

Test informatizados y su contribución a la acción evaluativa en educación

Computerized Tests and their Contribution to Evaluation in Education

Carlos Rodríguez Garcés
Universidad del Bío Bío
carlosro@ubiobio.cl

Marlene Muñoz Sepúlveda
Universidad del Bío Bío
marlene@ubiobio.cl

Víctor Castillo Riquelme
Universidad del Bío Bío
vicastil@alumnos.ubiobio.cl

Resumen

El sistema educativo nacional no sólo evidencia debilidades en su capacidad para instalar competencias cognitivas y procedimentales en su alumnado, sino que también en la calidad de sus sistemas evaluativos. Si bien evaluar es una tarea engorrosa y compleja para el profesorado, los Test Informatizados proporcionan una poderosa herramienta para mejorar en pertinencia, validez y rigor en la acción evaluativa. Los Test Informatizados son aplicaciones versátiles y amigables, que no sólo facilitan la tarea del docente al evaluar, sino que al permitir usar ítems calibrados y de métrica probada hace más objetiva y justa la certificación del dominio de la competencia, incorporando los nuevos aportes de la teoría de respuesta al ítem. Por ello, realizar acciones de apropiación tecnológica puede tributar fuertemente al mejoramiento de los niveles de eficiencia y calidad del proceso educativo en general.

Palabras clave

Test informatizados, generadores de test, evaluación, banco de ítems, sistemas evaluativos

Abstract

The national educational system not only reveals weaknesses in its ability to install cognitive and procedural skills in its students, but also in the quality of its assessment systems. While evaluation is a cumbersome and complex task for the teachers, the Computerized Tests provide a powerful tool to improve the evaluative action in terms of relevance, validity and security. The Computerized Tests are versatile and user-friendly applications, which not only facilitate the task of the teacher to assess, but in allowing the use of calibrated and tested metric items, it makes the certification of the proficiency attainable, more objective and fair; incorporating new contributions of the item response theory. Therefore, carrying out actions of technological appropriation; in general, may strongly increase the improvement of the efficiency levels and the quality of the educational process.

Keywords

Computerized test, generators test, evaluation, item bank, evaluative systems.

Introducción

Con el retorno a la democracia el Estado de Chile ha definido a la Educación como una estrategia central para el desarrollo, cristalizándose en iniciativas tales como la expansión de la enseñanza obligatoria, la instauración de la jornada escolar completa, la articulación de programas de fortalecimiento docente, una incipiente institucionalidad acreditadora de los procesos de formación inicial pedagógica (Programa INICIA), un sistema de evaluación de profesores en ejercicio (Evaluación Docente), de certificación de la oferta educativa (Consejo Nacional de Acreditación), así como en una fuerte inversión en tecnologías para la enseñanza. Conjunto de iniciativas que se suman a un reforzado Sistema de Medición de la Calidad de la Educación y los aprendizajes (SIMCE). Esfuerzos que si bien han logrado impactar sobre los niveles de cobertura, retención y asistencia del alumnado, están escasamente asociados a los niveles de logro escolar, el cual en el mejor de los casos evidencia modestas mejoras según lo constatado por las mediciones del SIMCE. En efecto, el sistema educativo chileno presenta profundas debilidades en su capacidad para instalar competencias cognitivas y procedimentales sobre su alumnado, especialmente en quienes provienen del sistema público municipal y familias con escaso capital cultural. Al respecto, las pruebas de rendimiento internacionales no solo posicionan al país en el grupo de menores logros de aprendizaje, muy por debajo al de países con igual nivel de desarrollo o ingreso per cápita, sino que también como uno de los países con mayor desigualdad educacional (Rosas & Santa Cruz, 2013). Desigualdad educativa que se constituye en el correlato natural de una profunda desigualdad socioeconómica y segregación territorial, a pesar de los sobresalientes y sostenidos indicadores macroeconómicos, y una eficiente política de reducción de la pobreza en los últimos 20 años (Contreras & Gallegos, 2011).

La política educativa basada en vouchers o subsidio a la demanda instalada en los 80' bajo la dictadura militar y continuada con cosméticas reformas en democracia segmentó al sistema escolar según la capacidad de pago de las familias, desregulando el mercado educativo con escasa intervención estatal. Como consecuencia de ello aumentó la matrícula en el sistema particular subvencionado, creado durante la reforma, sector al que migran masivamente los alumnos de mejor nivel socioeconómico y capital cultural que asistían a las escuelas públicas (Valenzuela, 2006; Mayol, Araya, & Azócar, 2011; Bellei, 2013). Evento que, profundizado por la capacidad institucional de seleccionar a los estudiantes, produjo un natural descreme (Epple & Romano, 1998) al captar al alumnado de mejor perfil educativo sin aportar mayor valor agregado y generar externalidades negativas a los establecimientos públicos, privándolos de las bondades del efecto grupo (McEwan & Carnoy, 2000) y restringiendo con ello la calidad de la educación a la que acceden los estudiantes más vulnerables (Hsieh & Urquiola, 2006).

El sistema educativo nacional evidencia problemas de equidad, inclusión y profundiza las brechas de calidad y segregación socioeducativa, dando cuenta de un modelo incapaz de instalar competencias y aprendizajes como valor agregado en su alumnado, especialmente en los más pobres; debilidad que es también extensible a los mismos sistemas evaluativos, caracterizados por la escasa pertinencia y capacidad predictiva que evidencian. En efecto, los sistemas evaluativos que las escuelas implementan para medir el nivel de aprovechamiento del currículum en determinados contextos educativos no se condice con la debilidad de los aprendizajes que muestran las evaluaciones externas,

evidenciando dificultades para discriminar eficientemente los niveles de logro del estudiantado con las garantías métricas de fiabilidad y validez exigibles a todo instrumento de medición. Así por ejemplo, al comparar el rendimiento escolar en relación al puntaje obtenido en las pruebas de selección universitaria, instrumento en sintonía con el currículum escolar, se observa que los puntajes asignados al primero son en promedio superiores a los obtenidos en cualquier test de selección, situación especialmente marcada en los establecimientos municipales, además de observarse una progresiva reducción del rango de las calificaciones por elevación de los valores inferiores (Rodríguez & Jarpa, 2014), haciendo de esta manera al rendimiento escolar del estudiante que asiste a establecimientos de mayor vulnerabilidad, perder capacidad para predecir rendimiento en las pruebas de selección universitaria. En los establecimientos particulares pagados, en cambio, el rendimiento escolar alcanzado correlaciona fuertemente con los puntajes en dichas pruebas, dando cuenta de sistemas evaluativos más confiables y de mayor capacidad predictiva.

Aun cuando la evaluación es considerada una fase decisiva en la acción pedagógica y que adquiere profunda relevancia en los modelos educativos actuales, las prácticas evaluativas docentes han privilegiado tareas reproductivas, de control instrumental y memorístico (Stiggins, 2004) configurándose como un proceso desvinculado de la enseñanza (Celman, 2005). A ello se suma la relevante evidencia en torno a que la evaluación, desde una perspectiva didáctica, está fuertemente relacionada con la emisión de juicios valorativos de carácter infundado sobre el trabajo y acciones de los estudiantes, contruidos desde los marcos axiológicos y epistemológicos del profesor de carácter subjetivo (Camilloni, 2005; Celman, 2005; Litwin, 2005; Darling-Hammond, 2001; Eisner, 1998). Esto hace poco rigurosa la acción evaluativa no siendo la calificación un real reflejo del nivel de aprovechamiento educativo logrado por el estudiante, sino más bien el fruto de diferenciadas políticas de asignación de notas, segmentados niveles de exigencia, así como una tendencia cada vez mayor a hacer subir artificialmente las calificaciones de los estudiantes (Contreras, Gallegos, y Meneses, 2009; Díaz, Himmel, y Maltes, 1990). Situación que tiende a ser más delicada dada la importancia que evidencian las calificaciones en los sistemas de selección a la educación superior como el implementado en Chile, donde uno de cada dos puntos del puntaje de postulación de un alumno a la Educación Superior es atribuible al factor rendimiento escolar, sea como Notas de Enseñanza Media (NEM) o ranking, situación que vulneraría la capacidad de dicho instrumento para seleccionar a los más aptos de forma de aprovechar adecuadamente la formación a recibir y que obtengan buenos resultados académicos (Donoso y Schiefelbein, 2007).

La pérdida de fiabilidad de la evaluación escolar, en especial cuando sobreestima el rendimiento del estudiante da cuenta de una doble problemática. Por un lado, agrega una complejidad adicional a la labor pedagógica al no detectar las debilidades de las conductas de entrada de los estudiantes que transitan a un determinado nivel, profundizando el retraso pedagógico al inhibir la acción remedial. Por otro lado, evidencia las dificultades que tiene la escuela en la certificación de sus sistemas evaluativos donde las calificaciones no serían el real reflejo del nivel de rasgo o aprovechamiento educativo, sino un indicador sobreestimado artificialmente que invisibiliza frente a la comunidad escolar los déficits en la calidad de los aprendizajes.

Medir los conocimientos que los estudiantes han alcanzado no se constituye en sí mismo en expresión de una mejor calidad de la educación, pero sí en una herramienta útil para cuantificar el nivel de logro alcanzado y discernir respecto de quiénes han adquirido las competencias y habilidades necesarias para otorgar certificación o promoción, así como también identificar cambios necesarios a introducir para mejorar los procesos de enseñanza. En consecuencia, enseñanza, aprendizaje y evaluación son procesos, si bien organizados e implementados en momentos diferentes, ineludiblemente complementarios e inseparables.

Aunque la evaluación es una medición difícil de determinar con exactitud, las investigaciones psicométricas dan cuenta que es posible reunir información definida con precisión y rigor, planificada, objetiva e interpretable en la medida que esté sustentada en ítems de calidad métrica probada. Sólo en esa medida la evaluación posibilitará el estudio de la competencia, calibrar los niveles de desempeño y con ello posibilitar una adecuada toma de decisiones acerca de los individuos y de los procesos implementados (Cronbach, 1963; Beeby, 1977; Thorndike, 1989; Nunnally 1995)).

La evaluación, como expresión de medición, es una fase decisiva en la acción pedagógica que adquiere profunda relevancia en los modelos educativos actuales y se constituye en un dominio de primer orden, especialmente para el diseño instruccional (Seels y Richey, 1994). Por una parte, en modelos centrados en el aprendizaje que propician la autonomía del estudiante, la evaluación brinda el estado de situación presente y lograda por el alumnado en la comprensión de un contenido o manejo de un procedimiento. Así mismo, al hacerse esta evaluación estandarizada y externa, calibra y rankea a los sistemas educativos y modalidades de enseñanza implementadas, siendo expresión de la eficiencia interna y calidad del proceso educativo.

No obstante los importantes atributos de polifuncionalidad y propiedades, una evaluación no está carente de dificultades por cuanto se trata de una medición relativa, probabilística e indirecta. Es decir, está referida en unidades arbitrarias, generalmente en la ejecución de un grupo de individuos, aunque mensurable no libre de error, y siempre expresada a partir de indicadores externos supuestamente relacionados con el rasgo de interés. A ello debemos agregar las fuertes aprehensiones en torno al control social o institucional, la selectividad y clasificación de los sujetos que como propósito extrínseco estas evaluaciones pueden tener.

En consecuencia la evaluación, en especial la generación de Test, siempre ha implicado una dificultad para la acción docente, pues el educador no sólo debe diseñar ítems válidos y confiables y estructurarlos coherentemente, sino que además calibrarlos en base a su dificultad y valoración. Es decir, en una evaluación los ítems o preguntas deben ser capaces de dimensionar cada uno de los niveles en que está expresado el rasgo evaluado, lo que implica la elaboración de un número adecuado y válido de ítems que garanticen que la puntuación obtenida es expresión del nivel de dominio o presencia del rasgo medido. Además se deben ponderar los ítems de acuerdo a sus niveles de dificultad, controlar el sesgo de adivinación, castigar eficientemente el azar, evitar la copia, etc. a fin de generar métricas evaluativas eficientes. Así mismo debe procurar reducir el sesgo terminológico y las connotaciones equivocadas del uso de las palabras al momento de formular las preguntas y sus categorías de respuestas, en el entendido que

el cómo y el qué preguntar tienden a ser decisiones arbitrarias y subjetivas del profesor (Morales, 2012). Riesgos de evaluaciones inadecuadas que pueden reducirse al aumentar el número de ítems, calibrando su dificultad, asignándole diferenciadas puntuaciones, aumentando el número de evaluaciones, así como aleatorizando tanto el ítem como las alternativas.

Todo este trabajo y cuidados que se han de tener pueden resultar por parte del docente engorroso e implicar un gran consumo de tiempo. Situación que no pocas veces tiende a ser desbordante, en especial cuando aumenta el número de evaluaciones, cursos y estudiantes. Investigaciones recientes señalan que docentes carecen de tiempo suficiente para realizar tareas pedagógicas referentes a la preparación de clases, elaboración y procesamiento de evaluaciones, preparación de material didáctico y planificación de actividades (De los Ríos, Valenzuela, Bellei, & Sevilla, 2009; Beyer, y otros, 2010). A ello se suman los déficits que evidencian los procesos formativos del profesorado, donde el nivel de desempeño del docente en ejercicio tiende a ser resultado principalmente del ejercicio docente fruto del ensayo y error, más que de la acción directa en la formación inicial recibida (Ortúzar, Flores, Milesi, & Cox, 2009; Rodríguez & Castillo, 2014). Además, modelos centrados en la acreditación de competencias individuales e institucionales, así como en la explicitación de resultados objetivos de los procesos de intervención realizados, hacen de la evaluación un tema especialmente sensible, presionando a las instituciones educativas a centrar la toma de decisiones sobre la base de información evaluativa clara y confiable. La acción evaluativa genera consecuencias personales y sociales, por ello las mayores exigencias psicométricas a los test. La necesidad de medir bien, según Olea, Abad, y Barrado (2010) se constituye en un requisito ineludible que está impulsando el desarrollo de nuevos tipos de tests y nuevos modelos psicométricos para estudiar las garantías que ofrecen sus aplicaciones. Solo el dominio y conocimiento del contenido disciplinar y de los factores implicados en la teoría de la medición permitirán subsanar o reducir los riesgos de una evaluación escasamente pertinente y confiable; y si bien, ésta puede ser concebida como una tarea compleja y delicada, las tecnologías de la información y comunicación ofrecen un conjunto de nuevas aplicaciones tendientes a facilitar esta tarea.

En efecto, las TIC en el ámbito educativo en general, y en la acción evaluativa en particular, evidencian un avance cada vez más notable, por lo que obviar su utilidad, sea por desconocimiento o subestimar su potencial didáctico, restaría significatividad a la acción educativa en un estudiantado cada vez más familiarizado a su uso cotidiano, profundizaría la brecha digital profesor-alumno (Gómez, Rodríguez, & Ibarra, 2013) y reduciría la eficiencia de la gestión educativa. Es en este contexto en que los Test Informatizados y en especial los Generadores de Test se convierten en un poderoso aporte que las TIC proporcionan a la acción docente. La evolución y abaratamiento de los costes de la tecnología informática ha ayudado a incrementar la eficiencia de la evaluación docente mediante la incorporación de aplicaciones informáticas amigables, así como la inclusión de nuevas funcionalidades, como son la generación automática de ítems, la aplicación adaptativa de un test o la corrección automática de respuestas (Olea, Abad, y Barrado, 2010).

Los test informatizados, atributos y requerimientos

Los Test Informatizados son evaluaciones generadas por un software especializado denominado Generador de Test que, siendo administradas en la generalidad de las veces por un ordenador, busca medir y graduar en forma sencilla y rápida los conocimientos, habilidades y aptitudes que presenta una persona, por ejemplo un estudiante. Como herramienta ha dejado de ser una técnica experimental para constituirse en un procedimiento utilizado por cientos de programas de medición que evalúan a millones de personas cada año, estimándose que un futuro ser evaluado por un ordenador sea tan natural como la evaluación escrita (Davey, 2005). No solo facilitan el trabajo desde el punto de vista de la reducción del tiempo, sino que brinda la posibilidad cierta de mejorar la validez y confiabilidad de la métrica evaluativa, reactivándose todo un campo investigativo referente a la aplicación de la informática a la teoría de los test (Olea, Ponsoda, y Prieto, 1998). Avances experimentados por la metodología evaluativa cimentados precisamente en el desarrollo tecnológico que sustituye la concepción meramente instrumental por una orientación científico técnica de obtener información para la toma de decisiones docentes, otorgando credibilidad y soporte ético a evaluados y evaluadores.

En términos generales el concepto Test Informatizados se reserva a aquellos instrumentos de propiedades psicométricas conocidas, cuyos ítems se presentan y responden desde un computador a fin de estimar el nivel del rasgo o rendimiento de una persona (Olea, Ponsoda, y Prieto, 1998). En consecuencia la correcta aplicación, así como la justicia en las decisiones que se toman a partir de los resultados que éstas generan no son observadas por el mero hecho de ser formuladas desde un terminal computacional, sino más bien porque la elaboración del ítems está en sintonía con una teoría psicológica y un modelo estadístico determinado (Muñiz, 2000).

Estos generadores o editores de Test, en cuanto a software informático especializados en la confección de pruebas, son herramientas modulares que permiten configurar varios parámetros sobre las preguntas o ítems y los tests (Romero, 2006). Tienen por principal característica ser aplicaciones versátiles, profundamente amigables con el usuario, siendo muy sencillo de utilizar incluso en aplicaciones de naturaleza compleja como los requerimientos de los Test basados en la Teoría de Respuesta al Ítem para la estimación de los parámetros (Muñiz, 2010) o los Test Adaptativos Informatizados (TAIs), que si bien pertenecen a la familia de los Test Informatizados, tienen especificidades no siempre operacionables desde cualquier Generador de Test. Si bien sus potencialidades dependerán del tipo de aplicación o software específico, subyace en ellos una estructura común, que podríamos organizar en cuatro fases: La Construcción de un Banco de Ítems, Construcción del Test, Administración del Test y Actualización del Banco de Ítems.

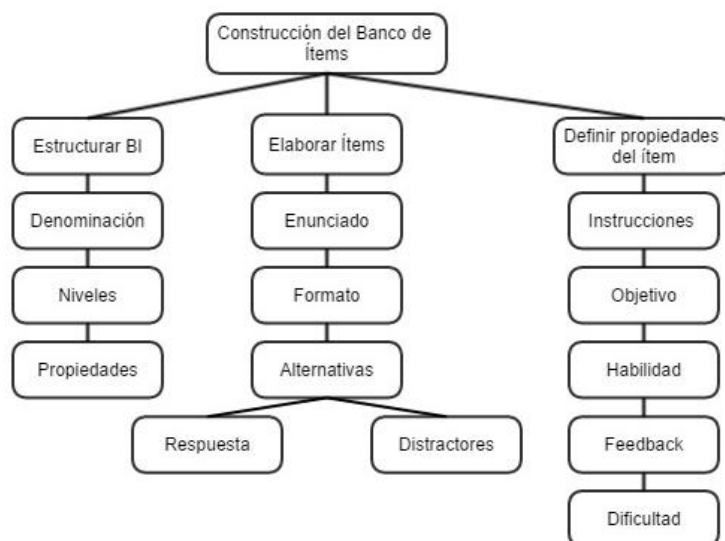
Diagrama 1: Flujo General Generadores de Test



Los Test Informatizados trabajan con bibliotecas o bancos de ítems conformado por un conjunto de preguntas preexistentes y/o generadas por el propio usuario o profesor, estructurados con base a uno o más patrones comunes, los cuales son almacenados y clasificados en función de las propiedades psicométricas y de contenido a las que hacen referencia. Estos bancos se pueden organizar en diversos niveles, propósitos, contenidos y formatos de preguntas, cualidad que los hace tremendamente funcionales para la labor pedagógica en las diversas áreas del conocimiento. Se pueden vincular tanto a respuestas como a preguntas diversos tipos de enlaces y formatos (enlaces web, archivos html, gifs, jpg, vídeos, etc.). Posibilita además la generación de base de ítems, incrustación de ecuaciones, gráficos y notación científica. Los test se pueden personalizar en cuanto a tipo de letra, cuerpo, color, etc.

Las ítems o preguntas se formulan en sus formatos clásicos (Verdadero/Falso, alternativas, términos pareados, respuestas breves o largas), además de otros no convencionales como lo serían los de expresión manipulativa proporcionando, además de dinamismo e interactividad a la aplicación del test, un más amplio ramillete de opciones de preguntas o ejercicios. Así por ejemplo colorear el área de un mapa, seleccionar una sección de un gráfico, despejar una fórmula, seguir con el mouse el recorrido de un trayecto o laberinto predeterminado; reconocer, mover o modificar un objeto; realizar correcciones gramaticales; reconocer, emitir sonidos o grabar verbalmente una determinada respuesta, entre otras. Ítems que al estar almacenados, parametrizados y clasificados con base a una estructura de propiedades conocidas por el examinador se pueden seleccionar, reutilizar y adaptar en función del tipo de evaluación requerida en cada momento.

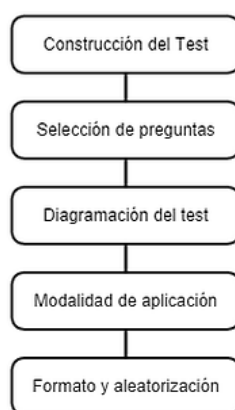
Diagrama 2: Proceso de Construcción del Banco de Ítems



Es precisamente esta polivalencia lo que permite su aplicación para la medición de una amplia gama de rasgos, conductas y competencias (Drasgow, 1999; Williamson, Mislevy, y Bejar, 2006), así como para la estimación de los análisis de procesos, estrategias o estructuras de conocimiento implicada en la resolución de los ítems por parte del examinado (Bejar, 2002), nivel de dominio en cuanto a razonamiento

matemático, analítico, visualización, analogías verbales, motricidad, entre otros. Misma funcionalidad que posibilita su uso en diversas áreas de la enseñanza, la formación profesional y la selección de personal; adaptándose con facilidad a modelos educativos no tradicionales, así como a diferentes teorías psicológicas del procesamiento de la información, haciendo surgir Test de diversa naturaleza. Los Test Adaptativos Informatizados (TAIs), puestos a prueba primariamente por Weiss (1974), consisten en la selección de forma automática y en dificultad dinámica de los ítems más apropiados para cada sujeto. Éstos son seleccionados por medio de un algoritmo aplicado por un ordenador según el nivel que progresivamente va manifestando en el transcurso del desarrollo de la prueba el examinado, con la sustancial reducción en los tiempos de aplicación que demandan los test no adaptativos (Olea y Ponsoda, 2003). En la actualidad, el Test of English as a Foreign Language (TOEFL), el Graduate Record Exam (GRE) o el Armed Services Vocational Aptitude Battery (ASVAB) disponen de versiones de esta modalidad (Drasgow, 1999; Rojas, 2001). Así también encontramos los Test basados en Modelos que son guiados por una particular teoría psicológica del aprendizaje y procesamiento de la información; y los Test Ipsativos, ampliamente utilizados en los test de personalidad y selección de personal que al presentar ítems con igual nivel de dificultad social reduce significativamente el sesgo por falseamiento de respuesta.

Diagrama 3: Proceso de Construcción del Test



Los Test Informatizados propician, a través de la automatización, una optimización del proceso clásico de aplicación de un Test al permitir la selección y presentación de los ítems de uno o más bancos, corrección de las respuestas, asignación de puntaje, e interpretación de la puntuación obtenida de forma mucho más rápida que de manera convencional. Agilización que se materializa en los menores tiempos requeridos para confección, aplicación y corrección de los exámenes, reduciendo significativamente los costes asociados al proceso.

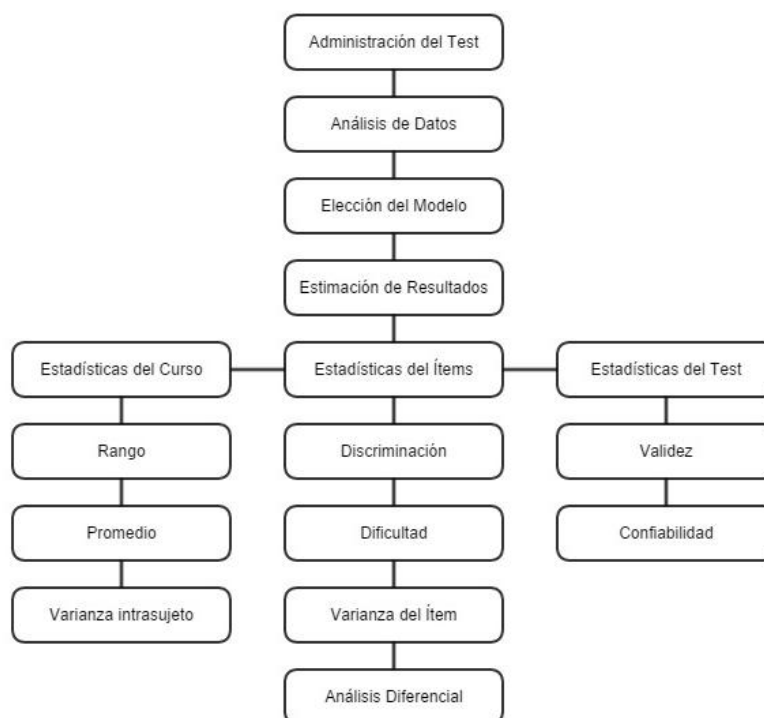
Los Test, una vez editados pueden aleatorizarse en cuanto al orden de las preguntas y/o alternativas consignando tantos formatos como alumnos sean. Esta aleatorización reduce la posibilidad de copia entre los alumnos pero facilita que las condiciones sean

similares para todos los evaluados, eliminando el sesgo por instrumentación. Modalidad de aplicación que estandariza más eficientemente las condiciones de control interno requeridas para dar validez a los resultados evaluativos, al homogeneizar las instrucciones, controlar el tiempo de aplicación, así como reducción de sesgos asociados a la copia, memorización, transmisión de contenidos, error de corrección, entre otros.

Así la evaluación se hace más confiable en sus puntuaciones, cuya aplicación puede operar sin la supervisión continua del profesor o la profesora, incluso ser utilizada tanto como proceso de auto-evaluación, así como para certificar los niveles comparativos de logro académico de un conocimiento o aptitud determinada.

Dada su funcionalidad en el test se pueden establecer tiempos de ejecución para la realización del examen (momento de acceso y tiempo máximo de ejecución), incluso un encriptado de seguridad en los exámenes del profesorado y de los estudiantes. Junto con gestionar desde el servidor el tiempo que tendrán los examinados para responder y la aleatorización de preguntas, se puede establecer el número máximo y mínimo de respuestas, e inclusive la corrección automática que posibilita al estudiante o examinador saber de forma inmediata cuál es la clasificación obtenida proporcionando el puntaje total, por ítem y por respuesta. Además puede retroalimentar al alumnado sobre cada pregunta y la razón por la cual la alternativa elegida por éste no es correcta, incluso poder intentarlo en más de una ocasión.

Diagrama 4: Proceso de Administración del Test



Los Test Informatizados brindan también la posibilidad de puntuar negativamente las respuestas incorrectas para penalizar la adivinación y corregir el efecto azar, asignar valoraciones diferenciadas tanto al ítem como a la respuesta, anular preguntas y/o

respuestas que pudieran resultar extremadamente fáciles, difíciles o que presentaran problemas en su formulación por ambigüedad o inducir a respuestas erróneas.

En su aplicación, además de los aciertos y errores, posibilita el registro de otro tipo de atributos del evaluado tales como los tiempos de reacción a un estímulo, agudeza visual, capacidad auditiva, coordinación motriz o destreza. Es precisamente esta polivalencia en la configuración del perfil psicotécnico del evaluado lo que determina su alta pertinencia, pues su funcionalidad y eficiencia opera con independencia relativa al componente cognitivo, procedimental o actitudinal a ser medido.

Con esta información se generan reportes automáticos, expresados en información cuantitativa, verbal, tabular o gráfica de forma inmediata, que posicionan al evaluado en un baremo específico respecto del grupo de referencia, además de gráficos y tablas con variada información (pregunta/respuesta + acertada/+fallada, mejor/peor alumno, alumno más próximo al aprobado, % de aprobados, etc.). Estadísticas referentes al nivel de aprovechamiento del curso y del alumno evaluado útiles tanto para la comunicación de resultados, como para la toma de decisiones respecto del proceso educativo con diferenciados niveles de desagregación. Estos índices resumen además de ser exportables a múltiples formatos y comunicados a variadas audiencias (alumnos, profesores, apoderados, directivos) posibilitan el estudio del comportamiento del mismo ítem, permitiendo la actualización de los baremos y parámetros del ítem a fin de conservar la vigencia de la métrica del Test.

Diagrama 5: Proceso de Actualización del Banco de Ítems



En un proceso cíclico, cada aplicación del test se constituye en una prueba piloto del reactivo que nutre de información al sistema para la correcta parametrización de los ítems contenidos en él. La información proveniente de aplicaciones sucesivas posibilita la recalibración, reformulación y mejora continua de la calidad de los bancos de ítems mediante un proceso de actualización rápido y amigable a nivel de usuario. En síntesis, los test informatizados no se constituyen únicamente en un repositorio estático de ítems de fácil acceso al momento de elaborar los test, sino que además en una herramienta que con escasos requerimientos brinda mayor seguridad a la acción evaluativa de la labor pedagógica.

A nivel de exigencias de hardware los Generadores de Test son programas livianos y versátiles que requieren mínimas exigencias en términos de procesador y memoria. Funcionan bajo plataformas de Sistemas Operativos Windows (2000 o superior) o

Macintosh, y aunque el tamaño varía dependiendo del número de ítems y test, su instalación precisa de escaso espacio disponible en disco.

Bajo la modalidad web funcionan con diversos navegadores (Chrome, Internet Explorer, Firefox) y si bien hay generadores que posibilitan el diseño de Test en formato impreso y web, algunos programas precisan de un navegador web para utilizar y acceder al menú de ayuda o soporte.

Hoy en día existe una amplia gama de programas generadores de pruebas informatizadas, que operan en línea y/o residentes en PC y permiten crear de forma sencilla distintos tipos de pruebas e ítems, aplicables en formato impreso (tradicional) o en modalidad web, siendo accesibles desde cualquier ordenador con conexión a Internet para su diseño y aplicación. Algunos de ellos incluso son de distribución gratuita, de uso no comercial, restringido al ámbito educativo, cuyas funcionalidades no se diferencian significativamente de las aplicaciones de pago o licenciadas.

Actualmente no se dispone de una metodología y/o evaluación consistente para estos Generadores de Test que posibilite a nivel de diseño metodológico evaluativo dimensionar los atributos o categorías de análisis más relevantes a tomar en cuenta, y menos una métrica o ranking referente a estos programas. A modo de ejemplo, principalmente por su presencia en la Internet dentro de los más utilizados y conocidos se pueden mencionar en los siguientes:

Quiz Press: Herramienta que permite crear test de examen en formato HTML. Admite seis tipos distintos de preguntas: respuesta múltiple, completar oraciones, respuesta corta y larga, verdadero o falso y términos pareados. Los test se realizan en modalidad web. Incluye efectos dinámicos además de un informe final que repasa los aciertos y los errores que ha cometido el usuario al contestar la prueba siguiendo las pautas del autor del test.

iTest: Herramienta especialmente diseñada para el profesorado que facilita la tarea de crear y corregir exámenes. Consta de dos herramientas: El cliente, parte que ve el alumno y en la que debe responder a las preguntas del examinador y el servidor, parte donde se almacenan las preguntas y respuestas.

Hot Potatoes: Conjunto de seis herramientas para elaborar ejercicios interactivos en modalidad Web. La información de cada tipo de ejercicio se guarda en un archivo específico. A partir del mismo se generará el documento final interactivo en formato HTML. El alumno no necesita tener instalado en su equipo el programa para realizar cada ejercicio. Dándose de alta en la página Web de Hot Potatoes el acceso es gratuito y sin restricciones.

Wondershare QuizCreator: Generador de Test con ítems de diversos formatos (elección múltiple, selección de imágenes, parejas de conceptos, ordenar conceptos para demostrar el entendimiento de una secuencia, etc. El usuario sólo ingresa las preguntas e indica cuál es la respuesta correcta y la aplicación crea todos los aspectos gráficos y de programación.

TestGIP Profesor: Genera y edita exámenes multimedia tipo test de respuestas múltiples. Incluye la posibilidad de vincular imágenes, vídeos, sonidos y objetos OLE/HTM tanto a las preguntas como a las respuestas. Posibilita su etiquetado, ordenación, aleatorización, penalización y corrección de la adivinación. Desarrollado por la Universidad Politécnica de Valencia.

TestGen 7.0: Es una aplicación de distribución restringida de propiedad de la editorial Pearson generadora de test y bancos de ítems, muy funcional, versátil y amigable. Atributos que lo han hecho muy utilizable en centros de estudios de los Estados Unidos. Consta básicamente de dos ventanas. La primera dispone del conjunto de archivos, bancos de ítems disponibles, la segunda posibilita editar o crear un banco o un test incorporando preguntas y alternativas según sea el caso. Disponible en <http://www.pearsoned.com/testgen>

ExamView: Software de la editorial MacGraw-Hill que permite crear pruebas impresas o en modalidad Web. Incluye un corrector ortográfico, editor de ecuaciones, pruebas en línea / guías de estudio, bancos de preguntas, función de traducción.

Conclusiones

La evaluación es un proceso relevante para constatar los niveles de aprovechamiento educativo que alcanzan los alumnos, determinar la eficiencia de la acción docente y certificar la calidad de los procesos formativos de las instituciones, en especial en contextos tan dinámicos, complejos y abiertos como el sistema educativo chileno.

Los modelos educativos centrados en el alumno, que ponen el acento en el aprendizaje, precisan de instrumentos y herramientas de medición evaluativas adecuadas y eficientes para conocer más fielmente los niveles de logro de los estudiantes y propiciar acciones remediales y ajustes curriculares basados en evidencia, relevando así el componente ético de garantizar aprendizajes y hacer más justa la calificación resultante de la acción evaluativa.

Evaluar es una tarea importante pero compleja y no exenta de dificultades, donde la posibilidad de elaborar y aplicar una medición confiable y válida puede resultar engorrosa y desbordante para el profesor, en especial cuando aumenta el número de evaluaciones, cursos y estudiantes.

En la actualidad la informática ha propiciado grandes cambios en las estrategias evaluativas, principalmente en la forma de construir, validar y aplicar tests psicológicos y educativos, así como en la forma de analizar los resultados que de sus aplicaciones se generan, por lo que se configuran en un eficiente y poderoso aporte que las TIC proporcionan a la acción docente.

Los Test Informatizados posibilitan la creación de bancos de ítems estructurados, estandarizados, calibrados y validados para la confección, en forma rápida y sencilla, de evaluaciones de diversa naturaleza y formato. No tan solo facilitan la acción evaluativa

a través de la actualización permanente de bancos de preguntas, seleccionando las más pertinentes en cada ocasión, sino que además simplifica, haciendo más amigable a nivel de usuario, engorrosas y complejas aplicaciones estadísticas y sistemas de información; espacio en el cual el profesor no se percibe adecuadamente preparado o bien no dispone del tiempo suficiente que la naturaleza de la tarea demanda.

Realizar procesos de apropiación de este tipo de tecnologías no tan solo es necesario para la gestión de la actividad docente, sino que además puede tributar fuertemente al mejoramiento de los niveles de eficiencia, calidad y logro del proceso educativo, en la convicción que no hay procesos de constatación de la eficacia de la acción pedagógica sin la adecuada y objetiva medición de los niveles de logro de los estudiantes.

Fecha de finalización de redacción del artículo: 4 de enero de 2014.-

Rodríguez Garcés, C. Muñoz Sepúlveda, M. Castillo Riquelme, V. (2014). Test Informatizados y su Contribución a la Acción Evaluativa en Educación. *RED, Revista de Educación a Distancia. Número 43*. 15 de Noviembre de 2014. Consultado el (dd/mm/aaaa) en <http://www.um.es/ead/red/43/>

Bibliografía

- Beeby, C. E. (1977). The meaning of evaluation. *Current Issues in Education Evaluation. Department of Education, Wellington*, N° 4. 68-78.
- Bejar, I. (2002). Generative testing: From conception to implementation. En S. Irvine, & P. (. Kyllonen, *Item generation for test development* (págs. 199-217). Mahwah, NJ: Lawrence Erlbaum Associates.
- Bellei, C. (2013). Estudio de la segregación socioeconómica y académica de la educación chilena. *Estudios Pedagógicos*, 39(1), 325-345.
- Beyer, H., Alvarado, J., Arellano, J., Aylwin, M., Brunner, J. J., Krebs, A., y otros. (2010). *Propuesta para fortalecer la profesión docente en el sistema escolar chileno*. Santiago de Chile: MINEDUC.
- Camilloni, A. (2005). La calidad de los programas de evaluación y de los instrumentos que los integran. En A. e. Camilloni, *La evaluación de los aprendizajes en el debate didáctico contemporáneo* (págs. 9-30). Buenos Aires: : Paidós Educador.
- Celman, S. (2005). ¿Es posible mejorar la evaluación y trasformarla en una herramienta de conocimiento? En A. e. Camilloni, *La evaluación de los aprendizajes en el debate didáctico contemporáneo* (págs. 35-66). Buenos Aires: Paidós Educador.
- Contreras, D., & Gallegos, S. (2011). Desigualdad salarial en América Latina: una década de cambios. *Revista CEPAL*(103), 27-45.

- Contreras, D., Gallegos, S., & Meneses, F. (2009.). *Determinantes de desempeño universitario: ¿Importa la habilidad relativa? Análisis ex - ante y ex – post de una política pública*. Santiago: Investigación preparada para el Consejo de Educación Superior.
- Cronbach, U. (1963). Course improvement through evaluation. *TEach.Coll.Rec.*, 64: 762-83.
- Darling-Hammond, L. (2001). *El derecho de aprender. Crear buenas escuelas para todos*. Barcelona: Ariel Educación.
- Davey, T. (2005). Computer-based testing. En B. Everitt, & D. (. Howell, *Encyclopedia of statistics in behavioral science*. Hoboken, NJ: Wiley.
- De los Ríos, D., Valenzuela, J. P., Bellei, C., & Sevilla, A. (2009). *Datos cuantitativos sobre la profesión docente en Chile*. Santiago de Chile: Centro de investigación avanzada en educación, Universidad de Chile.
- Díaz, E., Himmel, E., & Maltes, S. (1990). Evolución histórica del sistema de selección a las universidades chilenas, 1967-1989. En M. J. Lemaitre, *La educación superior en Chile: un sistema de transición*. Santiago: Corporación de Promoción Universitaria (CPU), Colección Foro de la Educación Superior.
- Donoso, S., & Schiefelbein, E. (2007). Análisis de los modelos explicativos de retención de estudiantes en la universidad: Una visión desde la desigualdad social. *Estudios Pedagógicos*, N° 33, Pags. 7-27.
- Dragow, F. y.-B. (1999). *Innovations in computerized assessment*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Eisner, E. (1998). *El ojo ilustrado. Indagación y mejora de la práctica docente*. Barcelona: Editorial Paidós.
- Epple, D., & Romano, R. (1998). Competition between private and public schools, vouchers and peer group effects. *American Economic Review*, 88(1), 33-62.
- Gómez, M. Á., Rodríguez, G., & Ibarra, M. S. (2013). Desarrollo de las competencias básicas de los estudiantes de educación superior mediante la e-evaluación orientada al aprendizaje. *Relieve*, (19)(1).
- Hsieh, C.-T., & Urquiola, M. (2006). The effects of generalized school choice on achievement and stratification: evidence from Chile's voucher program. *Journal of Public Economics*, 90, 1477-1503.
- Litwin, E. (2005). La evaluación: campo de controversias y paradojas o un nuevo lugar para la buena enseñanza. En A. e. Camilloni, *La evaluación de los aprendizajes en el debate didáctico contemporáneo* (págs. 11-34). Buenos Aires: Editorial Paidós.

- Mayol, A., Araya, J., & Azócar, C. (2011). Desigualdad y educación: la pertinencia de políticas educacionales que promuevan un sistema público. *Docencia*, 24-33.
- McEwan, P., & Carnoy, M. (2000). The Effectiveness and Efficiency of Private Schools in Chile's Voucher System. *Educational Evaluation and Policy Analysis*, 22(3), 213-239.
- Morales, P. (2012). *Análisis de ítems en las pruebas objetivas*. Madrid: Universidad Pontificia Comillas.
- Muñiz, J. (2000). *Teoría Clásica de los Tests*. Madrid: Pirámide.
- Muñiz, J. (2010). Las teorías de los tests: Teoría Clásica y Teoría de Respuesta a los Ítems. *Papeles del Psicólogo*, N° 31 pag: 57-66.
- Nunnally, J., & Bernstein, I. (1995). *Teoría Psicométrica*. Madrid: McGraw-Hill.
- Olea, J., & Ponsoda, V. (2003). *Tests adaptativos informatizados*. Madrid: UNED, Colección Aula Abierta.
- Olea, J., Abad, F., & Barrado, J. (2010). Tests informatizados y otros nuevos tipos de tests. *Papeles del psicólogo: revista del Colegio Oficial de Psicólogos*, Vol. 31, N° 1, 97-107.
- Olea, J., Ponsoda, V., & Prieto, G. (. (1998). *Test informatizados. Fundamentos y aplicaciones*. Madrid: Pirámide.
- Ortúzar, S., Flores, C., Milesi, C., & Cox, C. (2009). *Aspectos de la formación inicial docente y su influencia en el rendimiento académico de los alumnos*. Santiago de Chile: Instituto de Sociología y Centro de Estudio de Políticas y Prácticas en Educación, Pontificia Universidad Católica de Chile.
- Rodríguez, C., & Castillo, V. (2014). Calidad en la formación inicial docente: los déficits de las competencias pedagógicas y disciplinares en Chile. *Actualidades Investigativas en Educación*.
- Rodríguez, C., & Jarpa, C. G. (2014). Segmentación de la calidad evaluativa: Análisis del rendimiento escolar y su relación con el nivel de logro en pruebas de selección universitaria en Chile. *En prensa*.
- Rojas, A. (2001). Pasado, presente y futuro de los tests adaptativos informatizados: entrevista con Issac Bejar. *Psicothema*, 13, 685-690.
- Romero, C. y. (2006). Herramienta Autor para la Gestión de Tests Informatizados dentro del Sistema AHA! *Revista Iberoamericana de Informática Educativa*, N° 03; 43-54.

- Rosas, R., & Santa Cruz, C. (2013). *Dime en qué colegio estudiaste y te diré que CI tienes: Radiografía al desigual acceso al capital cognitivo en Chile*. Santiago de Chile: Ediciones UC.
- Seels, B., & Richey, R. (1994). *Instructional technology: The definition and domains of the field*. Washington, DC: Association for Educational Communications and Technology.
- Stiggins, R. (2004). New Assessment Beliefs for a New School Mission. *Phi Delta Kappan* 86, 22-27.
- Thorndike, R. L. (1989). *Psicometría Aplicada*. México: Limusa.
- Valenzuela, J. P. (2006). *Evolución de la segregación socioeconómica de los estudiantes chilenos y su relación con el financiamiento compartido*. Santiago de Chile: Fondo de Investigación y Desarrollo en Educación - Ministerio de Educación .
- Weiss, D. (1974). *Strategies of adaptive ability measurement*. University of Minnesota: Research report 74-5. Dep. Of Psychology.
- Williamson, D., Mislevy, R., & Bejar, I. (2006). *Automated Scoring of Complex Tasks in Computer Based Testing*. Mahwah, NJ: Lawrence Erlbaum Associates.