

Anexo N° 15/2013

Informe Proceso Evaluación PSU

Enero 2012 a enero de 2013

Dra. María Elena González

Marzo de 2013



Antecedentes

Contraparte Técnica

- MINEDUC
 - Juan Bravo
 - Francisca Dussailant
 - Francisco Lagos
- CRUCH
 - Hector Allende (UTFSM)
 - María Elena González (UTA)
 - Victor Salinas (USACH)

Consultores MINEDUC

- Christina Stage
- Ronald Hambleton



Acerca de Person

- Alta calificación de expertos responsables.
- Diseño metodológico con diversidad de técnicas para recolección, análisis de datos y resultados (cuali y cuantitativo).
- Intervención de diversos equipos, según el objetivo a cubrir.
- Disparidad con relación a la profundidad del trabajo según objetivo y al aprovechamiento de los datos y fuentes de información disponibles.
- Enfoque principalmente norteamericano, para estándares y como aproximación general.



Acercas del proceso

- Complejo: por el grado dispar de dominio de la temática, por las características de los actores involucrados y por la ausencia de otros.
- Extenso: requirió de una dedicación mucho mayor a la presupuestada inicialmente.
- Colaborativo: a partir de un enfoque muy técnico con la contraparte Mineduc y sus asesores.
- Accidentado: por las dificultades para que Pearson comprendiera el contexto educacional chileno, fuera sensible a éste y se ajustara a los requerimientos formales.



Acerca de la evaluación

Consideró los siguientes aspectos:

- evaluación de los procesos de construcción de las pruebas
- análisis de la puntuación de las pruebas, comunicación y uso de los resultados
- estudio de la validez

Aún cuando el proceso fue muy dificultoso, finalmente los resultados dan cuenta de un trabajo sistemático, riguroso y útil para la mejora del proceso



Acerca de las recomendaciones

Es importante diferenciar entre aspectos técnicos objetados, sugerencias para mejora y recomendaciones que involucran aspectos de política.

- Los aspectos técnicos objetados son de dos tipos:
 - aquellos relativos a estándares internacionales
 - aquellos referidos a opciones teóricas, metodológicas o culturales
- Las sugerencias para mejora prioritariamente se refieren a aspectos procedimentales, normativos o metodológicos
- Las recomendaciones que involucran aspectos de política tienen un impacto institucional, social o económico.



Recomendaciones Generales

- Contar con mayor y mejor documentación de los procesos asociados con la construcción, aplicación y análisis de las Pruebas de Selección Universitaria:
 - Interna: Desarrollo de manuales, documentos con normativas, estandarización de procesos, etc.
 - Externa: informes y reportes para “educación de los destinatarios” y clarificación de propósitos y usos.
- Instalar procedimientos de control de calidad: validaciones externas, auditorias, certificaciones.
- Mantener una línea de investigación permanente que asegure pertinencia y fundamente la toma de decisiones
- Capacitar permanentemente al recurso humano de manera de asegurar alta calificación

Acercas de los procesos de construcción de las pruebas





Desarrollo de ítems y pruebas

Aspectos técnicos objetados

- No contar con expertos externos al DEMRE para revisión y análisis de especificaciones para construcción de la prueba
- Utilizar diferentes escalas y criterios de calificación para la aprobación o el rechazo de ítems
- No contar con una descripción detallada de los criterios para la actualización del banco de ítems
- Uso de la “corrección por adivinación”
- Procedimientos para la toma de decisiones en la selección de ítems revisados.

Sugerencias para mejora

- Mejorar el proceso de revisión con tareas más específicas a los coordinadores
- Detallar más la información del proceso (identificar etapas y participantes)
- Desarrollar procesos estandarizados de capacitación y certificación de redactores de ítems
- Implementar procedimientos de auditoría que permitan detectar posible puntos de filtración de información
- Desarrollar chequeos sistemáticos que permitan evaluar el cumplimiento de los principios de confidencialidad y derechos de autor
- Contar con protocolos para el control del material en su distribución a los lugares de administración de pruebas.



Recomendaciones que involucran aspectos de política

- Mayor información pública acerca del proceso (de qué, para qué y a quienes).
- Situación de “inequidad” para modalidad Técnico-Profesional:
 - Alineación de las pruebas a los CMO de primeros y segundo medio.
 - Ponderación mayor para la modalidad Científico-Humanista que para la modalidad Técnico-Profesional.
- Asegurar equidad a distintas poblaciones (materiales e instalaciones)



Pilotaje de las preguntas

Aspectos técnicos objetados

- Propósito general del pilotaje de ítems
- Selección de los ítems
- Representatividad de la muestra
- Expectativas psicométricas de los resultados del pilotaje

Recomendaciones que involucran aspectos de política

- Revisión proceso completo
- Redefinir el propósito del pilotaje
- Rediseñar el pilotaje

Ensamblaje de las pruebas definitivas

Aspectos técnicos objetados

- Items “anclas”

Sugerencias para mejora

- Desarrollar estudios que entreguen información para decidir cómo ajustar el contenido de la PSU, así como también sus aspectos formales (edición), condiciones de administración, difusión y uso de los resultados de las pruebas.
- Documentar los criterios utilizados para armar las pruebas
- Utilizar un marco de Teoría de Respuesta al Ítem (TRI) que considere el uso de medidas de precisión.
- Automatizar el sistema de ensamblaje de pruebas

Recomendaciones que involucran aspectos de política

- Comunicar claramente el uso pretendido de los resultados de la PSU para disminuir el mal uso de estos resultados (p.e: desagregación de los puntajes por institución educativa y las comparaciones con otras instituciones)
- Someter la decisión de colocar más énfasis en la modalidad Científico-Humanista que en la modalidad Técnico-Profesional a una validación externa e incluir en los marcos teóricos de las pruebas la justificación de esta decisión.



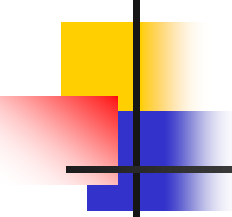
Administración del banco de ítems

Aspectos técnicos objetados

- Insuficiente información técnica acerca del banco de ítems (criterios psicométricos utilizados para la mantención y actualización del banco).
- Gestión del banco de ítems no se condice con las mejores prácticas a nivel internacional (se permite que los ítems sean modificados después del pilotaje, una vez que ya fueron ingresados al banco)

Sugerencias para mejora

- Realizar una inspección detallada del sistema del banco de ítems
- Mejorar y ampliar documentación del proceso
- Documentar los criterios para la revisión de ítems.



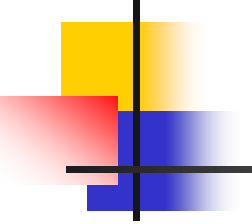
Selección de ítems del piloto para la forma operacional

Aspectos técnicos objetados

- Excepciones a las normas internacionales para selección de ítems, tales como el rango de dificultad TRI, y la tasa de omisión.
- Subjetividad en la selección de los ítems que puede afectar la comparabilidad de las formas de prueba entre administraciones.
- Ítems pilotados que son editados o modificados en forma previa a la aplicación operacional

Sugerencias para mejora

- Documentar en mayor detalle los procesos de selección de ítems considerando conjuntamente los distintos criterios.
- Modificar el software empleado en la construcción de ítems a fin de que los ítems desarrollados puedan ser cargados en el banco con la historia de sus modificaciones y usos en la administración.
- Aumentar la diversidad institucional y geográfica de los participantes para los procesos de revisión



VARIABLES ASOCIADAS AL DESEMPEÑO DIFERENCIAL DE ÍTEMS

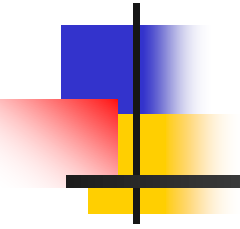
Aspectos técnicos objetados

- Altas discrepancias en el desempeño de ítems entre las administraciones piloto y operacionales, información que no es utilizada
- No considerar los resultados DIF del piloto

Sugerencias para mejora

- Eliminar la “corrección por adivinación”
- Revisar las condiciones del pilotaje (autoselección, tasas de motivación diferenciales, y representatividad de la muestra piloto)
- Ampliar los análisis DIF a los siguientes subgrupos: región, status socioeconómico y modalidad curricular.

Puntuación de las pruebas, comunicación y uso de los resultados





Cálculo de los puntajes estandarizados

Aspectos técnicos objetados

- Fórmula para corregir por adivinación
- Falta de un mecanismo para mantener la escala de la PSU a lo largo de los procesos de admisión
- Desconocimiento de la estructura, normas y propiedades psicométricas de los puntajes NEM
- Graves problemas en la documentación y adecuación técnica de las escalas

Sugerencias para mejora

- Abandonar la práctica de la corrección por adivinación y considerar la teoría de respuesta al ítem como un enfoque alternativo para tratar el comportamiento de adivinación del postulante

Recomendaciones que involucran aspectos de política

- Reemplazar el NEM con mediciones estandarizadas de desempeño académico en la enseñanza media tales como puntajes a partir de pruebas administradas nacionalmente
- Revisar completamente el sistema de puntajes de postulación
- Desarrollar pautas sobre los usos pretendidos e interpretación de las escalas de puntaje de la PSU y puntajes estandarizados
- Mantener una agenda de investigación para estudiar la estabilidad año tras año de las escalas primarias y secundarias de la PSU.



Confiabilidad y precisión de los puntajes de las pruebas

Aspectos técnicos objetados

- Informe de confiabilidad de la PSU es limitado ya que no proporciona información para establecer el tamaño aceptable de errores de medida, especialmente importante en las regiones críticas de la escala de la PSU en las que se toman las decisiones de admisión y otorgamiento de becas

Sugerencias para mejora

- Se recomienda el uso de TRI
- Revisar escala NEM
- Se recomienda utilizar el método Hofstee para fijar el puntaje de corte para fines de becas.

Modelo para asignación de beneficios sociales

- Se recomienda utilizar el método Hofstee para fijar el puntaje de corte para fines de becas.



Prueba de Ciencias

Aspectos técnicos objetados

- Equivalencia de puntajes de partes de pruebas (Biología, Física y Química)
- Procedimiento de regresión

Sugerencias para mejora

- Documentar la decisión de reportar un puntaje único para Ciencias, su racionalidad y la investigación que fundamenta dicha decisión.
- Enlazar los puntajes totales, más que usar el proceso actual de enlazar puntajes en secciones opcionales y luego sumar los puntajes enlazados con los puntajes de la porción común.
- Considerar el uso de métodos típicos de enlace
- Verificar la comparabilidad de puntajes (resultados de regresiones tales como promedios de notas universitarias sobre puntajes totales para los grupos que toman los módulos opcionales de Biología, Química y Física)

Recomendaciones que involucren aspectos de política

- Desarrollar pruebas separadas para Biología, Física y Química



Uso de los modelos TRI para calibrar, equiparar y puntuar las pruebas.

Aspectos técnicos objetados

- Documentación imprecisa y uso de terminología incorrecta
- Discrepancia entre lo documentado y lo que efectivamente se realiza
- El programa de pruebas de la PSU no mantiene una escala de información sobre una base anual ni calibra los ítems piloto en una escala común.
- Construcción y uso de conjuntos ancla

Sugerencias para mejora

- Garantizar la comparabilidad de las pruebas entre distintos años.
- Se recomienda la introducción de métodos que permitan equiparar los resultados de las pruebas entre años y formas operacionales aplicadas en un mismo año

Evaluación de software

Sugerencias para mejora

- Automatizar lo más posible los procesos relativos a la base de datos y estandarizar los análisis para eliminar o reducir la cantidad de intervención manual requerida
- Conservar las estadísticas de todas las administraciones de un ítem y poder visibilizarlas en orden cronológico.



Entrega y claridad de la información

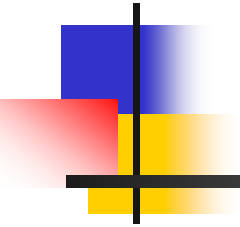
Aspectos técnicos objetados

- Reportes de resultados de la PSU no son interpretados adecuadamente por las distintas audiencias a los que están dirigidos (principalmente, estudiantes y docentes).

Sugerencias para mejora

- Rediseñar los informes
- Suspender la entrega de Informes Estadísticos de la PSU para los docentes

Validez de las pruebas y de sus resultados





Estructura interna

Aspectos técnicos objetados

- Si bien los análisis realizados permitieron identificar la presencia de un factor principal o dimensión latente para cada prueba de la PSU, las pruebas presentan evidencia de Funcionamiento Diferencial de Pruebas (por modalidad, dependencia, y contenido)

Sugerencias para mejora

- Realizar análisis adicionales para entender mejor el DTF
- Adoptar el marco TRI



Validez de contenido

Aspectos técnicos objetados

- Bajo nivel de alineamiento de la PSU con los Objetivos Fundamentales (OF) y los Contenidos Mínimos Obligatorios (CMO) del currículum
- Desconexión entre el propósito y uso de la PSU para seleccionar estudiantes para las universidades y su contenido, desconexión mayor para los estudiantes de la modalidad curricular Técnico-Profesional

Recomendaciones que involucran aspectos de política

- Revisar la política de usar el Marco Curricular como la base para el desarrollo de los marcos para las pruebas de la PSU.
- Desarrollar las formas de pruebas de la PSU para el 2013–2014 basados en la plena implementación de la reforma curricular del año 2009
- Revisar los tipos de ítems usados en las pruebas de la PSU para abordar el nivel bajo de complejidad cognitiva percibido que se encuentra en las pruebas (foco sobre habilidades de pensamiento de orden inferior y los Contenidos Mínimos Obligatorios).



Análisis de trayectorias

Aspectos técnicos objetados

- Equidad del puntaje en la prueba de la PSU entre años adyacentes se pone en riesgo al realizar cambios en las especificaciones de las pruebas de la PSU (Matemática aumentó las preguntas en el proceso de postulación del año 2012).
- **Sugerencias para mejora**
 - Llevar a cabo una equiparación del puntaje de la prueba anualmente.
 - Monitorear las tendencias de los resultados en poblaciones relevantes de estudiantes



Capacidad predictiva

Sugerencias para mejora

- Desarrollar estudios que respalden el uso de la PSU en los procesos de admisión, incorporando eventualmente nuevas variables a este proceso
- Incorporar criterios alternativos para la investigación de la validez predictiva, más allá de las notas y tasas de graduación



Consideraciones finales

- La mayor parte de los aspectos técnicos objetados merecen atención, ya sea para corregirlos o para dar evidencia que sustente la utilización de un criterio diferente al señalado en la evaluación
- Las sugerencias de mejora son bastante coincidentes con las recomendaciones del CTA y las que atañen al DEMRE son abordables aunque algunas requieren de ajustes mayores
- Aquellas recomendaciones que involucran aspectos de política son de mayor envergadura y tienen mayor impacto en la opinión pública. Con respecto a ellas es importante relevar los elementos técnicos que las sustentan y que no necesariamente fueron considerados por el equipo evaluador



Gracias
