APORTE TECNOLÓGICO A LA EVALUACIÓN: SOFTWARE PARA EL ANÁLISIS Y REPORTE DE RESULTADOS DE EXÁMENES

Cristian Ernesto Castañeda Sánchez
cristiancastaneda@xochicalco.edu.mx

Luis Ángel Contreras Niño
angel@uabc.edu.mx

Palabras clave: Software, Exámenes, Informes de resultados

Resumen

El propósito de la evaluación del aprendizaje es obtener información sobre los logros alcanzados en este aspecto de manera que puedan ser sometidos a juicio y sirvan como un elemento más para la toma de decisiones correspondientes. Al reportar la información que se produce a partir del análisis de los resultados de los exámenes para medir el aprendizaje se pretende cumplir con este propósito cuyas decisiones son tan diversas como los destinatarios de tales reportes. El presente documento tiene como objetivo dar a conocer el desarrollo y funcionamiento de un software libre y gratuito en una primera versión que analice los resultados de exámenes y que los integre en reportes para los destinatarios con la información pertinente según las necesidades que tienen a este respecto. El proceso se llevó a cabo en cinco etapas y con ello se pretende, entre otras cosas, reducir el empleo de recursos y tiempo en la elaboración de informes de resultados.

Introducción

La información que se obtiene tras la aplicación de exámenes para evaluar el aprendizaje, bien sea que esté interpretada con referencia a un criterio o a una norma, es empleada es utilizada con dos objetivos; uno, a menor escala, busca brindar información a los actores educativos más cercanos a las aulas, como son profesores, alumnos, directivos o padres de familia, sobre el logro académico y los aspectos que se relacionan con él, y de esta manera corregir o mejorar las prácticas; otro objetivo aportar elementos que apoye la toma de decisiones en los niveles medios y altos del sistema educativo (Ravela, 2002). Asimismo, Contreras (2000) y Sánchez (2007) resaltaron el uso de los resultados de las exámenes con fines de evaluación diagnóstica y formativa del aprendizaje de los alumnos, así como la acreditación y promoción de niveles escolares, o bien, como otro insumo al momento de evaluar programas educativos normales y avanzados. Condelli (2005) también afirmó que con los resultados se da cuenta de los logros de los programas educativos, las mejoras que son necesarias y el contexto en el que se ha llevado a cabo el proceso de enseñanza.

Dicha información puede ser presentada mediante diferentes formatos, los cuales representan la evidencia final del proceso de evaluación; uno de los formatos más usados son los llamados reportes de

resultados. Ya que, como se ha mencionado, el impacto de los resultados tiene resonancia en las decisiones que toman tanto al interior de las instituciones y las aulas, en las prácticas de enseñanza, como en la planeación de la educación desde fuera de las instituciones, por ejemplo en la elaboración de políticas educativas (Martínez, 2012; Ravela, 2002), es necesario poner atención a esta parte del proceso de evaluación que es la elaboración de los reportes.

Condelli (2005) señaló que para la elaboración de los informes de resultados es necesario tener en cuenta algunos aspectos que a continuación se resumen: (a) Destinatarios a los que van dirigidos y el objetivo de informar a dichos destinatarios; (b) Elección de las métricas que representen adecuadamente el desempeño de los examinados; (c) Descripción de los niveles de logro que se establecieron previamente y los puntos de corte de cada nivel; (d) Formato de presentación del documento y (e) Publicación y/o entrega de los resultados con la finalidad de lograr el impacto deseado.

Por todo lo anterior, se propuso desarrollar una herramienta digital con la que se pueda analizar los resultados de exámenes de referencia criterial y además, con los datos analizados, se produjeran los reportes de resultados dirigidos a algunos de los distintos destinatarios y con ello agilizar esta serie de procedimientos a fin de brindar la información con mayor oportunidad. En la presente ponencia se explica el proceso de elaboración del software llamado PARRERC en su primera versión. Para apoyar el diseño del software, se utilizaron los datos existentes de una aplicación de un examen. Dichos datos correspondieron a la aplicación de la versión cuatro de un examen de español para el egreso de la educación primaria en Baja California (Contreras, 2000). Se tomaron los resultados de 117 estudiantes de primer grado de secundaria de cuatro instituciones rurales del municipio de Ensenada. La idea de generar el software ya había sido documentada en otros estudios (Contreras, 2009; Sánchez, 2007) como un recurso que agilizara la entrega de la información estadística de manera oportuna.

Desarrollo

De acuerdo con algunos autores (Bronson, 2007; Alonso, Martínez & Segovia, 2005; Kendall & Kendall, 2005; Álvarez& Arias, 2002) las etapas de proceso de desarrollo de software son: análisis, diseño, codificación, comprobación, mantenimiento y documentación. El desarrollo del PARRERC que aquí se refiere ha seguido este mismo modelo y a continuación se describen sus etapas:

A) Análisis: en esta etapa se precisaron los aspectos que se pretendía solventar o apoyar con la creación del programa. Para lo cual se definieron los destinatarios a los que se les pretende entregar un informe de resultados; asimismo, se llevó a cabo una búsqueda en diferentes fuentes (documentadas y de comunicación personal) sobre las necesidades informativas que tienen los diferentes tipos de

destinatarios, de acuerdo a las decisiones que toman en su desempeño dentro del proceso educativo. profesional. Luego, se identificó la información que pudiera ser analizada cuantitativamente en un programa de cómputo, al menos en su primera versión.

En el marco conceptual generado de esta consulta, se definieron dos categorías de destinatarios, y se definieron los objetivos que se persigue al entregarles un tipo determinado de resultados. Las categorías son las siguientes:

- Usuarios de la información de nivel de macro planeación educativa: en la que se encuentran los destinatarios cuyas decisiones afectan a todo un el sistema educativo como son los responsables de la planeación educativa de un sistema, las instituciones de capacitación docente o los elaboradores de políticas públicas en materia de educación.
- Usuarios de la información de nivel de micro planeación educativa, cuya labor es la toma de decisiones al interior de los planteles educativos, es decir, los directivos, docentes, inspectores o asesores pedagógicos y los propios estudiantes.

  Lo anterior sirvió de base para la selección del tipo de métricas más pertinentes para reportar los resultados, de manera que su interpretación fuera lo más completa, sencilla y objetiva posible.

B) Diseño: Una vez definida la información que se analizaría con el software, se diseñaron dos modelos:

- un modelo de procesamiento de datos numéricos: consistió en la recopilación y elaboración de un banco de fórmulas matemáticas y fórmulas en Excel con las que se pudieran analizar los resultados de un examen. Luego, se definieron, mediante una matriz de análisis, las fórmulas con las que se obtendría la información para cada destinatario, tal como se estableció en la etapa anterior.
- Un modelo de interfaz de usuario y archivos de salida: consistió en una serie de representaciones gráficas de las posibilidades que se ofrecerían en el software: barra de menú de opciones, ventanas de configuración y archivos de salida (output) con los gráficos, tablas, y cuadros de texto correspondientes a las opciones que se seleccionaran.

  Ambos modelos se crearon con la finalidad de clarificar las especificaciones establecidas en la etapa de análisis, así como de simular la funcionalidad lógica del sistema que se pretendía crear.

C) Codificación o programación: Esta etapa corresponde a la traducción del diseño en código fuente y luego en código objeto, de manera que las especificaciones puedan tener funcionalidad un programa de cómputo del sistema. En ella, los modelos sirvieron como insumo para el desarrollo del software en lenguaje de programación Java.

D) Comprobación: Una vez que se realizó la programación, se puso en operación el producto y se compararon los resultados que se obtuvieron con los que se obtuvieron con el programa Excel al momento de diseñar los insumos, a fin de cotejar la exactitud de los cálculos.

E) Mantenimiento y documentación: son etapas en las que actualmente se encuentra el programa ya que la primera corresponde a la verificación continua de su funcionamiento y la segunda a la descripción del proceso de desarrollo y funcionalidad.

Respecto a los resultados que pueden producirse con el software a partir de los datos disponibles son los siguientes:

- Información de calidad psicométrica del examen: con el programa es posible producir valores sobre el examen en general, como la confiabilidad de medición del instrumento, la dificultad y la discriminación global; asimismo, se obtienen valores sobre la discriminación, dificultad y funcionalidad de los distractores de cada uno de los reactivos, además se pueden generar gráficas las que se observa la relación que existe entre la puntuación total en el examen y los aciertos y errores que tuvieron en cada ítem los examinados, de manera que se puedan visualizar los parámetros del ítem. Lo anterior se denomina "análisis gráfico de ítems" y se produce al selección una opción que pone en ejecución un software añadido al que aquí se presenta llamado PAGI diseñado con este objetivo (De Los Santos, 2010).

  Como se mencionó en un principio, el software fue probado con los resultados de un examen aplicado a estudiantes de secundaria. Las respuestas al instrumento en fueron analizadas con el PARRERC y con el programa IBM SPSS v. 22 con el fin de comprobar los resultados. Los datos del análisis psicométricos fueron los siguientes:

  Tabla 1. Valores de propiedades psicométricas obtenidas para comparación

Propiedades obtenidas	IBM SPSS v. 22	PARRERC
Índice de confiabilidad (Alfa de Cronbach)	.7	.7
Índice de dificultad global	.54	.54
Índice de discriminación global	.31	.30

- Resultados del desempeño de los examinados a nivel general: es posible obtener los puntajes totales y porcentajes de aciertos de cada uno de los examinados, o bien, los promedios de estas métricas por conjuntos de diversos tamaños, como grupo escolar, el total de examinados de una escuela, de una zona escolar o municipios (ver figura 1).

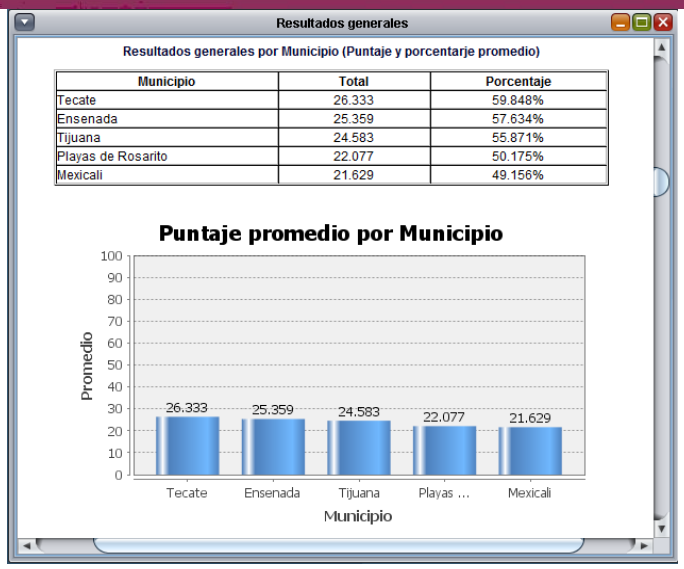

Figura 1. Resultados generales de los examinados agrupados por municipio.

- Resultados del desempeño de los examinados por unidad de aprendizaje o línea de formación: de igual forma que en el anterior tipo de resultados, se obtienen puntajes totales y porcentajes de aciertos en este caso desagregados por las unidades de aprendizaje o líneas de formación que fueron examinadas.

  En los dos tipos de análisis de resultados anteriormente descritos, es posible generar gráficas de la información analizada.

- Reportes de resultados dirigidos a diferentes destinatarios: los reportes de resultados son una parte importante del programa, en ellos se resumen los datos analizados y se presentan a algunos destinatarios. Es importante señalar que, al seleccionar las opciones que aparecen en "Reporte de resultados", se generan archivos de salida con un nivel de especificidad de la información distinto para cada tipo de destinatario. Sin embargo, esto no quiere decir que el destinatario deba recibir solo esa información, sino que a juicio del evaluador y de acuerdo con las funciones que ejerce se pueden entregar reportes con datos más específicos, por ejemplo, al asesor técnico pedagógico de una escuela se le pueden entregar el mismo reporte que se le entregan a un profesor de un grupo determinado. Los destinatarios a los que es posible reportar resultados del examen con el programa y el tipo de información que se les entrega se describe en la tabla 2.

Tabla 1. Destinatarios e información reportada (2)

Destinatario	Tipo de información
Alumnos y Padres de familia	Porcentaje de aciertos en el examen

	Puntaje total y porcentaje de aciertos en cada una de las unidades de aprendizaje
	Puntaje promedio de aciertos en el examen y en cada unidad de aprendizaje del grupo al que pertenece el examinado para fines de comparación
Docentes	Puntaje y porcentaje de aciertos general y de las unidades de aprendizaje de cada examinado y promedio global del propio grupo
	Puntaje y porcentaje promedio (general y de unidad de aprendizaje) de los demás grupos examinados del mismo grado y que pertenecen a la misma escuela con fines de comparación

Tabla 1. Destinatarios e información reportada (continuación)

Directivos del plantel	Puntaje y porcentaje promedio (general y de unidad de aprendizaje) de los grupos examinados de la propia escuela
	Puntaje y porcentaje promedio (general y de unidad de aprendizaje) de las escuelas que pertenecen a la misma zona escolar
Inspectores de zona – Asesores Técnico Pedagógicos(3)	Puntaje y porcentaje promedio (general y de unidad de aprendizaje) de las escuelas de la zona escolar en la que ejercen su labor y/o del municipio
Responsables de planeación educativa a nivel macro y secretario de educación	Puntaje y porcentaje promedio (general y de unidad de aprendizaje) de las zonas escolares y municipios evaluados

Toda esta información se concentra en archivos de salida y se presenta en tablas de contingencia, gráficas y cuadros de texto con interpretación de los datos numéricos. En la figura 2 se muestra uno de los archivos de salida.

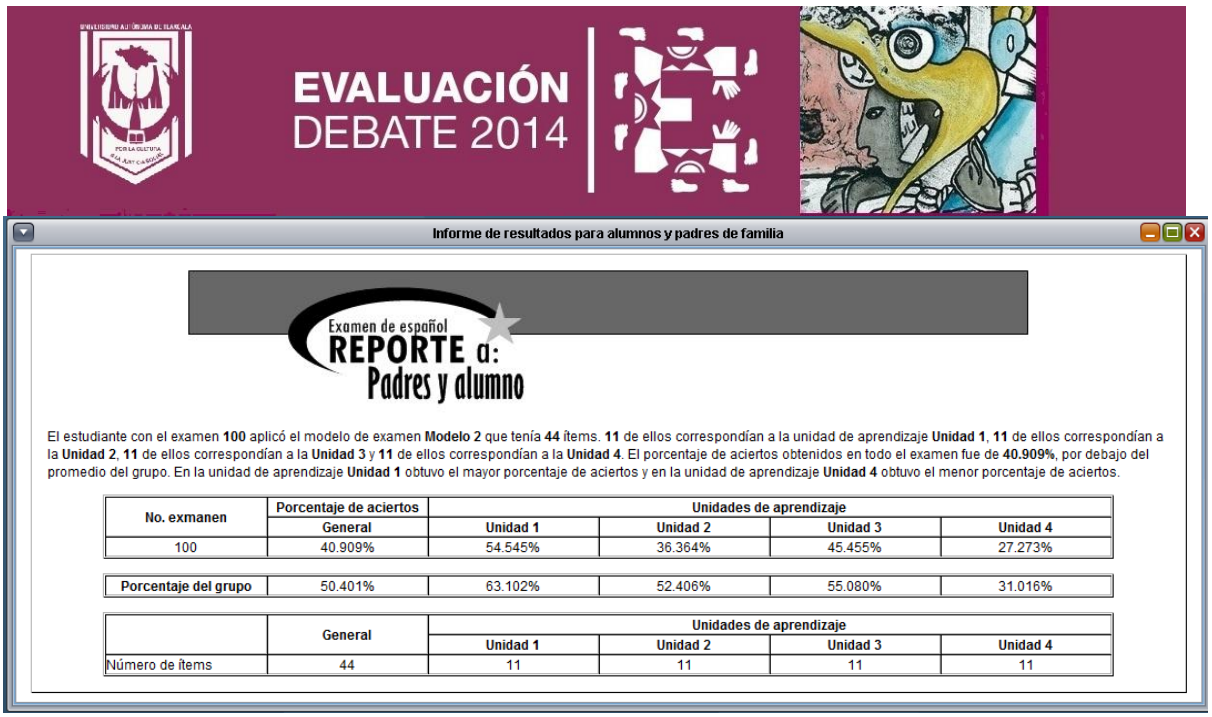

Figura 2. Archivo de salida con el reporte de resultados dirigido a padres de familia y alumnos.

Conclusiones

La evaluación, más allá de brindar simplemente información sobre los procesos educativos, tiene la encomienda de proporciona elementos para la reflexión a una comunidad a partir de la observación de la realidad en un área de interés específica. (Tiana, 1997). Al producir este software se intenta contribuir al desarrollo de la cultura evaluadora que entre sus fines está que los agentes educativos y, por lo tanto, usuarios de la evaluación, reconozcan los resultados, en este caso, de exámenes, como insumo útil para la toma de decisiones, lo cual supone, entre otras cosas, su interpretación oportuna (Himmel, 2010). Por este motivo, se pensó desde un principio que el PARRERC pueda ser usado por todo agente educativo interesado en la aplicación y obtención de resultados mediante exámenes de respuesta cerrada. Para ello, se requiere un entrenamiento sencillo en el uso del software, así como para la interpretación de algunos de los valores que se obtienen.

Con este software, no solo se obtiene información del desempeño de los examinados, de manera individual y en conjuntos de forma casi instantánea, lo cual ya es un apoyo considerable a diferentes agentes, sino que también es posible calcular valores que indiquen las propiedades del instrumento como la dificultad, la discriminación y la confiabilidad y se evalúe el contenido preciso que se quiere.

Sobre el aspecto de la calidad psicométrica que se adquiere con el software, es importante señalar que este programa es potencialmente un soporte valioso en lo que respecta a la generación de evaluaciones con las que recabe información que les permita a los profesores en servicio de todos los niveles, determinar con mayor precisión los logros de aprendizaje de sus estudiantes, ya que ellos mismos pueden desarrollar

sus propios exámenes y ajustarlos de manera que operen lo más cercano a la objetividad que pueda obtenerse con los cálculos estadísticos adecuados.

Con respecto al tipo de interpretaciones que se hacen de los resultados y que se reportan con el software, la idea original fue que dichas interpretaciones se realizaran siempre referidas a un criterio. Las métricas que se emplean para reportar el desempeño en el examen son puntajes totales y porcentajes de aciertos, los cuales ya son por sí mismas métricas que se pueden interpretar referidas a un criterio, puesto que es posible identificar el nivel de dominio de los examinados en la prueba (Jornet & González, 2009). No obstante, como ya se describió, en los reportes, se decidió que los reportes contengan también promedios de diversos conjuntos –lo cual conduce a una interpretación referida a una norma– ya que de esta manera, el destinatario puede tener más elementos útiles para el ejercicio de toma de decisiones que deba hacer.

Un aspecto que queda pendiente es que el software no se limita para el uso de educación de nivel básico, sin embargo, es necesario establecer nombres genéricos a las opciones, de manera que no cause confusión entre los usuarios de otros niveles.

Otro aspecto pendiente es una valoración exhaustiva de los propios reportes que se producen con el software por parte de más agentes educativos, de manera que sus aportes sirvan para la mejora del programa y se proyecte una nueva versión del mismo para beneficio de la comunidad educativa.

Referencias

Álvarez, J. & Arias, M. (2002). Análisis, diseño y mantenimiento del software. Madrid: UNED. Recuperado de: http://www.ia.uned.es/ia/asignaturas/adms/GuiaDidADMS/

Alonso, F., Martínez, L. & Segovia, F. (2005). Introducción a la ingeniería del software: modelos de desarrollo de programas. Madrid: Delta Publicaciones.

Bronson, G. (2007). C++ para Ingeniería y Ciencias. (2ª ed.). México: Thompson.

Condelli, L. (2005). Demonstrating results: developing state and local report cards for adult education. Washington: American Institutes for Research.

Contreras, L. A. (2000). Desarrollo y pilotaje de un examen de español para educación primaria del estado de Baja California. Tesis de Maestría. Universidad Autónoma de Baja California. Ensenada.

Contreras, L. A. (2009). Desarrollo y validación de un modelo para elaborar exámenes criteriales de gran escala alineados con el currículum. (Tesis inédita de doctorado). Universidad Autónoma del Estado de Morelos. Cuernavaca.

De los Santos, G. (2010). *Desarrollo, operación y evaluación de un módulo para capacitar a docentes en servicio para que mejoren sus exámenes de opción múltiple mediante el análisis gráfico de ítems*. (Tesis inédita de maestría). Universidad Autónoma de Baja California. Ensenada.

Himmel, E. (2010) Hacia una cultura de evaluación educacional. México. Recuperado de: http://www.ifie.edu.mx/Erika%20Himmel.htm.

Jornet, J. & González, J. (2009). Evaluación criterial: determinación de estándares de interpretación (EE) para pruebas de rendimiento educativo. Estudios sobre Educación, 16, 103 – 123.

Kendall, K. & Kendall, J. (2005). Análisis y diseño de sistemas. México: Pearson.

Martínez, F. (2013). El futuro de la Evaluación Educativa. *Sinética,* 40. Recuperado de: http://www.sinectica.iteso.mx/assets/files/articulos/40_el_futuro_de_la_evaluacion_educativa.pdf

Ravela, P. (2002). ¿Cómo Presentan sus Resultados los Sistemas Nacionales de Evaluación Educativa en América Latina? Programa de promoción de la reforma educativa en América Latina y el Caribe. Recuperado de: http://www.oei.es/calidad2/ravela.pdf

Sánchez, C. (2007). Desarrollo de un módulo de capacitación y asesoría para apoyar el análisis, reporte e interpretación de resultados derivados de la aplicación de pruebas criteriales de gran escala. (Tesis inédita de maestría). Universidad Autónoma de Baja California. Ensenada.

Tiana, A. (1997). Tratamiento y uso de la información en evaluación. Calidad y equidad en la educación. Recuperado de: http://www.campus-oei.org/calidad/tiana.htm