla 12. Clasificación del índice de correlación biserial

$r_b$	Clasificación del ítem
Igual o mayor que 0,40	Discrimina muy bien
Entre 0,30 y 0,39	Discrimina bien
Entre 0,20 y 0,29	Discrimina poco
Entre 0,10 y 0,19	Limite. Se debe mejorar
Menor de 0,10	Carece de utilidad para discriminar

### 7.1.3. Análisis de las opciones incorrectas o “distractores”

En un ítem, se denomina opción incorrecta o distractor a sus opciones incorrectas de respuesta. Como señalan Muñiz *et al.* (2005), es fundamental que todas las opciones incorrectas incluidas, en tanto opciones de respuesta al ítem, resulten “(...) igualmente atractivas para las personas evaluadas que desconocieran la respuesta correcta” (pág. 70). Analizar la distribución de las respuestas de los examinados, explica el funcionamiento de los distractores.

Por ejemplo, en un ítem, un índice de discriminación bajo puede explicarse porque alguno de los distractores fue elegido tanto por los individuos con bajo desempeño como por los de alto desempeño. En este caso, es conveniente cambiar dicho distractor por uno más adecuado y volver a pilotear el ítem. Además, es posible que algún distractor no sea elegido por los examinados (lo que se denomina *distractor vacío*), lo que también afecta el poder discriminativo del ítem.

## 7.2 Teoría de Respuesta al Ítem (IRT)

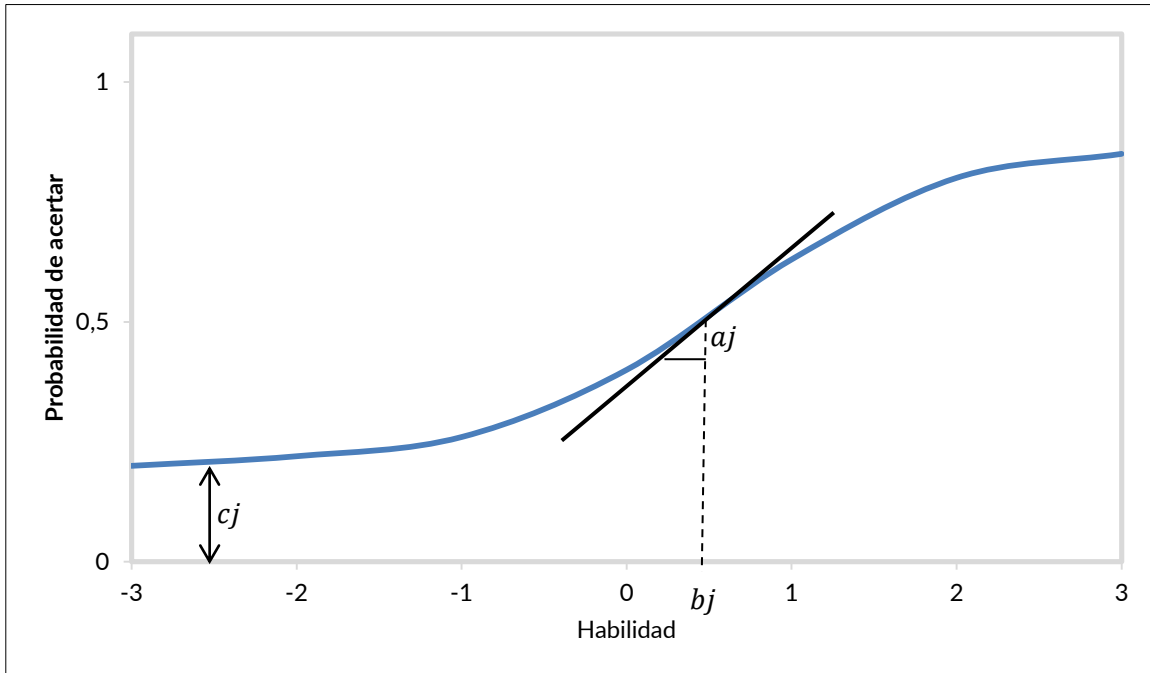
La IRT, constituye un nuevo enfoque dentro de la psicometría, cuyo objetivo es resolver problemáticas que escapan del alcance de la CTT. Tal como su nombre lo indica, el enfoque IRT se basa en las propiedades de los ítems más que en el test en términos globales.

La mayor contribución de este enfoque es que permite obtener mediciones invariantes respecto de los instrumentos utilizados y los sujetos implicados. Es decir, con el uso de IRT se obtendrán mediciones que no cambian en función del test usado e instrumentos de medida con propiedades independientes de los sujetos evaluados (Muñiz, 1997).

IRT busca estimar las características de los ítems, y además, estimar la probabilidad de que un estudiante responda correctamente. Para esto, los modelos IRT asumen la existencia de una relación funcional entre los valores de la variable que miden los ítems (dicho de otra forma, las puntuaciones de los sujetos en la variable medida) y la probabilidad de acertarlos, denominando a dicha función “Curva Característica del Ítem” (CCI).

En el Gráfico 6, se observa la CCI. En el eje de abscisas se representan los valores de la variable que mide habilidad necesaria para responder correctamente un ítem, la que se denota como  $\theta$ . En el eje de ordenadas se representa la probabilidad de responder correctamente el ítem, la que se denota como  $P(\theta)$ . Como se observa, sujetos con mayor habilidad, poseen mayores probabilidades de acertar el ítem.

Gráfico 6. Curva Característica del Ítem



Fuente: Elaboración propia

La CCI se compone de tres parámetros, los cuales pueden analizarse en conjunto o por separado. Si bien son los mismos criterios usados en CTT, su significado no es exactamente el mismo. El parámetro  $b_j$ , o índice de dificultad del ítem  $j$ , es el valor de  $\theta$  correspondiente al punto de máxima pendiente de la CCI, por lo tanto, la dificultad se mide en la misma escala de habilidad ( $\theta$ ). Por su parte, el parámetro  $a_j$  es el índice de discriminación del ítem  $j$  y su valor es proporcional a la pendiente de la recta tangente a la CCI en  $b_j$ . Es decir, cuanto mayor sea la pendiente, mayor será el índice de discriminación. Por último, el parámetro  $c_j$  es la aproximación a la probabilidad de acertar el ítem  $j$  al azar.

### 7.3 Procedimiento de análisis

Tras explicitar los modelos principales que se utilizan en los análisis psicométricos, esta sección describe el procedimiento realizado por el DEMRE para analizar, aceptar y rechazar preguntas en el Piloto PSU 2016.

Previo al análisis, se eliminan los casos de estudiantes que omitieron toda la prueba. Además, se identifican ítems cuyos distractores tienen un biserial superior a 0,200, estos se envían a los comités constructores para que verifiquen las claves y/o que no haya un error de impresión. Posteriormente, la primera fase, que llamaremos de “purificación del instrumento”, busca obtener resultados más



limpios y robustos. Para cada una de las pruebas, se analizan todas las preguntas a través de una correlación puntual con análisis de CTT. La correlación se realiza con el porcentaje de respuesta correcta de cada individuo que contesta el ítem. Tras el cálculo, se procede a eliminar todos los ítems biseriales puntuales son menores a 0,150.

La segunda fase se realiza con las preguntas aceptadas tras la purificación. En esta fase, se implementa un análisis IRT de un parámetro, específicamente el modelo de Rasch. La principal característica de este modelo es que los ítems solo difieren en el parámetro de dificultad, asumiendo que la discriminación para todos los ítems es igual a 1. La forma tradicional de estimar el modelo de Rasch, se presenta en la Ecuación 4.

Ecuación 4.

$$P_j(\theta) = \frac{1}{1 + e^{-(\theta - b_j)}}$$

En el análisis IRT, se calcula la *habilidad individual* de cada sujeto, en vez de calcular un puntaje propiamente tal. Para cada sujeto, esta habilidad se calcula sumando todas las probabilidades de responder correctamente (cada ítem). Es decir, si una prueba tiene 75 preguntas, la habilidad individual de una persona se calcula sumando 75 probabilidades de responder correctamente.

El cálculo de esta habilidad individual es necesario para la tercera fase, que consiste en analizar los resultados psicométricos de los ítems piloteados. Para la aceptación o rechazo de ítems de pilotaje, el principal criterio utilizado es el índice de discriminación calculado mediante CTT. Esta vez, la correlación no se realiza con el porcentaje de respuesta correcta de cada individuo, sino con el puntaje IRT (o habilidad individual), calculado previamente. Los criterios para aceptar los ítems se exponen en la Tabla 13.

Tabla 13. Condiciones para la aceptación de ítems

Biserial de la clave	Biserial de los distractores	Diferencia entre el biserial de la clave y los biseriales de distractores	Porcentaje de respuesta en los distractores	Característica del biserial de la clave, en el grupo superior (25% de sujetos con mayor habilidad)
≥ 0,250	≤ 0,200	≥ 0,150	> 3%	$rb_{clave} \geq rb_{distractores}$

Si sucede alguna de las situaciones planteadas en la Tabla 14, el ítem debe rechazarse. Cualquier otra situación diferente a las que aparecen en la Tabla 13 o 14, lleva al ítem al estado de “revisar”, lo que implica una mirada extensiva no solo de los estadísticos, sino también a sus componentes (como formato, contenido, clave, etc.).

Tabla 14. Condiciones para el rechazo de ítems

<b>Si se cumple una de las siguientes situaciones:</b>
Biserial clave ( $rb_{clave}$ ) < 0,225
Respuesta correcta > 90%
Respuesta correcta < 10%
Más de un distractor con respuesta seleccionada < 3%

Biserial clave ( $r_{b_{clave}}$ ) - Biserial distractor ( $r_{b_{distractor}}$ ) > 0,150

«Rechazar clave», si el biserial de algún distractor ( $r_{b_{distractor}}$ ) > 0,200.

«Distractor vacío», si:

- Cumple criterios de aceptación expuestos en Tabla 13 excepto porcentaje de respuesta >3%
- Respuesta de un distractor  $\leq$  3%

## 7.4 Funcionamiento Diferencial del Ítem (DIF)

Además del análisis CTT e IRT realizado a los ítems piloto, se realiza un estudio sobre el funcionamiento diferencial de los ítems (DIF, *Differential Item Functioning*). Este análisis DIF busca detectar posibles *sesgos* analizando la equivalencia entre grupos comparables de individuos que rinden la prueba. Dado que un instrumento de medición no debe estar afectado, en su función de medir, por las características del objeto de medida.

Desde la perspectiva CTT, se dice que un ítem funciona diferencialmente cuando examinados de igual nivel en la variable medida por el test, pertenecientes a diferentes grupos, tienen distinta probabilidad de resolverlo correctamente. Si un ítem no presenta DIF, implica que no hay sesgo, pero si el ítem presenta DIF existen dos posibles causas. Esto puede ocurrir por las diferencias reales que existen entre los grupos en el rasgo subyacente, llamado impacto, o porque el ítem está sesgado. Una de las investigaciones ante la presencia de DIF debe ser un análisis de contenido por parte de expertos en la materia, ya que es imprescindible estudiar las causas y encontrar una explicación teórica de la ocurrencia del mismo.

El proceso inicial para analizar el funcionamiento diferencial de los ítems toma como punto de referencia las variables que se consideran susceptibles de diferencias. Cada variable se categoriza en dos grupos diferentes: *grupo focal* y *grupo referencial*. Es arbitrario establecer la categorización de cada grupo, pero suele reservarse el término focal para el grupo minoritario o que, a priori, se considera posiblemente perjudicado por alguno de los ítems (Muñiz, 2003). Para el caso del análisis del pilotaje de la PSU, en la Tabla 15 se muestran las variables y grupos analizados que – considerando la realidad nacional- podrían presentar DIF.

Tabla 15. Variables y grupos considerados para el análisis DIF

Variable	Grupo Focal	Grupo Referencial
Sexo	Femenino	Masculino
Dependencia	Municipal	Particular Subvencionado
	Particular Subvencionado	Particular Pagado
	Municipal	Particular Pagado
Modalidad	Técnico-Profesional	Humanista-Científico
Zona	Norte (regiones: XV, I a VI)	Metropolitana
	Sur (regiones VII a XIV)	Norte (regiones: XV, I a VI)



Sur (regiones VII a XIV)

Metropolitana

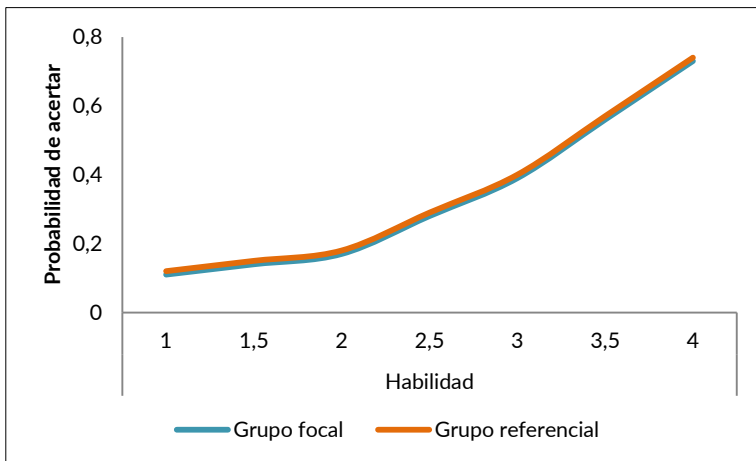
Para el análisis, se utilizan métodos para detectar comportamiento diferencial uniforme y no uniforme. Es decir, para evidenciar si existen diferencias de probabilidad de respuesta correcta y si esta probabilidad es constante o no entre los grupos estudiados.

#### 7.4.1. Funcionamiento Diferencial del Ítem uniforme y no uniforme

Como se mencionó, un ítem presenta un funcionamiento diferencial cuando la probabilidad de ser resuelto correctamente por los estudiantes que poseen el mismo nivel de habilidad varía en función de su grupo de pertenencia. Esto es observable cuando la curva de acierto de un ítem es diferente para distintas poblaciones (Moreira-Mora, 2008).

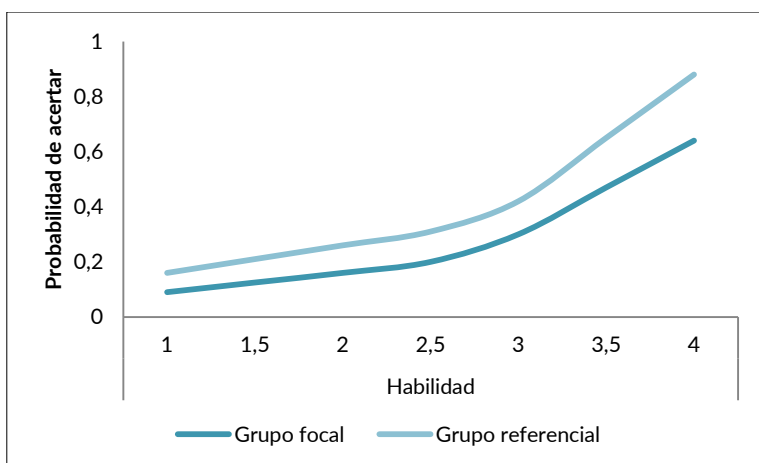
Ahora, la forma de esta curva varía dependiendo de la naturaleza del DIF. A modo de ejemplo, el Gráfico 7 muestra la curva de un Ítem sin DIF. Como se observa, no hay brechas entre los grupos y a medida que aumenta la habilidad, aumenta también la probabilidad de acertar el ítem.

Gráfico 7. Ejemplo de ítem sin DIF



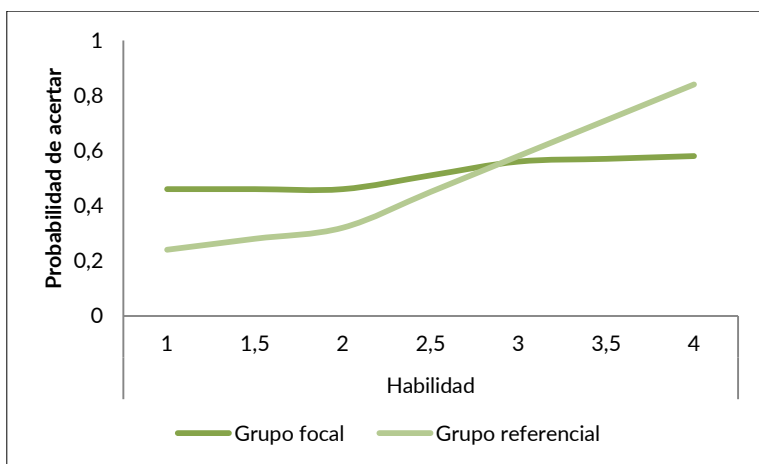
En contraposición, cuando no existe interacción entre el nivel del atributo medido y la pertenencia a un determinado grupo, se presenta funcionamiento diferencial *uniforme*. Es decir, cuando el ítem proporciona una ventaja constante para el mismo grupo, de un extremo a otro en el rango de desempeño (Penfield & Camilli, 2006). En este caso, las curvas de acierto del ítem son paralelas tal como lo muestra el Gráfico 8. En el gráfico, se observa que la curva del grupo referencial, se encuentra por sobre la curva del grupo focal, lo que indica que el ítem es más fácil para el primer grupo. La brecha entre ambas curvas supone que el ítem presenta DIF, ya que para una misma habilidad, la probabilidad de responder correctamente es siempre superior para el grupo de referencia.

Gráfico 8. Ejemplo de ítem con DIF uniforme



Por otra parte, el funcionamiento diferencial *no uniforme* se observa cuando la diferencia en las probabilidades de responder correctamente al ítem no es la misma a lo largo del *continuum* del atributo medido (Moreira-Mora, 2008) y, por consiguiente, las curvas no son totalmente paralelas. Dentro del DIF no uniforme existe una modalidad conocida como *crossing*, expuesta en el Gráfico 9, donde el grupo focal tiene ventaja en el extremo de menor habilidad, mientras que el referencial posee ventaja en el extremo de mayor habilidad.

Gráfico 9. Ejemplo de ítem con DIF no uniforme



Con el fin de detectar DIF uniforme, se empleó el método de Mantel-Haenszel. Se calculan, además, estimadores que determinan si el ítem favorece al grupo focal o referencial, e indican la magnitud de las diferencias entre ellos. En lo que respecta a magnitud, se utiliza la clasificación DIF promovida por el ETS. Finalmente, se calcula el estadístico de Breslow-Day, que es efectivo cuando existen diferencias no uniformes en los niveles de habilidad de los grupos. Todos estos cálculos, que se describen a continuación, fueron complementados con análisis de Regresión Logística.

#### 7.4.2. Mantel-Haenszel

La lógica que subyace al procedimiento Mantel-Haenszel (en adelante “MH”) es la siguiente: si el ítem no presenta DIF, la razón entre el número de personas que aciertan el ítem y las que lo fallan debe ser la misma en los dos grupos comparados a lo largo de todos los niveles de puntuación (Pérez



Gil, 2003). De tal modo, el método MH, distribuye los datos de los grupos en tantas tablas de contingencia como rangos de habilidad de los sujetos se definan para el análisis, con el propósito de comparar las probabilidades de acierto de un ítem (ver Tabla 16). Por lo tanto, el análisis se basa en  $K$  tablas de contingencia.

Tabla 16. Frecuencias absolutas y marginales de grupos en el nivel  $k$ .

Grupos	Tipo de respuesta		
	Aciertos (1)	Errores (0)	Marginales
Grupo de Referencia (R)	$A_k$	$B_k$	$n_{Rk}$
Grupo Focal (F)	$C_k$	$D_k$	$n_{Fk}$
Marginales	$n_{1k}$	$n_{0k}$	$N_{kk}$

- $A_k$ : Es la frecuencia absoluta del grupo referencial que acierta el ítem para el nivel  $k$ .
- $B_k$ : Es la frecuencia absoluta del grupo referencial que no acierta el ítem para la nivel  $k$ .
- $C_k$ : Es la frecuencia absoluta del grupo focal que acierta el ítem para el nivel  $k$ .
- $D_k$ : Es la frecuencia absoluta del grupo focal que no acierta el ítem para el nivel  $k$ .
- $n_{Rk}$ : Cantidad de individuos del grupo referencial para el nivel  $k$ .
- $n_{Fk}$ : Cantidad de individuos del grupo focal para el nivel  $k$ .
- $n_{1k}$ : Cantidad de individuos que acierta el ítem para el nivel  $k$ .
- $n_{0k}$ : Cantidad de individuos que no acierta el ítem para el nivel  $k$ .
- $N_k$ : Número total de la muestra.
- $k = 1, \dots, K$

Con el fin de probar la hipótesis nula, correspondiente a la ausencia de DIF, se postula que la proporción de respuesta correcta del grupo referencial y focal es el mismo para cada nivel de habilidad  $k$ . Mientras, la hipótesis alternativa indica que al menos en un nivel es distinto y por tanto, hay presencia de DIF. Esta hipótesis nula se somete a prueba mediante el estadístico MH, asociado a una prueba de significación, que distribuye según una  $\chi^2$  (Chi-cuadrado) con un grado de libertad descrito en la Ecuación 5.

Ecuación 5.

$$\chi_{MH}^2 = \frac{(|\sum_{k=1}^K A_k - \sum_{k=1}^K E(A_k)| - 0,5)^2}{\sum_{k=1}^K s^2(A_k)}$$

- $\sum_k A_k$  : es la suma de los aciertos del grupo referencial de cada una de los niveles j.
- $\sum_k E(A_k)$  : es la suma de las esperanzas matemáticas de  $A_k$  e igual a  $n_{Rk}n_{1k}/N_k$ .
- $\sum_k s^2(A_k)$  : es la suma de las varianzas de  $A_k$  e igual a  $n_{Rk}n_{Fk}n_{1k}n_{0k}/N_k^2(N_k - 1)$

Para rechazar o no la hipótesis nula, el DEMRE utiliza un nivel de significación ( $\alpha$ ) al 2,5%. Específicamente, si el estadístico de MH ( $\chi^2_{MH}$ ) es mayor que una  $\chi^2_{0.975,1}$  equivalente a 5,02389, se rechaza la hipótesis nula. Por consiguiente, existe evidencia estadística significativa para afirmar que el ítem analizado posee DIF.

El método MH, además, proporciona un estimador numérico que indica la dirección de las diferencias encontradas, es decir, cuál es el grupo favorecido cuando existe un funcionamiento diferencial. El estimador es el coeficiente  $\hat{\alpha}_{MH}$ , cuyos valores oscilan entre cero e infinito. Con el fin de obtener una forma más práctica de interpretación, se propone una transformación logarítmica del coeficiente  $\hat{\alpha}_{MH}$  (Holland & Thayer, 1985) a una escala simétrica con origen en cero, señalada en la **¡Error! No se encuentra el origen de la referencia..** En esta escala, un valor negativo indica que el ítem favorece al grupo de referencia, mientras que un valor positivo, al grupo focal. De igual manera, un valor igual a cero indica ausencia de DIF.

Ecuación 7.

$$\Delta_{MH} = -2,35 \ln(\hat{\alpha}_{MH})$$

Donde el estimador  $\hat{\alpha}_{MH}$  es el señalado en la Ecuación .

Ecuación 8.

$$\hat{\alpha}_{MH} = \frac{\sum_k \frac{A_k D_k}{N_k}}{\sum_k \frac{B_k C_k}{N_k}}$$

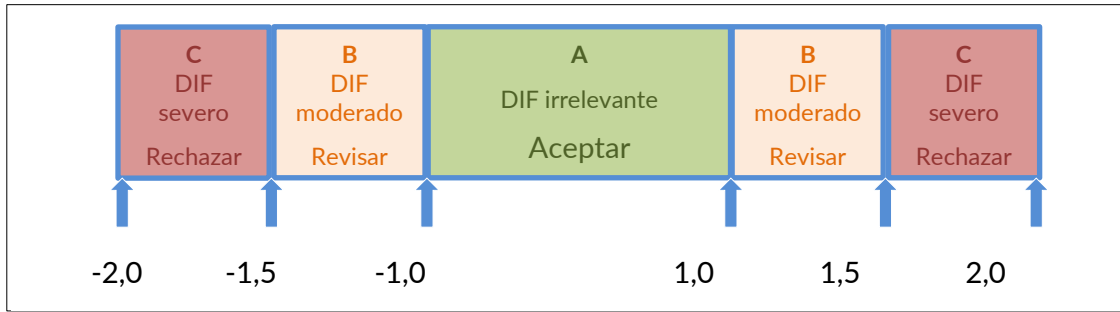
Complementariamente a lo señalado, el ETS (Zwick, 2012) propuso una escala jerárquica para los distintos valores del coeficiente  $\Delta_{MH}$  el cual indica la magnitud de las diferencias.

- $|\Delta_{MH}| < 1$  : es un ítem Categoría A, considerado con DIF *Despreciable o Irrelevante*.
- $1 \leq |\Delta_{MH}| < 1,5$  : es un ítem Categoría B, considerado con DIF *Moderado*.
- $|\Delta_{MH}| \geq 1,5$  : es un ítem Categoría C, considerado con DIF *Severo*.

Según estas categorizaciones, se sugieren ciertas decisiones asociadas al valor del estadístico, lo que se grafica en la Figura 8.



Figura 8. Rangos de la magnitud  $\Delta MH$  de un ítem DIF



### 7.4.3. Breslow Day

La prueba de Breslow-Day (BD) se utiliza para detectar DIF no uniforme (Camilli & Shepard, 1994). Esta prueba determina si la asociación entre la respuesta del ítem y los grupos es homogénea sobre el rango de valores del total de los puntajes. Basado en una distribución Chi-cuadrado con un grado de libertad y con la misma notación usada en el método Mantel-Haenszel (Penfield R. , 2003), el estadístico es el señalado en la Ecuación 6.

Ecuación 6.

$$BD = \frac{[\sum_k X_k (A_k - a_k)]^2}{\sum_{k=1}^K X_k S^2(A_k) - \frac{[\sum_{k=1}^K X_k S^2(A_k)]^2}{S^2(A_k)}}$$

- $a_k = \frac{n_{1k} - n_{Fk} - \psi(n_{1k} + n_{Rk}) \pm \sqrt{(n_{Fk} - n_{1k} + \psi n_{1k} + \psi n_{1k})^2 + 4(1 - \psi)(n_{Rk} n_{1k} \psi)}}{2(1 - \psi)}$
- $\psi = \frac{a_k(n_{Fk} - n_{1k} + a_k)}{(n_{Rk} - a_k)(n_{1k} - a_k)}$
- $S^2(A_k) = \left( \frac{1}{a_k} + \frac{1}{n_{Rk} - a_k} + \frac{1}{n_{1k} - a_k} + \frac{1}{n_{Fk} - n_{1k} + a_k} \right)^{-1}$

Del mismo modo que para el estadístico Mantel-Haenszel, el DEMRE prueba la hipótesis nula con un nivel de significación ( $\alpha$ ) al 2,5%. Específicamente, si el estadístico BD es mayor que una Chi-cuadrado equivalente a 5,02389, se rechaza la hipótesis nula. En efecto, existe evidencia estadística significativa para afirmar que el ítem analizado posee DIF no uniforme.

### 7.4.4. Regresión logística

Además de la detección de DIF mediante MH y BD, el análisis de funcionamiento diferencial se complementa con regresión logística. La regresión logística (RL) es un modelo lineal generalizado que, a través de una función, pretende explicar y predecir los valores de una variable dependiente dicotómica, a partir de los valores conocidos de una o varias variables independientes.

Mediante esta técnica, se busca determinar si en la función matemática necesaria para predecir las respuestas dicotómicas a un ítem (correcta o errada), se debe incluir un término referido al nivel de habilidad de los sujetos independiente de su grupo de pertenencia, o al grupo de pertenencia, o a la

interacción entre el grupo y la habilidad (Pérez Gil, 2003). Las fórmulas para calcular el DIF con RL se ilustran en las ecuaciones 7 y 8.

Ecuación 7.

$$p_j = \frac{f(x)}{1 + f(x)}$$

- $p_j$ : probabilidad de responder correctamente el ítem  $j$  (donde 1= "responde correctamente", 0= "responde incorrectamente").

Ecuación 8.

$$f(x) = e^{\tau_0 + \tau_1\theta + \tau_2G + \tau_3\theta G}$$

- $e$ : exponencial
- $\theta$ : habilidad observada (puntaje en la prueba)
- $G$ : grupo variable (0= "referencial" o 1= "focal")
- $\tau_0$ : representa el peso asociado a la intersección
- $\tau_1\theta$ : indica las diferencias de habilidades entre grupos de examinados, respecto de la probabilidad de acertar el ítem
- $\tau_2G$ : probabilidad de acertar respecto de la pertenencia a un grupo
- $\tau_3\theta G$ : interacción entre el grupo y la habilidad. Es decir, la probabilidad de no responder correctamente no se ve influenciada únicamente por pertenecer a un grupo, sino también por diferencias de habilidad (*crossing*).

Bao *et. al.* (2009), señalan las características de los coeficientes en términos de su significancia. Si  $\tau_1\theta$  es estadísticamente significativo, eso implica que los examinados de habilidad superior tienen mayor probabilidad de responder el ítem correctamente. Si  $\tau_2G$  es significativo, quiere decir que la probabilidad de responder correctamente es diferente en cada uno de los dos grupos, observándose DIF uniforme. Por último, si  $\tau_3\theta G$  es estadísticamente significativo, el ítem muestra amplias diferencias en el desempeño del grupo en ciertos niveles de habilidad, pudiendo observarse DIF no uniforme.

Con el fin de clasificar el DIF, se generan tres modelos sucesivos de regresión, para evaluar cada uno de los coeficientes de forma separada. Los valores de los coeficientes son los mismos de la Ecuación 8.

Ecuación 9. Modelo 1

$$f(x) = e^{\tau_0 + \tau_1\theta}$$

Ecuación 10. Modelo 2

$$f(x) = e^{\tau_0 + \tau_1\theta + \tau_2G}$$



Ecuación 11. Modelo 3

$$f(x) = e^{\tau_0 + \tau_1\theta + \tau_2G + \tau_3\theta G}$$

Considerando estos modelos, Zumbo (1999) proporciona una forma para medir el tamaño del efecto DIF, denominado  $\Delta R^2$ , que es la diferencia en los valores de R-cuadrado en cada uno de los modelos. Para el análisis DIF del piloto de la PSU, interesa únicamente saber si el ítem posee DIF, y en qué magnitud, por lo que esta diferencia se realiza entre el Modelo 3 y el 1, como se expone en la Ecuación 12.

Ecuación 12.

$$\Delta R^2 = R^2(M3) - R^2(M1)$$

Tras calcular este  $\Delta R^2$ , el DEMRE utiliza el criterio propuesto por Jodoin y Gierl (2001), para categorizar el DIF, con una lógica similar a la mostrada en la Figura 8.

- $\Delta R^2 < 0,035$ : es un ítem Categoría A, considerado con DIF *Despreciable o Irrelevante*.
- $0,035 \leq \Delta R^2 \leq 0,070$ : es un ítem Categoría B, considerado con DIF *Moderado*. Se rechaza la hipótesis nula (no hay diferencias entre grupos).
- $\Delta R^2 > 0,070$ : es un ítem Categoría C, considerado con DIF *Severo* e hipótesis nula rechazada.

#### **7.4.5. Regla combinada de decisión**

Como se mencionó, MH es conocido por su eficacia al detectar DIF uniforme, mientras que BD es útil para la identificación de DIF no uniforme. De tal forma, se combinan ambos estadísticos a un nivel de significancia de 2,5 con el resultado obtenido en RL para evaluar la presencia y el tipo de comportamiento diferencial. Es decir, si ninguno de estos estadísticos es significativo, el ítem no posee DIF.

### **7.5 Revisión cualitativa de ítems**

Tras el análisis estadístico de los ítems piloto, estos fueron revisados cualitativamente para indagar en las causas que pudieron alterar su comportamiento en la población. El análisis cualitativo se utiliza como una estrategia para reconocer la posible influencia del contenido curricular u otros aspectos de la prueba, como el formato o el lenguaje, en las respuestas de los examinados. De tal modo, y sobre todo tras el análisis DIF, es importante contemplar la posibilidad de realizar un estudio que incorpore la interacción cultura-evaluación (Moreira-Mora, 2008).

En el DEMRE, se realiza un análisis posterior, donde se confirman las claves y se evalúan los folletos, para confirmar que no existan problemas de impresión de imágenes, figuras o textos. Durante la

revisión, también se confrontan los estadísticos con el acta de certificación y ensamblaje<sup>9</sup>. En estas actas, los constructores y revisores realizan comentarios respecto de las preguntas y sus características, previo a la aplicación piloto.

Por último, los ítems considerados “límitrofes”, es decir, cuyos estadísticos se encuentren al límite de lo aceptado o rechazado, se revisan en conjunto entre la UDA y la UCP, combinando elementos disciplinares y técnicos.

## 7.6 Finalización del proceso

Tras la identificación de los ítems con características estadísticas óptimas para ser ensambladas, la UDA informa a la UCP de sus resultados. Los ejes de análisis y las recomendaciones incluyen una mirada extensiva, considerando todos los parámetros mencionados anteriormente. Se incluye necesariamente una visión conjunta sobre su dificultad, discriminación, distractores y posibles sesgos.

De tal forma, la UDA entrega a la UCP, la nómina de ítems aceptados y sus características. Esto quiere decir que éstos son óptimos según los criterios estadísticos y pueden ensamblarse en una prueba definitiva.

## 7.7 Análisis de los resultados en la forma ancla

Como se indicó en la sección 1, en el piloto 2016 se aplicó en cada disciplina una forma de las utilizadas en la prueba PSU Oficial de la Admisión 2016 (diciembre 2015). Esta forma, denominada “ancla”, se distribuyó aleatoriamente entre los estudiantes. El objetivo es contrastar los resultados psicométricos de esta forma ancla con los de la aplicación oficial y evaluar así el supuesto de estabilidad de los parámetros de los ítems. Este supuesto es esencial para asumir que las preguntas piloteadas se comportarán de manera similar en la aplicación oficial.

Las preguntas comparadas fueron clasificadas de acuerdo a la diferencia en dificultad y discriminación entre ambas aplicaciones. Por una parte, si existían 10 o menos puntos porcentuales de diferencia en dificultad, la pregunta se clasificó como diferencia aceptable. Por otra parte, para comparar el parámetro de discriminación entre ambas aplicaciones se utilizó la transformación Z de Fisher (Kullback, 1959), detallada en la Ecuación 147.. Dado que la discriminación es un parámetro basado en correlaciones, se utiliza esta transformación para realizar el test de comparación de muestras. Esto permite evaluar si la diferencia entre la correlación observada en la muestra piloto y la correlación poblacional observada en la aplicación oficial es o no estadísticamente significativa al 95% (utilizando el p-valor).

Ecuación 13

$$r' = \frac{1}{2} \log \left| \frac{1 - r_b}{1 + r_b} \right|$$

- $r_b$ : correlación biserial

---

<sup>9</sup> Para mayor información respecto de los procesos de ensamblaje de pruebas piloto, ver página 14 del *Informe Técnico PSU. Proceso de Construcción y Ensamblaje de Pruebas*. En línea, disponible en: <http://www.portaldemre.demre.cl/estadisticas/documentos/informes/2016-vol-2-proceso-de-construccion-y-ensamblaje-de-pruebas.pdf> [Consultado el 26 de diciembre de 2016].



Ecuación 147.

$$Z_k = \frac{(r'_{1k} - r'_{2k})}{\sqrt{\frac{1}{N_{1k} - 3} + \frac{1}{N_{2k} - 3}}}$$

- $r'_{1k}$ : transformación de la correlación biserial del ítem k obtenida en forma ancla
- $r'_{2k}$ : transformación de la correlación biserial del ítem k obtenida en forma oficial
- $N_{1k}$ : número de sujetos que respondieron el ítem k en forma ancla
- $N_{2k}$ : número de sujetos que respondieron el ítem k en forma oficial

La Tabla 17 muestra la cantidad de preguntas que cumplen los criterios descritos en el párrafo anterior. Así, por ejemplo, se observa que en el caso de matemática 73 preguntas se encuentran dentro del rango aceptable cuando se considera el criterio de dificultad, pues no tienen más de 10 puntos porcentuales de diferencia en este parámetro. En relación al parámetro de discriminación, en esta misma disciplina hay 70 preguntas sin diferencia significativa entre la aplicación piloto y la oficial.

Tabla 17. Cantidad de preguntas en rangos aceptables por disciplina

	Total preguntas	Dificultad	Discriminación
Lenguaje	75	75	66
Matemática	75	73	70
Historia, Geografía y Cs. Sociales	75	74	75
Ciencias módulo común	54	54	46

En base a los resultados de este estudio comparativo se puede concluir que no existen grandes diferencias en los parámetros de las preguntas entre ambas aplicaciones.

# Bibliografía

- Camilli, G., & Shepard, L. A. (1994). *Methods for identifying biased test items*. Thousand Oaks: Sage.
- Adriola, W. (2003). Descripción de los principales métodos para detectar el funcionamiento diferencial del ítem (DIF) en el área de la evaluación educativa. *Revista de Pedagogía Bordón*, 177 - 189.
- Bao, H., Dayton, M., & Hendrickson, A. (2009). Differential Item Functioning Amplification and Cancellation in a Reading Test. *Practical Assessment, Research & Evaluation*, 14(19).
- DEMRE. (s.f). *Casos especiales*. Recuperado el 26 de septiembre de 2016, de Prueba de Selección Universitaria: <http://psu.demre.cl/inscripcion/casos-especiales>
- Holland, P., & Thayer, D. (1985). *An alternative definition of the ETS delta scale of item difficulty*. Educational Testing Service, Research Report . NJ: Princeton.
- Jodoin, M. G., & Gierl, M. J. (2001). Evaluation Type I error and power rates using an effect size measure with the logistic regression procedure for DIF detection. *Applied Measurement in Education*, 329-349.
- Kullback, S. (1959). *Information Theory and Statistics*. New York: Dover Publications.
- Mantel, N., & Haenszel, W. (1959). Statistical aspects of the analysis of data from retrospective studies of disease. *Journal of the National Cancer Institute*.
- MINEDUC. (2013). *Bases Curriculares de la Formación Diferenciada Técnico-Profesional aprobadas por Consejo Nacional de Educación*. Recuperado el 28 de octubre de 2016, de Unidad de Currículum y Evaluación. Ministerio de Educación: [http://ww2.educarchile.cl/UserFiles/P0001/File/CR\\_Articulos/Especial\\_Curriculum\\_2013/nuevas\\_bases\\_curriculares\\_tp\\_julio2013.pdf](http://ww2.educarchile.cl/UserFiles/P0001/File/CR_Articulos/Especial_Curriculum_2013/nuevas_bases_curriculares_tp_julio2013.pdf)
- Moreira-Mora, T. (2008). El funcionamiento diferencial del ítem: un asunto de validez y equidad. *Avances en Medición*(6), 5-16.
- Moreno, R., Martínez, R., & Muñiz, J. (2005). *Análisis de los ítems*. Madrid : La Muralla.
- Muñiz, J. (1997). *Introducción a la Teoría de Respuesta a los Ítems*. Madrid: Pirámide.
- Muñiz, J. (2003). *Teoría Clásica de los Test*. Madrid: Ediciones Pirámide.
- Penfield, D., & Camilli, G. (2006). Differential Item Functioning and Item Bias. En C. R. (Eds.), *Handbook of Statistics Psychometrics* (págs. 125-167). Amsterdam: Elsevier.
- Penfield, R. (2003). Applying the Breslow-Day test of trend in Odds Ratio heterogeneity to the analysis of nonuniform DIF. *The Alberta Journal of Educational Research*.
- Pérez Gil, J. A. (2003). *Funcionamiento Diferencial de los Ítems (DIF)*. *Desarrollos actuales de la medición: Aplicaciones en evaluación psicológica*. Sevilla: Dpto. de Psicología Experimental. Universidad de Sevilla.
- Sánchez, A. (2016). *Comunicación Ética de Resultados de Pruebas: El Plan Nacional para la Evaluación de los Aprendizajes (Planea) de México*. [Documento interno]. INEE: Instituto Nacional para la Evaluación de la Educación, México.



- Vivanco, M. (2005). *Muestreo Estadístico. Diseño y aplicaciones*. Santiago de Chile: Universitaria.
- Zumbo, B. (1999). *A handbook on the theory and methods for differential item functioning: Logistic regression modeling as a unitary framework for binary and likert-type (ordinal) item scores*. Ottawa: Directorate of Human Resources Research and Evaluation Department of National Defense.
- Zwick, R. (2012). *A Review of ETS Differential Item Functioning Assessment Procedures: Flagging Rules, Minimum Sample Size Requirements, and Criterion Refinement*. Research Report ETS RR-12-08. Princeton: Educational Testing Service.



# Anexos

## Anexo 1. Afiche informativo, Chillán

**Inscripciones**

**Piloto PSU**

**Rendición**

**Resultados de Puntajes**

**Período de Postulación**

**Resultados de Selección**

**Período de Matriculas**

# Este 8 y 9 de septiembre,

**No te quedes fuera del Piloto PSU 2016!**  
**Una buena oportunidad para:**

- » Poner a prueba tus conocimientos en un ensayo generado por el DEMRE
- » Ensayar para la PSU bajo las mismas condiciones de aplicación real
- » Recibir un puntaje simulado, acorde a tu desempeño

**Asegura tu cupo inscribiéndote al Proceso de Admisión 2017, en Demre.cl**

## Proceso de Admisión 2017 Prueba de Selección Universitaria

**DEMRE**  
DEPARTAMENTO DE EVALUACIÓN, MEDICIÓN Y REGISTRO EDUCACIONAL  
UNIVERSIDAD DE CHILE

Mesa de Ayuda  
Teléfono: (+56) 2 2978 3806  
[www.mesadeayuda.demre.cl](http://www.mesadeayuda.demre.cl)

Más información  
[www.demre.cl](http://www.demre.cl)

[f](#) [@demre\\_chi](#) [v](#) [DEMRECHILE](#)





Anexo 2. Características de establecimientos educacionales elegidos para pilotaje, en regiones seleccionadas

Región	Indicador	Categorías	Frec.	%
IV	Ruralidad	No rural	79	98,8
		Rural	1	1,3
	Dependencia	Particular pagado	5	6,3
		Particular subvencionado	50	62,5
		Municipal	25	31,3
	Rama	Científico humanista diurno	55	68,8
		Técnico comercial	9	11,3
		Técnico industrial	11	13,8
		Técnico servicios	5	6,3
	Total			80

V	Ruralidad	No rural	98	100
		Rural	0	0,0
	Dependencia	Particular pagado	18	18,4
		Particular subvencionado	49	50,0
		Municipal	31	31,6
	Rama	Científico humanista diurno	68	69,4
		Técnico comercial	8	8,2
		Técnico industrial	10	10,2
Técnico servicios		12	12,2	
Total			98	100

Región	Indicador	Categorías	Frec.	%
R.M.	Ruralidad	No rural	323	99,4
		Rural	2	0,6
	Dependencia	Particular pagado	59	18,2
		Particular subvencionado	209	64,3
		Municipal	57	17,5
	Rama	Científico humanista diurno	213	65,5
		Técnico comercial	38	11,7
		Técnico industrial	39	12,0
		Técnico servicios	35	10,8
	Total			325

VIII	Ruralidad	No rural	106	95,5
		Rural	5	4,5
	Dependencia	Particular pagado	13	11,7
		Particular subvencionado	68	61,3
		Municipal	30	27,0
	Rama	Científico humanista diurno	77	69,4
		Técnico comercial	11	9,9
		Técnico industrial	14	12,6
Técnico servicios		9	8,1	
Total			111	100



### Anexo 3. Distribución de Establecimientos Educativos en Proceso de Admisión 2016.

La presente distribución de datos se realizó a partir de los establecimientos educativos de donde provienen los estudiantes egresados el 2015 que rindieron la PSU Admisión 2016. Esta distribución considera únicamente las variables contenidas en el marco muestral. Por lo tanto, contiene las regiones IV, V, R.M y VIII, establecimientos de todas las dependencias y solo de las ramas HC Diurno, TP Comercial, Industrial y Servicios.

De tal modo, se consideran 3250 establecimientos en todo el país, donde solo 2120 pertenecen a las regiones del marco muestral. Además, una parte de estos establecimientos son polivalentes (con más de una rama de enseñanza), por lo tanto, en la variable rama hay más casos debido a la participación de este tipo de establecimientos.

Tabla 18. Distribución de Establecimientos Educativos en Proceso de Admisión 2016, según las variables que componen el marco muestral del Piloto PSU 2016.

		Frecuencia	Porcentaje
Rama educativa	Científico Humanista Diurno	2.167	59,4%
	Técnico Comercial	481	13,2%
	Técnico Industrial	519	14,2%
	Técnico Servicios	480	13,2%
Total		3.647	100%
Dependencia	Particular Pagado	401	12,4%
	Particular Subvencionado	1.848	56,8%
	Municipal	1.001	30,8%
Total		3.250	100%
Región	Metropolitana	1.173	55,3%
	Coquimbo	167	7,9%
	Valparaíso	425	20%
	Bío-Bío	355	16,7%
Total		2.120	100%
Ruralidad	Rural	175	94,6%
	No rural	3.075	5,4%
Total		3.250	100%



#### Anexo 4. Distribución de la primera prueba en la aplicación piloto, por comuna

Comunas que iniciaban actividades con la prueba de Lenguaje y Comunicación	Comunas que iniciaban actividades con la prueba de Matemática
Coquimbo	La Serena
Valparaíso	Ovalle
Quillota	Viña del Mar
Concepción	San Felipe
Talcahuano	Quilpué
Santiago	Los Ángeles
El Bosque	Chillan
Ñuñoa	La Cisterna
Providencia	La Florida
San Miguel	Las Condes
Puente Alto	Maipú
Melipilla	Quilicura
	San Bernardo

## Anexo 5. Registro de asistencia a Piloto PSU 2016



### PRUEBAS DE PILOTAJE - PROCESO DE ADMISION 2017 NÓMINA DE ASISTENCIA POR UNIDAD EDUCATIVA

<b>UNIDAD EDUCATIVA</b>	19 - 310
<b>PRUEBAS APLICADAS</b>	Matemática / Prueba Electiva / Lenguaje y Comunicación
<b>FECHA / HORA</b>	Martes 6 de Sept - 08:00 Hrs / Martes 6 de Sept - 11:30 Hrs / Miércoles 7 de Sept - 08:00 Hrs

RUT	NOMBRES	LENGUAJE Y C.	MATEMÁTICA	ELECTIVA
	ALVARADO CADIZ DUAM MANUEL ANGEL	P	P	P
	APARICIO RODRIGUEZ DANIEL MATIAS	P	P	P
	ARAOS MONTECINOS IGNACIA ABRIL	P	P	A
	BASAEZ VALDIVIESO FELIPE	A	A	A
	CARVAJAL SOLIS MARIA JESUS	A	P	P
	COLIPAN MACIAS SORAYA MILLARAY LLARETA	P	P	P
	CONTRERAS ANNONI NATALIA	P	P	P
	DURAN ALFARO MARIA CAROLINA	A	P	A
	ESCOBAR LEIGHTON FERNANDA ELIZABETH	P	P	P
	FARFAN CACERES VICTOR ENRIQUE	A	A	A
	FUENTES TORREBLANCA MARTA PAULA DENISSE	P	P	P
	GALLARDO BARRA SEBASTIAN IGNACIO	A	A	A
	GARCIA PARKER MARIA-JESUS	A	A	A
	GONZALEZ CANCINO JOSE TOMAS	A	A	A
	GUTIERREZ REYES EUGENIO IGNACIO	P	P	P
	HIDALGO AYLWIN PATRICIO EDUARDO	A	P	P
	LANAS SENTIS MARIA JESUS	P	P	A
	MANSILLA HENRIQUEZ NICOLAS RAFAEL	P	P	P
	MERCADO SUAZO NATALIA ISABEL	P	P	P
	PARRA RUIZ MARIA JOSE	P	P	P
	PEREZ SAAVEDRA CONSUELO JESUS	A	A	A
	REYES CROXATTO FRANCISCA IGNACIA	P	P	A
	RODRIGUEZ LYMEBURG GUSTAVO ANDRES	P	P	P
	ROJAS LAVEZZOLO CRISTOBAL RENE	A	P	P
	ROMUSSI GONZALEZ FLORENCIA ANTONIA	P	P	P
	SABAG MANSILLA BENJAMIN IGNACIO	A	A	A
	SUAREZ MONTENEGRO CAMILA CONSTANZA	P	P	P
	TEPPER HACKE FRANCISCO JOSE	P	P	P
	VALENZUELA SILVA ANDRES IGNACIO	A	P	P
	VILLA PARRA ALONDRA DE LOS ANGELES	P	P	A
	ZUÑIGA FUENTES NICOLAS IGNACIO	P	P	P

NOTA: P = PRESENTE ; A = AUSENTE



## Anexo 6. Informe Piloto PSU 2016 para establecimientos educacionales

### INFORME DE RESULTADOS [Nombre establecimiento]

**[RBD - Código enseñanza]**

En el mes de Septiembre se realizó la aplicación piloto PSU 2016 en 23 comunas, pertenecientes a cuatro regiones del país (Coquimbo, Valparaíso, Metropolitana y Bio-Bío). En esta aplicación participaron 888 unidades educativas y 23.543 estudiantes. Este informe muestra los resultados generales de los estudiantes de su establecimiento en cada una de las pruebas. A modo de referencia, se presentan los resultados nacionales, los de su región y los de su comuna.

El rendimiento se reporta como Porcentaje de Respuestas Correctas Promedio de cada prueba. Dado que en la aplicación piloto se "inventan" preguntas para utilizar en aplicaciones PSU Oficiales futuras, no es posible reportar puntajes agrupados en escala PSU.

Lenguaje y Comunicación	51%	NACIONAL
Matemática	27%	
Historia, Geografía y Cs. Sociales	39%	
Ciencias	36%	

Lenguaje y Comunicación	51%	REGION
Matemática	26%	
Historia, Geografía y Cs. Sociales	38%	
Ciencias	35%	

Lenguaje y Comunicación	53%	COMUNA
Matemática	28%	
Hist., Geo. y Cs. Sociales	41%	
Ciencias	38%	

Lenguaje y Comunicación	41%	ESTABLECIMIENTO
Matemática	19%	
Hist., Geo. y Cs. Sociales	34%	
Ciencias	19%	

### INFORME DE RESULTADOS [Nombre establecimiento]

**[RBD - Código enseñanza]**

Para interpretar en forma más precisa el rendimiento de sus estudiantes, se comparan los resultados de su establecimiento en el piloto PSU con los obtenidos por los estudiantes de establecimientos educacionales similares al suyo.

Establecimientos similares al suyo son aquellos de dependencia Municipal y que imparten la rama educativa Científico-Humanista.

41% NACIONAL 34% REGIONAL 49%	19% NACIONAL 37% REGIONAL 23%	34% NACIONAL 38% REGIONAL 36%	19% NACIONAL 38% REGIONAL 34%
Lenguaje y Comunicación	Matemática	Historia, Geografía y Ciencias Sociales	Ciencias

En caso de que menos de 6 estudiantes de su establecimiento hayan rendido las pruebas, no es posible entregar mayor detalle sobre el rendimiento obtenido pues se corre el riesgo de incurrir en imprecisiones técnicas.

### RESULTADOS POR PRUEBA [Nombre establecimiento]

**[RBD - Código enseñanza]**

**LENGUAJE Y COMUNICACIÓN**

Distribución de Respuestas Correctas (RC).

La prueba piloto de Lenguaje y Comunicación tuvo un total de 75 preguntas. La siguiente tabla indica la cantidad de estudiantes de su establecimiento ubicados en cada tramo de Respuestas Correctas.

0-15	16-30	31-45	46-60	61-75
0	6	4	0	0

Cantidad total de estudiantes de su establecimiento: 10

**Resultados por Sección y Habilidad**

La prueba de Lenguaje y Comunicación tiene cuatro secciones y evalúa cinco habilidades, definidas en el currículum de Educación Media. El siguiente diagrama indica el Porcentaje Promedio de Respuestas Correctas que obtienen los estudiantes de su establecimiento que participaron en la aplicación piloto en cada una de estas secciones y habilidades.

<b>Secciones</b>	Comprensión Lectora	39%	<b>Habilidades</b>	Analizar e Interpretar	53%
	Conexión	66%		Analizar y Sintetizar	34%
	Plan de Redacción	36%		Evaluar Información	40%
	Vocabulario	43%		Interpretar Información	33%
				Recuperar Información	43%

NOTA: Si las tablas incluyen el símbolo "-" quiere decir que menos de 6 estudiantes de su establecimiento participaron en el piloto PSU en la prueba respectiva. Si esto es el caso, no es posible entregar detalles sobre el rendimiento obtenido pues se corre el riesgo de incurrir en imprecisiones técnicas.

### RESULTADOS POR PRUEBA [Nombre establecimiento]

**[RBD - Código enseñanza]**

**MATEMÁTICA**

Distribución de Respuestas Correctas (RC).

La prueba piloto de Matemática tuvo un total de 25 preguntas. La siguiente tabla indica la cantidad de estudiantes de su establecimiento ubicados en cada tramo de Respuestas Correctas.

0-15	16-20	21-25	46-60	61-75
10	10	0	0	0

Cantidad total de estudiantes de su establecimiento: 20

**Resultados por Eje Temático y Habilidad**

La prueba de Matemática evalúa cuatro ejes temáticos y tres habilidades, todos definidos en el currículum de Educación Media. El siguiente diagrama indica el Porcentaje Promedio de Respuestas Correctas que obtienen los estudiantes de su establecimiento que participaron en la aplicación piloto en cada una de estas secciones y habilidades.

<b>Ejes Temáticos</b>	Números	19%	<b>Habilidades</b>	Aplicación	19%
	Álgebra	18%		ASE*	23%
	Geometría	20%		Comprensión	19%
	Datos y Azar	20%			

\*ASE: Análisis, Síntesis, Evaluación

NOTA: Si las tablas incluyen el símbolo "-" quiere decir que menos de 6 estudiantes de su establecimiento participaron en el piloto PSU en la prueba respectiva. Si esto es el caso, no es posible entregar detalles sobre el rendimiento obtenido pues se corre el riesgo de incurrir en imprecisiones técnicas.

### RESULTADOS POR PRUEBA [Nombre establecimiento]

**[RBD - Código enseñanza]**

**HISTORIA, GEOGRAFÍA Y CIENCIAS SOCIALES**

Distribución de Respuestas Correctas (RC).

La prueba piloto de Historia, Geografía y Ciencias Sociales tuvo un total de 25 preguntas. La siguiente tabla indica la cantidad de estudiantes de su establecimiento ubicados en cada tramo de Respuestas Correctas.

0-15	16-20	21-25	46-60	61-75
-	-	-	-	-

Cantidad total de estudiantes de su establecimiento: 0

**Resultados por Eje Temático y Habilidad**

La prueba de Historia, Geografía y Ciencias Sociales evalúa cuatro ejes temáticos y tres habilidades, todos definidos en el currículum de Educación Media. El siguiente diagrama indica el Porcentaje Promedio de Respuestas Correctas que obtienen los estudiantes de su establecimiento que participaron en la aplicación piloto en cada uno de estos ejes y habilidades.

<b>Ejes Temáticos</b>	Chile y América en Perspectiva	-%	<b>Habilidades</b>	Aplicación	-%
	Democracia y Desarrollo	-%		ASE	-%
	El Mundo en Perspectiva Histórica	-%		Comprensión	-%
	Espacio Geográfico	-%			

\*ASE: Análisis, Síntesis, Evaluación

NOTA: Si las tablas incluyen el símbolo "-" quiere decir que menos de 6 estudiantes de su establecimiento participaron en el piloto PSU en la prueba respectiva. Si esto es el caso, no es posible entregar detalles sobre el rendimiento obtenido pues se corre el riesgo de incurrir en imprecisiones técnicas.

### RESULTADOS POR PRUEBA [Nombre establecimiento]

**[RBD - Código enseñanza]**

**CIENCIAS**

Distribución de Respuestas Correctas (RC).

La prueba piloto de Ciencias tuvo un total de 10 preguntas. La siguiente tabla indica la cantidad de estudiantes de su establecimiento ubicados en cada tramo de Respuestas Correctas.

0-20	21-40	41-60	61-80
-	-	-	-

Cantidad total de estudiantes de su establecimiento: 0

**Resultados por Materia y Habilidad**

La prueba de Ciencias evalúa tres materias y cuatro habilidades, todos definidos en el currículum de Educación Media. El siguiente diagrama indica el Porcentaje Promedio de Respuestas Correctas que obtienen los estudiantes de su establecimiento que participaron en la aplicación piloto en cada una de estas materias y habilidades.

<b>Materias</b>	Biología	-%	<b>Habilidades</b>	Aplicación	-%
	Física	-%		ASE	-%
	Química	-%		Comprensión	-%
				Recapitulación	-%

\*ASE: Análisis, Síntesis, Evaluación

NOTA: Si las tablas incluyen el símbolo "-" quiere decir que menos de 6 estudiantes de su establecimiento participaron en el piloto PSU en la prueba respectiva. Si esto es el caso, no es posible entregar detalles sobre el rendimiento obtenido pues se corre el riesgo de incurrir en imprecisiones técnicas.