

# Primer informe del Comité de Expertos del Sistema Único de Admisión

---

Consejo de Rectores  
de las Universidades Chilenas



CONSEJO DE RECTORES DE  
LAS UNIVERSIDADES CHILENAS



# Índice

Presentación	4
Prólogo	6
Análisis y recomendaciones	9
Propuesta 1. Revisar y definir con claridad las interpretaciones y usos deseados de la batería de pruebas de la PSU	10
Propuesta 2. Revisar la extensión del contenido comprendido en la batería de pruebas de la PSU	12
Propuesta 3. Estudiar la posibilidad de dividir la actual Prueba de Matemáticas en dos pruebas con diferentes contenidos y niveles de dificultad	14
Propuesta 4. Revisar los contenidos de la prueba de Lenguaje y particularmente la evaluación de la escritura en esta prueba	18



Propuesta 5. Estudiar la posibilidad de dividir la actual prueba de Ciencias en pruebas separadas de Biología, Física y Química	23
Propuesta 6. Estudiar la adopción de modelos basados en la teoría de respuesta al ítem para calcular los puntajes y establecer la comparabilidad entre distintas Pruebas de Selección Universitarias a través del tiempo	27
Propuesta 7. Nombrar un Comité Técnico Internacional y un Comité Asesor Nacional	31
Epílogo: Síntesis de recomendaciones	32
Bibliografía	33
Apéndice: Bitácora de actividades y comunicaciones del Comité	38

## Presentación

El Consejo de Rectores de las Universidades Chilenas (CRUCH) ha dado muestras permanentes de su compromiso con la calidad y mejora continua de su sistema de admisión y los instrumentos de selección. Para ello, ha generado instancias técnicas especializadas y periódicamente ha realizado estudios y evaluaciones, tanto internas como externas.

A partir de 2013, el organismo del CRUCH encargado de velar por el desarrollo y gestión del sistema de selección y admisión a las universidades del Consejo de Rectores y a universidades privadas adscritas, es el Sistema Único de Admisión (SUA).

La estructura orgánica del SUA contempla, entre otras instancias, un Comité de Expertos, cuyo propósito es "asesorar técnicamente al SUA (en materias de selección y admisión universitaria) y a su Directorio en aquellas materias de su competencia, para su adecuada resolución o pronunciamiento por parte del pleno del Consejo" (Resolución Exenta N°122/2016). Para estos efectos, el Comité de Expertos puede realizar evaluaciones técnicas de los instrumentos que se utilizan, estudiar propuestas y revisar el desarrollo del proceso de admisión en su conjunto. Debe estar conformado por especialistas de reconocido prestigio en materias de admisión universitaria, cuatro nacionales y dos extranjeros.

En la Sesión N° 590 del CRUCH, del jueves 30 de noviembre de 2017 en la Universidad de La Frontera, Temuco, el Directorio del SUA presentó al pleno del Consejo nombres de especialistas nacionales e internacionales. En la ocasión se analizó la propuesta y se efectuó una votación. En virtud de la experiencia, trayectoria, formación y compromiso con esta temática, se ratificaron a las siguientes personas: Alejandra Mizala, Universidad de Chile; Salomé Martínez, Universidad de Chile; Verónica Santelices, Pontificia Universidad Católica de Chile; José Luis Saiz, Universidad de La Frontera; Rebecca Zwick, investigadora del Educational Testing Service (ETS), Estados Unidos; y Michal Beller, Rectora del Levinsky College of Education, Israel. Todos(as) fueron convocados(as) y aceptaron; posteriormente, la profesora Beller excusó su participación por problemas de agenda.

El 5 de abril de 2018 el Vicepresidente Ejecutivo del CRUCH, Rector Aldo Valle, y la Directora del SUA, María Elena González, dieron la bienvenida a los(as) integrantes nacionales del Comité de Expertos y dialogaron en torno a los requerimientos y necesidades del sistema de admisión del CRUCH, con miras a delinear el trabajo del año.

La solicitud del Consejo de Rectores al Comité de Expertos fue una evaluación integral del sistema, es decir, llevar a cabo un diagnóstico general, un análisis crítico



del mismo y la propuesta de estudios y de cambios que estimaran convenientes, a fin de corregir los actuales instrumentos de selección o de incorporar otros.

Sin embargo, el Comité de Expertos definió que su trabajo durante el primer año se focalizaría en analizar y evaluar la factibilidad de implementar las propuestas de ajustes a las Pruebas de Selección Universitaria (PSU) presentadas por el Departamento de Evaluación, Medición y Registro Educacional (DEMRE) de la Universidad de Chile, organismo encargado de administrar el proceso de admisión por mandato del CRUCH.

La semana del 14 de mayo de 2018 se constituyó formalmente el Comité de Expertos. El programa de trabajo contempló distintas reuniones entre sus miembros, así como espacios para escuchar las propuestas de cambio al sistema de admisión -específicamente de la PSU- del DEMRE, y de voceros(as) de la CONFECH (Confederación de Estudiantes de Chile) en conjunto con la CONES (Coordinadora Nacional de Estudiantes Secundarios).

Los(as) expertos(as), también participaron del Seminario "Desafíos y nuevas perspectivas para la admisión universitaria", realizado el 17 y 18 de mayo en la Universidad Técnica Federico Santa María, Valparaíso. El evento convocó a rectores, vicerrectores académicos, vicerrectores de pregrado, directores de docencia y delegados de admisión de las universidades del SUA.

A partir de mayo de 2018 el Comité de Expertos trabajó en evaluar las propuestas del DEMRE y, el 26 de diciembre de 2018, entregó un informe con sus recomendaciones.

## Prólogo

Este documento es el resultado de la discusión y evaluación de los miembros del Comité de Admisión con respecto a la propuesta del DEMRE (2018). Los miembros activos del comité fueron Salomé Martínez, Alejandra Mizala, Verónica Santelices y Rebecca Zwick<sup>1</sup>. En las siguientes páginas, se discute cada una de las propuestas de DEMRE en forma separada. Independientemente de la decisión alcanzada sobre cada uno de los cambios específicos propuestos, es importante que los cambios realizados a la batería de pruebas de la PSU se realicen en forma progresiva, que las modificaciones sean secuenciadas e, idealmente, que se implementen una a la vez. Además, información apropiada y avanzada debería proporcionarse a profesores y a estudiantes de educación media, familias e instituciones de educación superior.

Los miembros del comité concordaron en que abordar las debilidades de la prueba de Matemáticas y de la prueba de Lenguaje es una prioridad. Es urgente dividir la prueba de Matemáticas en dos pruebas separadas de contenido diferenciado (básico y avanzado) y excluir los actuales *proxies* de escritura (conectores y plan de redacción) de la prueba de Lenguaje. En paralelo, se debería considerar la implementación de una prueba de Escritura para mostrarle a las escuelas la importancia de enseñar redacción durante la educación media. Mientras todos, los cuatro miembros, concuerdan en la necesidad de tener una prueba básica de matemáticas obligatoria, no hubo consenso en si la prueba avanzada de Matemáticas debería ser obligatoria para todos los estudiantes o solo para estudiantes que postulan a programas del área de matemáticas.

Los miembros del comité no llegaron a un acuerdo en cuanto a si se debe implementar la Teoría de Respuesta al Ítem para analizar y asignar puntaje a la PSU. Los distintos miembros del comité sopesaron las ventajas y desventajas de forma diferente. No obstante, concordaron en la necesidad de estudiar su implementación y probar diferentes modelos, procedimientos de estimación y posibles desafíos comunicacionales que puede acarrear al personal de DEMRE y a otros actores relevantes la transición hacia tal sistema. Es importante mencionar que al usar los modelos IRT que podrían ser apropiados para la PSU, los puntajes ya no tendrían una relación simple con el número de respuestas correctas.

---

<sup>1</sup> El apéndice que acompaña este documento es un detalle de las reuniones y comunicaciones que sostuvieron los miembros del comité entre mayo y diciembre del 2018, y entre miembros del comité y profesionales del SUA durante este mismo periodo.



Los beneficios de reemplazar la actual prueba de Ciencias por tres pruebas de Ciencias independientes que midan Biología, Física y Química son menos claros y requieren mayor estudio antes de implementar la idea. Algunos miembros del comité apoyaron la propuesta de DEMRE de informar la disciplina de contenido avanzado escogida por el estudiante y el puntaje obtenido por el examinado en los ítems de tal disciplina específicamente (combinando ítems básicos y avanzados), junto con el puntaje total obtenido en la prueba. Sin embargo, otros miembros del comité sugirieron aumentar el número de ítems en la sección común de la prueba, con el fin de alcanzar un nivel de confiabilidad que permitiera informar el puntaje de la sección de contenido básico en forma separada e independiente del puntaje obtenido en la sección avanzada.

Salomé Martínez

Alejandra Mizala

María Verónica Santelices

Rebecca Zwick







# **ANÁLISIS Y RECOMENDACIONES**

# Propuesta 1. Revisar y definir con claridad las interpretaciones y usos deseados de la batería de pruebas de la PSU

Con el fin de abordar el primer punto presentado en el documento de DEMRE (abril, 2018), sugerimos que SUA encabece un estudio para explorar la Teoría de Acción, o Modelo Lógico, de los puntajes de la PSU, incluyendo los usos, tanto predefinidos como no predefinidos, que habrían de darles los diferentes actores interesados incluyendo los diseñadores de pruebas y los usuarios de los puntajes de pruebas. Los Estándares para Pruebas Educativas y Psicológicas (American Educational Research Association [AERA], la American Psychological Association [APA] y National Council on Measurement in Education [NCME]; 2014) plantean la importancia de definir el uso de los puntajes de las pruebas, lo cual debería ser realizado en forma explícita para la prueba de admisión a las universidades chilenas.

El estudio podría ser organizado por profesionales del SUA, en colaboración con personal del CRUCH o analistas de DEMRE, pero debe nutrirse de lo informado por los líderes de los diferentes grupos interesados, incluyendo las autoridades de educación superior, responsables de la toma de decisiones en el Ministerio de Educación y, tal vez, miembros del sector de enseñanza media (colegios y liceos).

Para definir los "usos esperados" de un sistema de evaluación, la medición o prueba se puede considerar como equivalente a una intervención o programa educacional. Este concepto o estrategia, característica de la literatura sobre Evaluación de Programas y Políticas (Rossi, Freeman y Lipsey, 2003) explora en profundidad las expectativas de aquellos participantes del diseño de la evaluación y de quienes en forma activa utilizan sus puntajes, así como los mecanismos a través de los cuales se espera alcanzar los objetivos, es decir los usos esperados. Las importantes consecuencias que tienen las mediciones tales como la PSU, para los estudiantes en Chile, sugieren que la Teoría de Acción debería ser explícitamente declarada y examinada.

Es esencial que se definan los usos que se le debería dar a la información obtenida de un instrumento o medición, con el fin de educar a los posibles usuarios respecto de las interpretaciones apropiadas de la información que entrega tal instrumento (NCME Statement, 2018). La Teoría de Acción, o Modelo Lógico, también sirve de base para cualquier iniciativa de validación. Un proceso adecuado de validación requiere que los usos de los puntajes del instrumento sean definidos y explicitados

lo más tempranamente posible. El concepto de validez, entendido a la luz de los Estándares (AERA, APA y NCME, 2014) y de Kane (2016), presenta un proceso de validación "como una discusión compleja" que incluye tanto evidencia cuantitativa como cualitativa, pero que siempre se concibe e implementa para un "uso" e "interpretación" específicos de los puntajes. Desde esta perspectiva, no es posible validar un instrumento, o prueba, en el vacío, ignorando los usos que se les da a sus puntajes en el ámbito educativo y psicológico. Además, la definición explícita de este "uso" específico, o "usos" específicos, y su comunicación activa permite una mejor definición de los criterios que deben ser considerados en estudios de validez predictiva y validación de estándares.

La Teoría de Acción es "el proceso por el cual los componentes del programa supuestamente afectan los resultados y las condiciones en las que se cree que operan estos procesos " (Donaldson, 2007, p.22). Cuando la Teoría de Acción no está presente, los expertos aconsejan utilizar múltiples fuentes para desarrollarla (Rogers, Petrosino, Huebner y Hacsí, 2000). Tales fuentes incluyen las teorías implícitas en aquellas personas cercanas al diseño y a la implementación del programa, la documentación de las actividades del programa, observación y exploración de los supuestos más relevantes. El objetivo es crear una teoría de programa simple, o un plan, que relacione las actividades (o componentes) del programa, o evaluación, con los objetivos esperados. Esta Teoría de Acción puede ser utilizada para priorizar las preguntas de evaluación (Donaldson y Gooler, 2003). La mayor parte de las teorías del programa se resumen en un diagrama que representa una cadena causal (Rogers et al., 2000).

Una vez que la Teoría de la Acción, o Modelo Lógico, es establecida, se debe centrar la atención en explorar la evidencia empírica que dé soporte (o no) al argumento, o argumentos, de validez implícito en tal teoría. Los argumentos de validez estarán dados por los "usos" o los "efectos esperados" especificados en la teoría de acción.

## Propuesta 2. Revisar la extensión del contenido comprendido en la batería de pruebas de la PSU

En relación a la propuesta del DEMRE de revisar el contenido considerado dentro de la batería de pruebas de la PSU, creemos que este tema se debería abordar en profundidad. Discutimos este tema de manera más extensa en la próxima sección sobre Matemáticas (sección tres) y tratamos algunos temas relacionados en las secciones cuatro y cinco para Lenguaje y Ciencias. La definición del contenido que cubre la prueba debería ser analizada con actores relevantes, tales como autoridades de la educación superior, representantes de instituciones de enseñanza media y el Ministerio de Educación, y no necesariamente ser limitados a los contenidos que incluye el Currículum Nacional de Enseñanza Media vigente en Chile.

El contenido de la prueba también debería estar en estrecha correspondencia con los usos predefinidos y las consecuencias esperadas según lo definido por el estudio anteriormente sugerido; es decir, el contenido debe ser consistente con los usos predeterminados. Por ejemplo, si el consenso es enfatizar el uso predictivo de los puntajes de la prueba con el fin de predecir el éxito académico en las universidades, entonces los actores relevantes deberían decidir qué contenido es el más adecuado para predecir el rendimiento universitario. Si el objetivo es proveer de información diagnóstica sobre ciertas habilidades generales a todo tipo de instituciones de educación superior, los actores relevantes en universidades e instituciones técnicas deberían decidir cuáles son aquellas habilidades generales y también el nivel de logro esperado ya sea para efectos de admisiones, remediales y/o cualquier otro tipo de decisión educacional. En lo sucesivo, describimos algunos de los estudios que se han llevado a cabo sobre estos dos usos potenciales en los últimos diez a quince años.

El poder predictivo de la actual batería de pruebas de la PSU, según lo señalado por Manzi et al. (2010), parece ser algo menor que el de pruebas similares en los Estados Unidos (Mattern, Patterson, y Kobrin, 2012) y Suecia (Wikström y Wikström, 2017)<sup>2</sup>.

---

2 De las dos principales pruebas de admisión a universidades estadounidenses, SAT y ACT, las mismas han sido validadas relacionando el rendimiento en las pruebas con el rendimiento académico medido a través del promedio de notas del primer y cuarto año de universidad y por los niveles de graduación and persistencia (Geiser y Santelices, 2007; Zwick, 2017). En general la correlación entre los puntajes de SAT (Verbal o Lectura Crítica y Matemáticas combinados) y las notas de primer año de universidad fluctúa entre 0.3 y 0.6, dependiendo de las características de los estudios, con un promedio de 0.4 (Ramist, Lewis y McCamley-Jenkins, 1994; Zwick, 2002).

Sin embargo, no existe suficiente información para realmente tener una comprensión acabada del poder predictivo de la actual PSU ya que los métodos usados en los análisis oficiales hasta el momento han sido limitados a la correlación promedio entre los puntajes de la prueba y las notas del primer año de universidad por institución. Además, los estudios de validez no son realizados en forma sistemática. El último estudio de validez predictiva que incluyó a todas a las instituciones del CRUCH se realizó utilizando datos del año 2015 (Consejo de Rectores, 2018).

La investigación en cuanto a validez debería ser realizada en forma sistemática por el organismo fiscalizador de la calidad de la prueba y/o encargado de su diseño o aplicación. Por otro lado, la investigación de validez debería llevarse a cabo examinando no solo la relación entre los puntajes de la prueba y el promedio de notas del primer año de universidad, sino también otros indicadores de transición exitosa hacia la educación superior tanto académicos, como sociales. Este tipo de análisis requiere que los investigadores y los profesionales del SUA o de otros centros de investigación y universidades tengan acceso a los indicadores académicos y sociales de los estudiantes de todas las instituciones que usen los puntajes de la prueba. La información debería estar disponible, sin incluir identificación personal, para todas las partes interesadas que la requieran. Actualmente, sería posible que el SIES (Servicio de Información de Educación Superior) recolectara los campos de información relevantes como parte del proceso continuo a través del cual recopila la información de matrículas y la información con respecto al número de créditos registrados y el número de créditos aprobados por los estudiantes.

En los últimos diez años, los diseñadores de la prueba han hecho énfasis en alinear el contenido de la prueba al currículum de enseñanza media del área científico-humanista sin dejar de lado la capacidad predictiva de los puntajes de la prueba. Sin embargo, el informe Pearson (2013) criticó la magnitud en que la PSU cubre, en realidad, el currículum de educación media. Otro tema importante de abordar es que el contenido de la PSU actualmente no está alineado con el currículum del área vocacional, el cual comprende aproximadamente un 45% de la matrícula de enseñanza media.

---

Mattern, Patterson y Kobrin (2012) encontraron una correlación entre SAT-Matemáticas y las notas de cursos de matemáticas de primer año de universidad de 0.52. SUA implementó un estudio de validez predictiva que encontró una correlación corregida por restricción de rango de 0.58 para el 2013, 0.62 para el 2014 y 0.56 para el 2015 entre la PSU de Matemáticas y el promedio ponderado acumulado de los estudiantes de primer año (SUA, 2017). El mismo estudio encontró correlaciones menores para la PSU de Lenguaje durante el mismo periodo (0.23 en 2013, 0.20 en 2014 y 0.18 en 2015).

## Propuesta 3. Estudiar la posibilidad de dividir la actual Prueba de Matemáticas en dos pruebas con diferentes contenidos y niveles de dificultad

La alta dificultad de la PSU de Matemáticas (PSU-M) queda demostrada en el Informe de DEMRE (2018a; Sección g.), según el cual se puede observar una distribución asimétrica del número de respuestas correctas en las aplicaciones de la prueba desde el 2004 al 2018. Por ejemplo, en el 2016 un puntaje de 450 puntos en la PSU-M corresponde a solo 19 respuestas correctas de un total de 75, y la media de 496 puntos corresponde a solo 22 respuestas correctas. Si se considera que cada ítem tiene 5 alternativas, estos puntajes son similares a responder al azar (lo que corresponde a 15 respuestas correctas). Es importante mencionar que 450 puntos es el puntaje PSU mínimo requerido para postular a carreras de universidades que participan del SUA. Debido a la naturaleza asimétrica de la distribución, el proceso de normalización usado para computar los puntajes genera una situación en que los estudiantes cuyos puntajes son mayores a la mediana obtienen finalmente pequeñas diferencias en los puntajes normalizados, aunque tengan diferencias importantes en el número de respuestas correctas. Esto es problemático puesto que ésta es la subpoblación relevante para efectos de selección. Es importante añadir que la normalización es una transformación hecha para comparar diferentes cohortes y no cambia el ranking.

La propuesta del DEMRE se sustenta en un estudio realizado con el cohorte del 2013 (DEMRE, 2017), en que la distribución de los puntajes de la PSU-M fue analizada para diferentes subpoblaciones de estudiantes, distinguiendo, además, según tipo de contenido. Los autores distinguen: Contenido Básico (CB), correspondiente a los Contenidos Mínimos Obligatorios de Matemáticas para 1° y 2° medio según el Currículum de Enseñanza Media, y Contenidos Avanzados (CA), correspondientes a los Contenidos Mínimos Obligatorios de Matemáticas para 3° y 4° medio según el Currículum de Enseñanza Media. En este informe, los autores investigan la confiabilidad, dificultad y validez predictiva de los dos subconjuntos de ítems. Si bien este estudio aborda muchos temas importantes relacionados a la propuesta, una limitación relevante es que fue realizado con el cohorte del 2013, momento en el cual el descuento por pregunta incorrecta aún estaba vigente. No obstante, es probable que las conclusiones del estudio aun así sean ciertas para aplicaciones actuales de la PSU. En efecto, el informe del CMM (Lacourly, San Martín, Amaya y

Uribe, 2017), en el que se desarrolló una metodología para comparar dos modalidades de pruebas diferentes (con y sin descuento por respuestas incorrectas), muestra que en general la información obtenida en ambos escenarios está altamente correlacionada. Aun así, sugerimos enérgicamente que se realice un estudio con los nuevos datos.

## Dificultad de la Prueba

En DEMRE (2017) se observa que la distribución de puntajes de la prueba es altamente asimétrica para los estudiantes de las modalidades Científico - Humanista (CH) y Técnico - Profesional (TP) de los establecimientos municipales (M) y los colegios particular - subvencionados (PS), los que representan un 90% de la matrícula. Además, existe una importante brecha de rendimiento entre los estudiantes de establecimientos SH y los estudiantes de establecimientos TP. Al considerar solo los ítems CB, la distribución de puntajes se mantiene no simétrica, pero como lo muestra la Tabla 4, existe un pequeño aumento en los porcentajes de respuestas correctas para las diferentes subpoblaciones de estudiantes.

Con respecto de los estudiantes de colegios particulares pagados (P), la PSU-M muestra un comportamiento anómalo, puesto que su distribución es uniforme. Esto puede estar relacionado con el hecho de que las tareas que la prueba pide realizar se pueden entrenar y a que la prueba presenta una baja dificultad para esta población. El porcentaje promedio de respuestas correctas para este grupo es de 65, el cual es mucho mayor que el 32% de respuestas correctas correspondientes a los estudiantes CH de establecimientos PS, que es el segundo grupo de más alto rendimiento, según modalidad curricular y tipo de establecimiento.

## Valor Predictivo

Un factor importante que debe ser tomado en consideración en la decisión de modificar los contenidos de la prueba es si el cambio disminuye su validez predictiva. En DEMRE (2017), los autores miden el rendimiento académico usando el promedio de notas de los estudiantes en su primer año de universidad. En la tabla 7a (p. 31), se muestra que para todos estudiantes de universidades del SUA, la validez predictiva de la prueba de Matemáticas tiene una correlación de 0.19 con el promedio de notas, y de 0.18 cuando solo son considerados los ítems CB. Por el contrario, la correlación entre el promedio de notas del primer año de universidad y el promedio de Notas de Enseñanza Media (NEM) es de 0.31. Es importante mencionar que la correlación de la PSU-M no es consistente con la informada en SUA (2018). En este informe, la correlación de la PSU-M y el promedio de notas del primer año

de universidad es 0.26 (ver página 40 del informe). Esta discrepancia se debe a diferencias metodológicas<sup>3</sup>.

En DEMRE (2017), se consideran diferentes modelos de regresión. Se establece que los ítems CB y CA no representan una contribución significativa a la varianza del promedio de notas de los alumnos en su primer año de universidad en algunas carreras. El estudio concluye que el valor predictivo de la prueba de Matemáticas es bajo (por ejemplo, de acuerdo a Cohen, 1988), mientras que, si solo se consideran los ítems CB, no hay cambios significativos en el valor predictivo de la prueba.

## Conclusiones y recomendaciones

El comité propone repetir el estudio de DEMRE (2017) considerando el cohorte PSU del 2016, por lo que solicitamos que SUA ponga a disposición los datos necesarios. En este nuevo estudio, recomendamos que se individualicen carreras que puedan ser de interés para estudiantes de alto rendimiento en matemáticas, tales como Ingeniería Civil, licenciaturas en Matemáticas o Física. Además, es importante considerar diferentes grupos de universidades según su selectividad y otras características con el fin de evaluar en forma más precisa el efecto de los ítems CA en sus carreras. Esto puede generar información sobre la utilidad de los ítems en la forma en que están contruidos actualmente y proporcionar una mejor comprensión sobre la construcción de los ítems CA para la selección en tales carreras.

El comité recomienda que la prueba de matemáticas de la batería de PSU sea dividida y propone la introducción de dos pruebas nuevas: una prueba de matemáticas Básica y una prueba de matemáticas Avanzada. Debido a que la PSU es construida basándose en un marco curricular, cada población de estudiantes debe ser evaluada de acuerdo a su propio currículum, de otra forma se generan importantes interrogantes en cuanto a oportunidades de aprendizaje y equidad. En consecuencia, tener una prueba obligatoria que considera solo los contenidos hasta segundo medio, que corresponde al currículum común para todos los estudiantes, es una decisión razonable. Este tema es particularmente importante en matemáticas, área donde se presentan problemas significativos en el rendimiento de los alumnos.

---

3 En el análisis de validez predictiva de DEMRE (2017) se consideraron solo los 45.530 estudiantes que se graduaron de educación media en el 2012 y que ingresaron a universidades del Consejo de Rectores y universidades privadas que participan del SUA en 2013. De éstos, 34.552 ingresaron a carreras con 30 o más estudiantes. Éstos constituyeron la muestra para el análisis de regresión múltiple. No hubo corrección por restricción de rango. En SUA (2018), se consideraron carreras con más de 25 estudiantes, y la muestra de estudiantes consideró no solo a aquellos graduados en el 2012 (ver SUA, 2018, página 21).



Conceptualizar y construir dos nuevas pruebas de matemáticas, una Básica y otra Avanzada, puede ser altamente beneficioso términos de selección. Los CB y CA se pueden evaluar de mejor manera con dos pruebas que pueden ser acortadas de forma significativa (a diferencia de la cantidad actual de 75 ítems), lo cual puede afectar positivamente el rendimiento. Además, nos parece que la prueba Básica podría incluir contenidos curriculares de 7° y 8° básico, lo cual entregaría mejores oportunidades para empezar a evaluar las habilidades matemáticas definidas en el currículum, las cuales, en teoría, deben tener una mayor correlación con el rendimiento en los programas de educación superior, para los cuales se desarrollan currículos en su mayoría de acuerdo a modelos basados en competencias.

Otro tema que se debe abordar es la modalidad en que los estudiantes pueden rendir estas pruebas. Por ejemplo, una propuesta reciente de DEMRE es contar con una prueba de matemáticas Básica y otra Avanzada y que cada estudiante pueda escoger cuál prueba rendir (pero que no sea posible que rinda ambas). Este diseño puede ser atractivo, ya que la prueba Avanzada de Matemáticas termina siendo obligatoria para solo para algunos alumnos y se ahorra tiempo y recursos al hacer que los estudiantes escojan solo una prueba. No obstante, existen importantes consideraciones al respecto que se deben tener en cuenta al seguir este diseño. Tener solo una prueba obligatoria aumenta considerablemente su extensión y alcance de contenidos y probablemente aumenta significativamente su dificultad. Además, si los estudiantes de alto rendimiento no toman la prueba Básica de Matemáticas, la prueba puede terminar con una distribución de puntajes menor, lo cual se traduciría eventualmente en una prueba de baja dificultad, enviando un mensaje negativo al sistema escolar. Escoger entre ambas pruebas puede llevar a cierto grupo de estudiantes con mayor aversión al riesgo (especialmente mujeres, como se señala en Arias, Mizala y Meneses, 2017) a escoger la prueba Básica de Matemáticas, lo cual puede incrementar la disminución de la matrícula de grupos específicos de estudiantes en carreras del área matemática.

El comité identifica dos alternativas distintas con respecto a la prueba de CA. La primera alternativa es que esta prueba sea obligatoria para todos los estudiantes que postulan a carreras del área matemática. Esta alternativa tiene la desventaja de que muchos establecimientos de enseñanza media, particularmente aquellos de menos recursos, podrían no enseñar el currículum avanzado de matemáticas y, por consiguiente, los estudiantes de aquellos establecimientos tendrían menos oportunidades de aprendizaje. Además, puede propiciar un grupo más segregado tanto de postulantes, como de estudiantes en los programas de educación superior. La segunda alternativa es que ambas pruebas, la Avanzada y la Básica, sean obligatorias. La prueba más avanzada no sería relevante para un grupo de instituciones y programas de educación superior, pero al menos los estudiantes tendrían la oportunidad de mostrar su conocimiento y los establecimientos escolares no estarían poco interesados en enseñar contenido avanzado de matemáticas. La decisión entre ambas alternativas no es obvia y requiere de un cuidadoso análisis.

## Propuesta 4. Revisar los contenidos de la prueba de Lenguaje y particularmente la evaluación de la escritura en esta prueba

El temario de la sección de Lenguaje y Comunicación de la PSU está organizado en torno a dos elementos: procesos y temas. Procesos se refiere a los aspectos genéricos del currículum evaluados, mientras que temas se refiere a aspectos específicos de categorías genéricas de textos. Son dos los procesos de escritura que la prueba pretende medir, (i) el valor de la escritura como una actividad de expresión personal creativa y reflexiva que permite al individuo organizar ideas, presentar información, interactuar con diversas realidades y que entrega una oportunidad de elaborar conscientemente una visión de mundo personal y (ii) el uso adecuado de un vocabulario variado, que incorpore algunos usos y matices de palabras, expresiones y terminología menos frecuentes de acuerdo al contexto, propósito y receptor.

El "Temario de la Prueba de Lenguaje y Comunicación" del DEMRE indica que estos procesos se miden indirectamente. La primera sección es de selección múltiple y se centra en el mapeo de conectores. En ésta, los estudiantes deben seleccionar uno de cinco conectores para completar diez oraciones descontextualizadas. En la segunda parte, llamada Plan de Redacción, los estudiantes deben ordenar la secuencia de cinco oraciones de modo de componer un texto coherente. La tarea se centra en medir la capacidad de los estudiantes de ordenar un texto ya escrito. Esta forma de evaluar escritura se basa en ciertas suposiciones que deben ser explicadas en detalle y analizadas. Primero, se asume que la tarea de escribir puede ser reducida al uso de conectores y vocabulario. Segundo, se supone que escribir es general, es decir, que una forma de redactar sirve para comunicar temas diferentes y que, por lo tanto, escribir no necesita ser una actividad situada o significativa. Tercero, se asume que no es necesario evaluar la producción de la escritura en forma directa, que puede ser evaluada indirectamente (Navarro, Ávila Reyes y Gómez Vera, 2018).

Este método de evaluación solo considera las dimensiones más visibles de la escritura: ortografía, puntuación y gramática. No obstante, estas dimensiones no son suficientes para evaluar si un texto comunica ideas de forma exitosa, lo cual es una función fundamental de la escritura (Perelman, 2018). Además, este tipo de evaluación pide tareas que son reproducibles y que se pueden copiar y entrenar. Por el contrario, de acuerdo a los expertos, escribir es una actividad situada, lo que significa que las actividades que implica son cultural y situacionalmente específicas

y cambiantes. Escribir es complejo, ya que incluye múltiples dimensiones que requieren de conocimientos y habilidades diferentes; escribir materializa complejos procesos de reflexión (Navarro, 2018).

El carácter situado y complejo de la escritura significa que ésta no puede ser evaluada como si fuera una competencia genérica, simple y transversal desconectada de contexto, receptores y propósitos específicos. La redacción se materializa en textos o formatos discursivos que satisfacen funciones comunicativas en un ambiente social. En otras palabras, una adecuada evaluación de la redacción requiere preguntar por formatos discursivos que sean significativos y pertinentes para los estudiantes y que traten de temas específicos que les sean familiares (Navarro, 2018).

## La importancia de la escritura

La escritura es muy relevante hoy en día ya que está involucrada en la mayoría de las actividades, tareas y eventos importantes. Esto es particularmente cierto en la educación formal, ya que les permite a los estudiantes demostrar lo que han aprendido comunicar sus ideas. La escritura es parte de la formación integral de los estudiantes de educación superior; involucra varias de las llamadas competencias del siglo 21 tales como pensamiento crítico, comunicación efectiva, argumentación, uso y evaluación de información y fuentes, solución de problemas, creatividad, etc. (Bazerman, 2013). El extenso uso de la escritura en la sociedad del conocimiento hace necesario que se incorpore la redacción en los instrumentos que miden la capacidad de completar exitosamente un programa de estudios en un establecimiento de educación superior. De hecho, los países desarrollados incluyen pruebas de escritura en sus sistemas educacionales (Perelman, 2018).

Además, lo que se mide y la forma en que se mide envía fuertes mensajes al sistema escolar. El resultado de usar selección múltiple para evaluar redacción en pruebas estandarizadas de alto impacto, tales como la PSU, es que no se refuerza la enseñanza y la práctica de escritura en enseñanza media (Slomp 2008). El déficit de este tipo de aprendizaje durante la enseñanza media ha provocado que un número de establecimientos chilenos de educación superior incorporen programas de escritura académica (Preiss, Castillo, Grigorenko y Manzi, 2013 a; Preiss, Castillo, Flotts y San Martín, 2013b). Incluir la evaluación de la escritura en las pruebas de admisión a instituciones de educación superior podría proveer información relevante a estas iniciativas de alfabetización académica.

Lo anterior también se relaciona con la validación de las pruebas, ya que la validación supone el hecho de la prueba esté alineada con los objetivos educacionales relevantes y que, al menos, no limite los dominios educativos enseñados (AERA, APA y NCME 2014).

## Cómo asignar puntaje a una prueba de escritura

Los instrumentos que se usan para evaluar la redacción en general definen explícitamente un receptor y un género discursivo específico (Perelman 2018). Un texto escrito puede ser evaluado con rúbricas analíticas u holísticas. Las rúbricas analíticas usualmente incluyen 5 o 6 dimensiones, tales como contenido, ortografía, vocabulario, cohesión, coherencia y estructura, cada una con 3-5 niveles de desempeño. Un problema que presenta este tipo de evaluación es de qué forma cuantificar cada dimensión. También es difícil determinar la ponderación de cada dimensión en la evaluación final, así como alcanzar un alto nivel de acuerdo entre los analistas. En el caso de las rúbricas holísticas, la evaluación involucra una lectura más o menos rápida. La atención se centra más en las fortalezas que en las debilidades. No se evalúa por dimensiones separadas, sino que más bien se evalúa si el texto comunica exitosamente información e ideas a un lector predefinido. Idealmente, debería ser evaluado por un panel, lo cual aumenta la validez del instrumento en términos de imparcialidad.

## Experiencia chilena en Evaluación de la Escritura

La prueba SIMCE evalúa la habilidad escritora de los estudiantes de 6° básico, especificando un propósito, una situación y un lector. La prueba está alineada con los objetivos de aprendizaje de las bases curriculares del Ministerio de Educación. Lo que se evalúa preferentemente es la capacidad de comunicar un mensaje con claridad. Esta prueba evalúa globalmente los textos escritos utilizando una guía holística<sup>4</sup>.

Aparte de la prueba SIMCE, existen estudios piloto realizados por DEMRE y algunas universidades. DEMRE desarrolló un proyecto de investigación para evaluar la posibilidad de implementar un instrumento que evalúe en forma directa la escritura en la prueba de admisión universitaria. Los investigadores crearon la prueba CODICE que involucra preguntas de opción múltiple para evaluar comprensión lectora y preguntas abiertas basadas en el mismo texto (con respuestas de extensión mediana) para evaluar la escritura. Entre los años 2006 y 2009, implementaron pruebas piloto y concluyeron que incluir estas preguntas abiertas puede mejorar el proceso de selección universitaria. Además, se concluyó que el uso de preguntas abiertas con referencia a un texto dado puede equilibrar las diferencias de condiciones culturales que se relacionan con el conocimiento previo de los estudiantes. Asimismo, pueden disminuir el impacto que tiene la capacitación para una prueba, lo que sucede por ejemplo si se utiliza un ensayo (Hernández, 2012).

---

4 En forma complementaria, se usan pautas analíticas para entregar información a los colegios y liceos.

La Pontificia Universidad Católica de Chile implementó una Prueba de Comunicación Escrita para sus estudiantes de pregrado. Los estudiantes deben escribir un ensayo escogiendo de entre tres temas distintos. Se evalúan aproximadamente 10 dimensiones usando una rúbrica analítica. La prueba no tiene incidencia en temas de admisión, sin embargo, es una prueba de alto impacto ya que aprobarla es un requisito de graduación. Si el estudiante no aprueba, debe asistir a clases adicionales para mejorar su habilidad escritora. Preiss et al. (2013a), usando los datos de un cohorte de estudiantes, encontró que esta prueba es un predictor significativo de las notas en la universidad a lo largo del tiempo, incluso luego del control por variables individuales y académicas. Además, Preiss et al. (2013b) utilizando datos similares encontraron que la información originada en la prueba de escritura complementa la información que entrega la estandarizada PSU.

Asimismo, un grupo de cuatro universidades participó en un estudio piloto en el cual instrumentos que complementan la prueba cognitiva utilizada en el proceso de admisión universitaria eran desarrollados y testeados en estudiantes universitarios y de enseñanza media. Estos instrumentos incluían un Ensayo de Pensamiento Crítico y un Ensayo de Reflexión Personal. El Ensayo de Pensamiento Crítico está basado en la Prueba de Comunicación Escrita mencionada anteriormente, el formato de la prueba es el mismo, pero el encabezado y también las instrucciones fueron modificados. El estudio muestra que estos instrumentos pueden complementar las actuales pruebas de selección universitaria (principalmente la PSU de lenguaje), aun cuando el Ensayo de Pensamiento Crítico cuenta con relativa baja confiabilidad (Santelices et al., 2010).

## Experiencia internacional en la evaluación de la escritura

El National Assessment of Educational Progress o NAEP (Programa Nacional de Evaluación del Progreso en el Sistema Educativo de los Estados Unidos) evalúa la escritura en una muestra representativa de estudiantes de 4º, 8º y 12º grado en los Estados Unidos. La tarea presenta un género y un receptor determinado (por ejemplo, una carta al director del colegio). En la prueba, se deben completar diferentes tareas de redacción en un tiempo determinado y, para evaluarla, se utiliza una rúbrica holística.

La prueba opcional de escritura del programa SAT pide un ensayo como respuesta a una lectura, el cual debe ser realizado en un tiempo determinado.

El Tercer Estudio Comparativo y Explicativo (TERCE), desarrollado por el Laboratorio Latinoamericano de Evaluación de la Calidad de la Educación (LLECE) administra una prueba a una muestra de estudiantes de 3º y 6º básico en 15 países de Latinoamérica y el Caribe. Esta prueba incluye escritura y utiliza el formato de cartas con propósitos comunicativos específicos.

## Conclusiones y recomendaciones

Existen buenos argumentos que apoyan evaluar la escritura en las pruebas que miden competencias para completar un programa de estudios en instituciones de educación superior. La escritura es parte de la formación integral de los estudiantes, les permite comunicar ideas y conocimientos y está conectada con las habilidades necesarias para participar en la sociedad de la información, la comunicación y el conocimiento.

Lo que actualmente se evalúa en la parte de lenguaje de la PSU es un proxy de escritura basada en una serie de suposiciones que no son correctas. La forma en que se evalúa la escritura involucra tareas susceptibles de ser reproducidas, copiadas y entrenadas, lo cual envía la señal de no fortalecer la enseñanza y el ejercicio de la escritura durante la educación media.

Existen buenos ejemplos, incluso en Chile, de pruebas de escritura que cumplen con un número de características adecuadas. No obstante, su aplicación implica mayores costos monetarios (es necesario preparar evaluadores para asignar puntaje a las pruebas y asegurar su confiabilidad) y mayor tiempo de corrección, lo cual puede ser complejo para la PSU, dado el corto tiempo en que los resultados deben estar disponibles.

Una posibilidad a evaluar es la introducción de una prueba de escritura que se pueda rendir durante el 4° año de enseñanza media, ya que una prueba de este tipo no requiere todos los contenidos de enseñanza media, lo que permite un mayor tiempo para la corrección. Ésta puede considerarse como prueba de estar habilitado para la educación superior y no como parte de la batería de tests que entrega el puntaje de ingreso a la educación superior. Esto es relevante ya que en general este tipo de pruebas tienen un menor grado de confiabilidad. Algunas universidades ya están realizando este tipo de pruebas a los estudiantes que ingresan a primer año y ofrecen programas remediales a quienes que no la aprueban. Aprobar esta prueba es un requisito de titulación. La implementación de esta prueba de escritura enviaría un mensaje a los establecimientos educacionales acerca de la importancia de enseñar escritura durante la educación media.

Por el momento, es necesario considerar no incluir las secciones de conectores y plan de redacción dentro de la prueba de lenguaje, ya que no miden las competencias de escritura de los estudiantes y generan incentivos no deseados para la educación media.

## Propuesta 5. Estudiar la posibilidad de dividir la actual prueba de Ciencias en pruebas separadas de Biología, Física y Química

En esta sección, se discuten los temas de dimensionalidad y confiabilidad de la PSU de Ciencias. Ambos temas son relevantes para las afirmaciones de validez del uso de puntajes de la PSU de Ciencias para efectos de admisión a programas de educación superior. Basándonos en esta discusión, se presentan recomendaciones en cuanto a la propuesta del DEMRE de dividir la actual prueba de Ciencias en tres pruebas separadas (DEMRE, 2017).

### Estructura Interna - Dimensionalidad

Existe preocupación con respecto a reportar un puntaje único de ciencias. La versión actual de la prueba incluye (1) una sección básica de 54 ítems de tres diferentes Ciencias (18 ítems de Biología, 18 ítems de Química y 18 ítems de Física) y (2) una sección avanzada de 26 ítems de sólo una disciplina (ya sea Biología, Física o Química) para los estudiantes científico-humanistas (DEMRE, 2016). Cada estudiante escoge cuál disciplina tomar basándose en sus habilidades y en los requisitos de admisión de la carrera que le interesa estudiar. Los estudiantes técnico-profesionales responden un set distinto de 26 ítems alineadas con el currículum de la modalidad profesional.

La pregunta respecto de si la actual prueba mide en realidad sólo un constructo (prueba unidimensional) o múltiples constructos (prueba multidimensional) se relaciona directamente con la forma en que se debería informar el desempeño en una prueba estandarizada. Los puntajes únicos se usan para informar el desempeño en una prueba unidimensional, mientras que puntajes en subescalas se justifican cuando las pruebas miden más de una dimensión. La dimensionalidad de una prueba usualmente se evalúa utilizando diferentes fuentes de evidencia que permiten comparar la definición teórica de un constructo con la evidencia empírica de la estructura interna de la prueba (por ejemplo, análisis factorial exploratorio y confirmatorio).

En este caso, la prueba de Ciencias combina preguntas de Biología, Física y Química, lo que podría sugerir que en cuanto a contenidos la prueba no es unidimensional. La evidencia de los análisis factoriales es algo contradictoria (DEMRE, 2018a,

2018b). Aunque los resultados de análisis factoriales confirmatorios realizados por profesionales del DEMRE muestran índices de ajuste del modelo que apoyan la estructura unidimensional, los resultados de los análisis factoriales exploratorios indican que el primer factor de la prueba de Ciencias tiende a explicar una proporción de la varianza (entre el 9% y 25% dependiendo de la forma de la prueba) que está por debajo de la proporción de la varianza esperada explicada por sólo un factor en pruebas unidimensionales<sup>5</sup>. Los miembros del comité interpretan que esta evidencia sugiere una potencial multidimensionalidad en la prueba de Ciencias, por lo cual no se aconseja que se informe solo un puntaje. Se deberían considerar formas alternativas de informar los desempeños y/o habilidades.

## Confiabilidad

Junto con la dimensionalidad de las pruebas, es importante considerar la precisión de la evaluación y de las subescalas para decidir qué puntajes y subescalas se deberían informar. En otras palabras, se debe poner mucha atención a la confiabilidad de la prueba, y sus subdimensiones, para tomar tal decisión.

En los documentos de los que dispusieron los miembros del Comité (DEMRE, 2017, página 6), podemos ver que la confiabilidad de un “puntaje de prueba compuesto” formado por todos los ítems de las secciones básica y avanzada de ciencias (un total de 44) de una disciplina específica (ya sea Biología, Física o Química) está dentro del rango que se considera apropiado para reportar puntajes a nivel individual para la disciplina escogida, 0.90 a 0.92 dependiendo de la disciplina (ver la siguiente tabla). No se encontró información de la confiabilidad de los ítems comunes básicos de Ciencia en la información proporcionada a los miembros del Comité.

**Tabla. Confiabilidad de la Prueba de Ciencias y del Puntaje Compuesto**

Confiabilidad (Alfa de Cronbach)	Prueba Ciencias (80 preguntas)	Sección Materia Específica (44 preguntas)
PSU Ciencias - Biología	0.94	0.90
PSU Ciencias - Física	0.95	0.92
PSU Ciencias - Química	0.95	0.91

Fuente: DEMRE, (2017). *Propuesta del Departamento de Evaluación Medición y Registro Educativo (DEMRE) al Consejo de Rectores de las Universidades Chilenas (CRUCH) de Mejoras a la PSU*. Santiago: Chile.

<sup>5</sup> Stevens (1996) recomienda un 75% o más para las pruebas unidimensionales, pero Henson y Robertson (2006), en su revisión de los análisis factoriales exploratorios, indican una varianza explicada promedio de 52% en pruebas unidimensionales.



## Conclusiones y recomendaciones

Los miembros del Comité apoyan la propuesta del DEMRE de informar la disciplina de contenido avanzado escogida por el estudiante y el puntaje obtenido por el examinado en los ítems de tal disciplina específica (combinando los ítems básicos y avanzados, "puntaje compuesto"). Ésta es de hecho una de las sugerencias incluidas en el Informe Pearson (2013)<sup>6</sup>.

Los miembros del Comité, sin embargo, no concuerdan en si hay o no suficiente evidencia para considerar la prueba de Ciencias realmente como una prueba "multidimensional". Mientras que algunos miembros del Comité se inclinan a considerar la prueba de Ciencias como una prueba multidimensional basándose mayormente en los resultados de análisis factoriales exploratorios, otros están dispuestos a considerar la estructura unidimensional como una posibilidad dados los resultados de análisis factoriales confirmatorios. Incluso si se le considera una prueba multidimensional, no hay suficiente evidencia para realmente interpretar y nombrar las dimensiones que la versión actual de la prueba de Ciencias puede estar midiendo.

Según la discusión, los miembros del comité no llegan a acuerdo sobre qué otro puntaje adicional se debería informar. Mientras algunos apoyan que se entregue el puntaje obtenido en los 80 ítems de Ciencias, basándose en los resultados de los Análisis Factoriales Confirmatorios, otros piensan que se debería informar el puntaje de los ítems básicos. Esto último solo sería correcto si esta sección tuviese una confiabilidad aceptable, lo cual está aún por ser confirmado. De acuerdo con los miembros del comité que apoyan esta iniciativa, habría de entregarse el puntaje de los ítems básicos de ciencia junto con el "puntaje compuesto" (la combinación de ítems básicos y avanzados en la disciplina que cada estudiante ha escogido).

Es importante considerar que ambos, tanto el puntaje de los ítems básicos de ciencia y el "puntaje compuesto" pueden medir un constructo multidimensional. Es necesario realizar más estudios para explorar esta hipótesis.

---

<sup>6</sup> En este nuevo escenario, es importante reflexionar sobre cómo reaccionará el proceso de admisión de educación superior a la disponibilidad de nueva información y cómo estas decisiones pueden impactar en las prácticas escolares y en las oportunidades de los estudiantes para aprender las tres ciencias. Creemos que incluso si algunos programas deciden requerir "Secciones Avanzadas" específicas de los solicitantes (por ejemplo, las escuelas de Medicina pueden decidir solicitar la sección Avanzada de Biología), todos los estudiantes todavía necesitarán aprender las tres Ciencias porque aún habrá una "Prueba de ciencia básica" que los evalúa a todos. Sin embargo, el efecto del potencial requisito de ciertas "Secciones avanzadas" se debe monitorear de cerca para garantizar que todas las escuelas secundarias y los estudiantes tengan incentivos para enseñar y aprender biología, física y química.



La introducción de las tres pruebas de Ciencias separadas requiere de mayor análisis y consideración. Esta modificación puede tener el objetivo de mejorar la validez predictiva de la batería de pruebas PSU para algunos programas de educación superior específicos, pero tal argumento requiere de un análisis más exhaustivo de los potenciales aumentos de la validez predictiva que pueden generarse como consecuencia de la introducción de una prueba de Matemáticas “Básica” y una “Avanzada”. Además, se debe estimar detenidamente el número de potenciales usuarios de estas tres pruebas de separadas de Ciencia con el fin de examinar si existen usuarios potenciales que justifiquen los costos del desarrollo, pilotaje y aplicación estas tres nuevas pruebas nacionales.

## Propuesta 6. Estudiar la adopción de modelos basados en la teoría de respuesta al ítem para calcular los puntajes y establecer la comparabilidad entre distintas Pruebas de Selección Universitarias a través del tiempo

Actualmente, el puntaje de la PSU es una transformación del número de ítems que son respondidos correctamente. El documento “Áreas Clave para Revisar en la batería de PSU” de fecha abril de 2018, plantea la idea de que los métodos para otorgar puntaje basados en la teoría de respuesta al ítem (IRT, *Item Response Theory*) podrían ser ventajosos.

Realizar una transición hacia un puntaje basado en TRI tendría implicancias sustanciales en distintas áreas. Se necesitaría un nuevo software, junto con nuevos procesos para analizar, otorgar puntaje, ensamblar, equiparar y realizar control de calidad a los ítems. Los requisitos del personal, incluyendo el entrenamiento necesario, probablemente también se verían afectados. La transición probablemente tomaría 2 o 3 años y requeriría una investigación previa sustancial, incluyendo investigación de las diferentes alternativas de modelos IRT y procedimientos de estimación. Finalmente, se necesitaría una serie de publicaciones para explicar el cambio a la comunidad, en especial, a los estudiantes de enseñanza media. Éste no es un asunto menor, puesto que según los modelos IRT que serían de utilidad para la PSU, los puntajes ya no estarían relacionados de manera simple con la cantidad de respuestas correctas.

Los modelos de respuesta al ítem que, en general, se usan en ítems de respuesta dicotómica (que se consideran correctas o incorrectas) son el modelo logístico de un parámetro (1PL), el modelo logístico de dos parámetros (2PL) y el modelo logístico de tres parámetros (3PL). En estos modelos de IRT, la probabilidad de respuesta correcta se representa como una función del nivel de competencia del examinado y de las propiedades del ítem.

El modelo de 1PL considera solo la dificultad del ítem, el 2PL permite además que los ítems difieran en discriminación (el grado en que el ítem puede distinguir entre examinados altamente competentes y examinados de baja-competencia) y el 3PL incluye un tercer parámetro que representa la posibilidad de que un examinado de muy baja-competencia responda correctamente el ítem (lo que a veces se denomina parámetro de pseudo-advinanza o *guessing parameter*). Se asume que

las respuestas de los ítems no son afectadas por ningún otro factor (tales como contexto oposición del ítem) y que son independientes, a un nivel de habilidad dado.

El uso de la asignación de puntajes basada en IRT puede, sin embargo, tener sustanciales ventajas. Puede facilitar el proceso de *equating* de múltiples formas de prueba, lo cual haría más fácil la posibilidad de aplicar la PSU dos o más veces al año. Además, puede facilitar la creación de bancos de ítems, así como el proceso de ensamblaje de las formas de pruebas. En especial, IRT facilita la construcción de pruebas que pueden, de forma precisa, distinguir entre examinados a distintos niveles de la escala de puntajes. Por ejemplo, si existe un puntaje de corte específico usado para efectos de admisión, un modelo IRT puede simplificar el desarrollo de una prueba que produzca resultados precisos en el área cercana a ese puntaje de corte. Un cambio hacia IRT sería esencial si hubiese un interés por desarrollar versiones flexibles o de etapas múltiples de ciertas porciones de la PSU. Este tipo de pruebas se ajustan a la dificultad de las preguntas incluidas en la prueba basándose en respuestas recibidas en preguntas anteriores (van der Linden y Glas, 2010; Yan, von Davier y Lewis, 2014). Sin embargo, es importante mencionar que todas estas potenciales ventajas de IRT aplican solo si los datos de la prueba se ajustan al modelo IRT y si hay suficientes datos para obtener estimados estables de los parámetros del modelo. En consecuencia, pruebas de ajuste del modelo son esenciales a la hora de tomar una decisión sobre si se debería adoptar un enfoque IRT y, en caso de ser así, qué modelo debería ser implementado.

Actualmente, hay un interés por aplicar el modelo de Rasch a los datos de la PSU. Sin embargo, los resultados del análisis del mismo DEMRE, que se pueden apreciar en las imágenes 5, 7, 9, 11 y 13 del informe *Respuesta a carta No. 19/2018 del SUA*, muestran que pocos ítems de la PSU se ajustan a este modelo. Este hecho no es sorprendente puesto que los ítems en pruebas de selección múltiple usualmente varían en discriminación y pueden ser respondidos correctamente adivinando. Tal como lo indicara un experto en evaluación educacional hace 35 años, los supuestos del modelo de Rasch, que el uso del azar es mínimo y que las discriminaciones de los ítems son iguales, en general, no se cumplen en el caso de los datos de desempeño en educación. Estas suposiciones, de hecho, "son contrarias al sentido común y a la gran cantidad de evidencia empírica acumulada a lo largo de los últimos 80 años" (Traub, 1983). El modelo de Rasch es atractivo por su simpleza, pero sus propiedades atractivas aplican solo si los datos se ajustan al modelo.

En lo siguiente, pasamos a revisar los modelos de escalas que utilizan programas de evaluación y medición prominentes. Cuatro de éstos son evaluaciones de puntaje en grupo y no son pruebas de alto impacto para los estudiantes que las rinden: la prueba Programme for International Student Assessment (PISA), la prueba National Assessment of Educational Progress (NAEP), la prueba Trends in International

Mathematics and Science Study (TIMSS) y la evaluación Progress in International Reading Literacy Study (PIRLS). Tres son pruebas de admisión de alto impacto para los examinados: las pruebas ACT, SAT y Graduate Record Examinations (GRE). Todos estos siete programas utilizan ya sea un modelo 2PL o uno 3PL para los ítems de selección múltiple o utilizan enfoques que no son modelos IRT. Es importante mencionar que, con anterioridad, la prueba PISA usaba el modelo de Rasch, pero lo dejó de lado por ser inadecuado.

**PISA, NAEP, TIMSS y PIRLS:** De acuerdo con el informe técnico PISA del 2015 (Organisation for Economic Cooperation and Development, 2017, p. 142), “en el pasado, se ha habido preocupación con respecto a la incapacidad del modelo de Rasch de abordar en forma adecuada las complejidades de los datos de PISA (Kreiner y Christensen, 2014; Oliveri y von Davier, 2011, entre otros). Otros estudios nacionales e internacionales usan modelos IRT más generales (Mazzeo y von Davier, 2014; von Davier y Sinharay, 2014). La evaluación National Assessment of Educational Progress (NAEP), por ejemplo, usa el modelo IRT de 3 parámetros y el modelo de crédito parcial generalizado o MCPG (Allen, Donoghue y Schoeps, 2001) de la misma forma que lo hacen el Trends in International Mathematics and Science Study o TIMSS y el Progress in International Reading Literacy Study o PIRLS (Martin, Gregory y Stemler, 2000). Con el fin de abordar las preocupaciones sobre el uso del modelo de Rasch, la prueba PISA 2015 implementó el modelo logístico de 2 parámetros, 2 PLM, (Birnbaum, 1968) para respuestas evaluadas en forma dicotómica y el modelo de crédito parcial generalizado (Muraki, 1992) para ítems con más de dos categorías de respuestas ordenadas.

**ACT y SAT:** estas pruebas de admisión a pregrado, que se rinden en EE.UU., no usan IRT para informar el puntaje. Los puntajes informados son transformaciones de los puntajes reales obtenidos a través de un proceso de *equating*.

**GRE:** La Prueba General GRE se usa para admisión a programas de posgrado en EE.UU., así como también a ciertas escuelas de negocios y escuelas de derecho. La versión actual la prueba GRE, que fue lanzada el año 2011, usa escalas desarrolladas a través de la teoría de respuesta al ítem. De acuerdo con la documentación sobre la prueba GRE (Robin y Steffen, 2014):

“Dos modelos alternativos IRT fueron considerados: el modelo de tres parámetros, el cual se preocupa de la discriminación y la dificultad de la pregunta y el uso de la respuesta al azar (lo cual puede ocurrir con preguntas que requieren que los examinados seleccionen de entre un número limitado de respuestas posibles), y el modelo de dos parámetros, el que se preocupa solo de la dificultad y la discriminación. Análisis de los datos tanto de las pruebas de Razonamiento Verbal, como de Razonamiento Cuantitativo indicó que el modelo IRT de dos parámetros proporcionaría un buen ajuste con los datos. Este resultado había sido anticipado ya que la prueba

actualizada reducía de forma importante el uso de simples preguntas de selección múltiple y, por consecuencia, la posibilidad de adivinar las respuestas correctamente. Con menos parámetros que considerar, el modelo de dos parámetros requiere de muestras más pequeñas que las que utilizaba anteriormente el modelo de tres parámetros [usado por el programa GRE]. Por consiguiente, el modelo IRT de dos parámetros fue escogido para la versión actualizada de la Prueba General. (p. 3.3.4)

## Conclusiones y recomendaciones

Debido a sus vastas implicancias, la decisión de adoptar o no puntajes por IRT para la PSU requiere un análisis cuidadoso. En términos de la selección del modelo, los propios análisis realizados por el DEMRE sugieren que adoptar el modelo de Rasch es altamente riesgoso. Usar un modelo que no se ajusta adecuadamente a los datos da como resultado representaciones imprecisas de las propiedades de los ítems y de las habilidades de los examinados. El Informe de Evaluación Pearson (2023), que recomienda que la PSU haga uso de modelos IRT, sugiere el uso del modelo de tres parámetros (p. 62). Vale la pena, además, considerar las experiencias del programa GRE, que eligió usar el modelo de dos parámetros en su versión más recientemente actualizada, en caso de que la PSU decida adoptar una asignación de puntaje por IRT. La evaluación de cualquier procedimiento de asignación de puntaje propuesto para la PSU debería incluir investigación de ajuste del modelo, precisión de puntaje y validez de los puntajes como predictores de rendimiento académico universitario.



## Propuesta 7. Nombrar un Comité Técnico Internacional y un Comité Asesor Nacional

El comité concuerda en que sería altamente beneficioso para el sistema de admisión nombrar un Comité Técnico Internacional, compuesto por seleccionados expertos de renombre mundial, que pueda aconsejar en materias clave y ayudar a construir un plan estratégico para el nuevo sistema de admisión. Elogiamos al DEMRE por nombrar un Comité Internacional y un Comité Asesor Nacional y por mantener conferencias internacionales durante estos últimos años.



## Epílogo: Síntesis de recomendaciones

- 1 Es necesario avanzar en el desarrollo de la Teoría de Acción detrás de la Prueba de Selección Universitaria, que ayuda a explicitar los usos intencionados y no intencionados de los puntajes.
- 2 Se recomienda separar la actual prueba de Matemáticas en dos pruebas: una prueba básica y una prueba avanzada.
- 3 Se recomienda eliminar la medición indirecta de la escritura en la prueba de Lenguaje y Comunicación e implementar una medición directa de escritura, en el corto o mediano plazo, la que podría realizarse una vez cursado 3° año de Enseñanza Media.
- 4 En el área de Ciencia, se recomienda reportar la disciplina elegida y un puntaje adicional que refleje el desempeño en dicha área de los alumnos provenientes de establecimientos secundarios científico-humanista.
- 5 Se recomienda avanzar en el desarrollo e implementación de un plan de estudios que analice en profundidad los desafíos que un eventual uso de la Teoría de Respuesta al Ítem implicaría para la modelación, estimación, precisión, confiabilidad, validez, equivalencia y comunicación del desempeño en la PSU.



# Bibliografía

- ♦ American Educational Research Association, American Psychological Association & National Council on Measurement in Education (AERA, APA & NCME) (2014). *Standards for educational and psychological testing*. Washington, D.C.: American Educational Research Association.
- ♦ Allen, N. L., Donoghue, J. R., & Schoeps, T. L. (2001). The NAEP 1998 technical report, NCES 2001-509, Office of Educational Research and Improvement, National Center for Education Statistics, U.S. Department of Education, Washington, DC: National Center for Education Statistics.
- ♦ Arias, O. Mizala, A. y F. Meneses. (2017). "Brecha de género en matemáticas: El sesgo de las pruebas competitivas" mimeo CEA, Ingeniería Industrial, Universidad de Chile.
- ♦ Bazerman, C. (2013). Understanding the Lifelong Journey of Writing Development. *Infancia y Aprendizaje: Journal for the Study of Education and Development* 36(4), 421-44.
- ♦ Birnbaum, A. (1968). Some latent trait models and their use in inferring a student's ability. En F. M. Lord y M. R. Novick (eds.), *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.
- ♦ DEMRE. (2017). Evidencia preliminar para la reducción de contenidos en la PSU de matemáticas. Santiago: Chile.
- ♦ DEMRE. (2018). Temario de Prueba de Lenguaje y Comunicación. Santiago: Chile.
- ♦ DEMRE. (2018). Key areas for review in the PSU Test Battery. Santiago: Chile.
- ♦ DEMRE. (2016). *Pruebas de Selección Universitaria. Informe Técnico, Volumen I, Características Principales y Composición*. Santiago: Chile.
- ♦ DEMRE. (2017). Propuesta del Departamento de Evaluación Medición y Registro Educativo (DEMRE) al Consejo de Rectores de las Universidades Chilenas (CRUCH) de Mejoras a la PSU. Santiago: Chile.
- ♦ DEMRE. (2018a). Respuesta a carta N° 19/2018 del SUA. Santiago: Chile. Santiago: Chile.

- ♦ DEMRE. (2018b). Follow up on the factor analysis results reported in Table 16 and 17. Santiago: Chile.
- ♦ Donaldson, S. (2007). *Program theory-driven evaluation science: strategies and applications*. Nueva York, N.Y.: Lawrence Erlbaum Associates Publishers.
- ♦ Donaldson, S. & Gooler, L. (2003). Theory-driven evaluation in action: lessons from a \$20 million statewide work and health initiative. *Evaluation and Program Planning*, 26(4), 355–366. [https://doi.org/10.1016/S0149-7189\(03\)00052-](https://doi.org/10.1016/S0149-7189(03)00052-)
- ♦ Geiser, S., & Santelices, M.V. (2007). Validity of high school grades in predicting student success beyond the freshman year: School record vs. standardized tests as indicators of four-year college outcomes. Research & Occasional Paper Series CSHE.6.07. Center for Studies on Higher Education, University of California, Berkeley. Recuperado de [http://cshe.berkeley.edu/publications/docs/ROPS.GEISER.\\_SAT\\_6.12.07.pdf](http://cshe.berkeley.edu/publications/docs/ROPS.GEISER._SAT_6.12.07.pdf).
- ♦ Henson, R. K. & Roberts, J. K. (2006) Use of Exploratory Factor Analysis in Published Research. Common Errors and Some Comment on Improved Practice. *Educational and Psychological Measurement*, Vol. 66(3), 393-416.
- ♦ Hernández, J. (2012) Experiencia Prueba CODICE y el desarrollo de instrumentos complementarios para la selección universitaria. En Santelices, V., Ugarte JJ, y P. Kyllonen (eds.), *Admisión a la Educación Superior: Mediciones Complementarias. Publicaciones Educación Superior Vol 1*, Ministerio de Educación.
- ♦ Kane M. (2016). Explicating validity. *Assessment in Education: Principles, Policy & Practice*, 23(2), 198-211. <https://doi.org/10.1080/0969594X.2015.1060192>
- ♦ Kreiner, S. & Christensen, K. B. (2014), Analyses of model fit and robustness: A new look at the PISA scaling model underlying ranking of countries according to reading literacy. *Psychometrika*, 79(2), 210-231.
- ♦ Lacourly, N., San Martín, J., Amaya, J. & Uribe, P. (2017). INFORME DEMRE 2014-2016. *Technical Report of the Center for Mathematical Modelling of Universidad de Chile*.

- ♦ Manzi, J., Bosch, A., Bravo, D., del Pino, G., Donoso, G. & Pizarro, R. (2010). Validez diferencial y sesgo en la predictividad de las pruebas de admisión a las universidades chilenas (PSU). *Revista Iberoamericana de Evaluación Educativa*, 3(2), 30-48. Recuperado de <http://www.rinace.net/riee/numeros/vol3-num2/art2.pdf>.
- ♦ Martin, M. O., Gregory, K. D., & Stemler, S. E. (Eds.). (2000). *TIMSS 1999 Technical Report*, International Study Center, Boston, MA.
- ♦ Mattern, K., Patterson, B. & Kobrin, J. (2012). The Validity of SAT® Scores in Predicting First-Year Mathematics and English Grades. *Research Report 2012-1*. Recuperado de <https://files.eric.ed.gov/fulltext/ED563105.pdf>
- ♦ Mazzeo, J. & von Davier, M. (2014), Linking scales in international large-scale assessments. En L. Rutkowski, M. von Davier and D. Rutkowski (eds.), *Handbook of International Large-Scale Assessment: Background, Technical Issues, and Methods of Data Analysis*. Boca Raton, FL: CRC Press.
- ♦ Muraki, E. (1992). A generalized partial credit model: Application of an EM algorithm, *Applied Psychological Measurement*, Vol. 16(2), pp. 159-177.
- ♦ Navarro, F. (2018) Más allá de la alfabetización académica: las funciones de la escritura en educación superior. En M.A. Alves & V. Iensen Bortoluzzi (eds.), *Formacao de professores: Ensino, linguagens e tecnologias*. Porto Alegre, Editora Fi.
- ♦ Navarro, F., Ávila Reyes, N. y Gómez Vera, G. (2018) Validez y justicia social: hacia la evaluación situada y significativa en pruebas estandarizadas de escritura. Documento de trabajo.
- ♦ NCME Statement, 2018. Recuperado de [https://higherlogicdownload.s3.amazonaws.com/NCME/c53581e4-9882-4137-987b-4475f6cb502a/UploadedImages/Documents/NCME\\_Position\\_Paper\\_on\\_Theories\\_of\\_Action\\_-\\_Final\\_July\\_\\_2018.pdf](https://higherlogicdownload.s3.amazonaws.com/NCME/c53581e4-9882-4137-987b-4475f6cb502a/UploadedImages/Documents/NCME_Position_Paper_on_Theories_of_Action_-_Final_July__2018.pdf)
- ♦ Oliveri, M. E. y von Davier, M. (2014). Toward increasing fairness in score scale calibrations employed in international large-scale assessments, *International Journal of Testing*, 14(1), pp. 1-21, doi:10.1080/15305058.2013.825265.

- ♦ Organisation for Economic Cooperation and Development (2017). PISA 2015 Technical Report. Recuperado de <http://www.oecd.org/pisa/data/2015-technical-report/>
- ♦ Pearson. (2013). Final Report Evaluation of the Chile PSU.
- ♦ Perelman, C. (2018) Towards a New NAPLAN: testing to the teaching activity. Surry Hills: NSW Teacher Federation.
- ♦ Preiss, D., Castillo, J. C., Grigorenko, H. & Manzi, J. (2013a). "Argumentative writing and academic achievement: A longitudinal study". *Learning and Individual Differences*, 8, 204-211.
- ♦ Preiss, D., Castillo, J. C., Flotts, P. & San Martin, E. (2013b). Assessment of argumentative writing and critical thinking in higher education: Educational correlates and gender differences. *Learning and Individual Differences*, 28, 193-203.
- ♦ Robin, F., & Steffen, M. (2014). Test design for the GRE revised General Test. En Wendler, C. & Bridgeman, B. (Eds.), *The research foundation for the GRE revised General Test: A compendium of studies*, (pp. 3.3.1-3.3.12). Princeton, NJ: Educational Testing Service.
- ♦ Rogers, P., Petrosino, A., Huebner, R. & Hacsí, T. (2000). Program theory evaluation: practice, promise, and problems. *New Directions for Evaluation*, 2000(87), 5-13. <https://doi.org/10.1002/ev.1177>
- ♦ Rossi, P., Freeman, H. & Lipsey, M. (2003). Evaluation. A systemic approach. Thousand Oaks, C.A.: Sage Publications.
- ♦ Santelices, V. Ugarte, J., Flotts, P., Radovic, D., Catalán, X. & P. Kyllonen (2010) Medición de atributos no cognitivos para el Sistema de Admisión a la Educación Superior en Chile. *Revista Iberoamericana de Evaluación Educativa*, 3(2), 49-75.
- ♦ Slomp, D. H. (2008) Harming not helping: The impact of a Canadian standardized writing assessment on curriculum and pedagogy. *Assessing Writing*, 12.
- ♦ Stevens, J. (1996) *Applied Multivariate Statistics for the Social Sciences* (3rd ed.) Mahwah, NJ: Lawrence Erlbaum.



- ♦ Traub, R. E. (1983). A priori considerations in choosing an item response model. En R. K. Hambleton (Ed.), *Applications of item response theory*. Vancouver, BC: Educational Research Institute of British Columbia.
- ♦ Van der Linden, W. J., & Glas, C. A. W. (Eds.). (2010) *Elements of adaptive testing*. New York: Springer.
- ♦ Von Davier, M. & Sinharay, S. (2014). Analytics in international large-scale assessments: Item response theory and population models. En L. Rutkowski, M. von Davier, & D. Rutkowski (Eds.), *Handbook of international large-scale assessment: Background, technical issues, and methods of data analysis*. Boca Raton, FL: CRC Press.
- ♦ Yan, D., von Davier, A. A., & Lewis, C. (Eds.). (2014) *Computerized multistage testing*. Boca Raton, FL: CRC Press.
- ♦ Wikström, C., Wikström, M. (2017). Group differences in student performance in the selection to higher education: tests vs grades. *Frontiers in Education*, 2(45).<https://doi.org/10.3389/feduc.2017.00045>.
- ♦ Zwick, R. (2017). *Who gets in? Strategies for fair and effective college admissions*. Cambridge, M.A.: Harvard University Press.



# Apéndice: Bitácora de actividades y comunicaciones del Comité

**Apéndice:** bitácora de actividades y comunicaciones del Comité

**Miembros:** los miembros activos del Comité de Expertos son Salomé Martínez, Alejandra Mizala, Verónica Santelices, y Rebecca Zwick. El siguiente es un resumen de las reuniones y comunicaciones del comité con SUA.

## Reuniones:

**Reunión vía Skype, 22 de junio del 2018:** Las participantes fueron Salomé Martínez, Alejandra Mizala, Verónica Santelices y Rebecca Zwick. El comité discutió su rol y su respuesta a la solicitud de María Elena González de fecha 13 de junio del 2018 (ver más abajo).

**Reunión vía Skype, 24 de septiembre del 2018:** Las participantes fueron Alejandra Mizala, Verónica Santelices y Rebecca Zwick. El comité discutió el informe y determinó quién realizaría el borrador de cada sección.

**Reunión vía Skype, 26 de noviembre del 2018:** Participaron Salomé Martínez, Alejandra Mizala, Verónica Santelices y Rebecca Zwick. El comité revisó cada sección del borrador del informe.

## Comunicaciones con el SUA:

**12 de junio:** se recibe de parte de Tatiana Diener la versión en inglés del documento de SUA "Committee of Experts Unified Admissions System" (Comité de Expertos del Sistema Único de Admisiones), de fecha 31 de mayo, con la propuesta de una cronología de trabajo para el comité y una lista de productos esperados.

**13 de junio de 2018:** Se recibe un mensaje de María Elena González en el que solicita las respuestas del comité al documento del 31 de mayo e indica que las solicitudes de asistencia se pueden dirigir a Elisa Zenteno.



**26 de junio de 2018:** El comité envía un mensaje a María Elena González informando, entre otros:

*"Hemos concluido que nuestra primera prioridad debe ser abordar los siete temas descritos en el documento de abril del 2018 "Áreas Clave que Revisar en la Batería de pruebas de la PSU". Creemos que podemos tener un informe sobre estos temas para nuestra próxima reunión de comité de enero del 2019, asumiendo que podemos contar con los materiales que necesitamos."*

**29 de junio del 2018:** El comité envía una solicitud de varios informes y análisis a Elisa Zenteno.

**5 de julio del 2018:** El comité recibe un mensaje de Eliza Zenteno pidiendo aclarar nuestro mensaje del 26 de junio de 2018 y enfatizando la importancia de temas concernientes a los estudiantes de establecimientos técnico-profesionales.

**20 de julio del 2018:** El comité envía un mensaje a Elisa Zenteno y a María Elena González proponiendo considerar los temas concernientes a estudiantes de establecimientos técnico-profesionales en el 2019.

**23 de julio del 2018:** el comité recibe un mensaje de Elisa Zenteno en el que señala que se está de acuerdo con la propuesta del comité de considerar los temas concernientes a estudiantes de establecimientos técnico-profesionales en el 2019, además de informar que estaba en vías de enviar los documentos que habíamos solicitado el 29 de junio.

**30 de julio del 2018:** Elisa Zenteno envía al comité los documentos solicitados en español.

**6 y 7 de agosto del 2018:** El comité intercambia comunicaciones con Elisa Zenteno con respecto del contenido de los documentos enviados el 30 de julio y la necesidad de que éstos sean traducidos al inglés.

**29 de agosto del 2018:** El comité recibe un mensaje de parte de Eliza Zenteno con la traducción al inglés solicitada.

**1 de octubre de 2018:** El comité envía un mensaje a Elisa Zenteno con preguntas técnicas sobre los documentos del DEMRE.



**22 de octubre del 2018:** María Elena González reenvía a Leonor Varas el mensaje del comité de fecha 1 de octubre.

**9 de noviembre del 2018:** El comité recibe un mensaje de María Elena González en el que reenvía una respuesta del DEMRE a las preguntas que enviara el comité el 1 de octubre del 2018.

**11 de noviembre del 2018:** El comité recibe un mensaje de María Elena González con un informe sobre los estudiantes que asisten a establecimientos técnico-profesionales.

**13 de noviembre del 2018:** El comité recibe un mensaje de María Elena González con una propuesta de agenda y cronograma de trabajo para el comité en lo sucesivo.

**26 de noviembre del 2018:** se discute la agenda y el cronograma de trabajo propuesto por SUA. El SUA propone que solo un miembro del Comité viaje a Arica a presentar el informe el 10 de enero. Los miembros del comité concuerdan en que todos los miembros deben estar presentes cuando se discuta el informe del Comité y que, por lo tanto, no es posible realizar la presentación el 10 de enero como lo sugiere el SUA. Los miembros del comité acuerdan entregar un informe el 20 de diciembre para su edición y traducción a cargo del SUA y buscar una fecha alternativa en diciembre para presentar el informe a un grupo de rectores en Santiago.

**26 de noviembre al 11 de diciembre del 2018:** Verónica se comunica con Tatiana Diener y María Elena para agendar la reunión. Miembros del SUA informan que sostener la reunión durante diciembre o los quince primeros días de enero era imposible, puesto que la mayoría de los rectores estarían ocupados con temas de finanzas y actividades de fin de año.

El 28 de enero, se presentan las principales conclusiones del informe a un subgrupo de autoridades académicas.







CONSEJO DE RECTORES DE  
LAS UNIVERSIDADES CHILENAS