

# Sistema de Evaluación Inteligente para Medir Habilidades de Razonamiento Matemático

## Intelligent Evaluation System to Measure Mathematical Reasoning Skills

Roberto Contreras-Juárez \*  
Etelvina Archundia-Sierra  
Héctor D. Ramírez-Hernández  
Nelva B. Espinoza-Hernández  
V. Judith Hernández-Moyotl

Benemérita Universidad Autónoma de Puebla, México

En el presente trabajo de investigación se proporcionan las bases que sustentan el diseño, desarrollo e implementación de un sistema de evaluación inteligente sobre los fundamentos de un Test Adaptativo Informatizado. El desarrollo de este prototipo de prueba adaptativa permitirá medir habilidades de razonamiento matemático, incorporando al resultado final, un perfil cognitivo que puede ser empleado para dar seguimiento del desempeño académico de los estudiantes a lo largo de su estadía en la institución de educación superior. En la parte teórica se proporcionan los principios básicos de una prueba adaptativa, la psicometría de los reactivos, la construcción de los niveles de desempeño a partir de procesos cognitivos basados en la Taxonomía de Bloom, así como la descripción y diseño de un algoritmo inteligente para la selección de reactivos. Para la parte experimental se utilizaron reactivos calibrados y probados en pruebas escritas aplicadas a una muestra de más de 10.000 estudiantes. Además, se explica el desarrollo del sistema de evaluación inteligente utilizando una arquitectura cliente servidor de tres niveles, implementado en el lenguaje de programación java, utilizando MySQL como sistema gestor de base de datos y HTML para la interfaz con el usuario y se proporcionan algunos de los resultados obtenidos.

**Palabras Clave:** Procesos cognitivos; Test adaptativo informatizado; Algoritmo de evaluación inteligente; Algoritmo adaptativo.

This research work provides the basis for the design, development and implementation of an intelligent evaluation system on the bases of a Computerized Adaptive Test. The development of this adaptive test prototype will allow the measurement of mathematical reasoning skills, incorporating to the final result, a cognitive profile that can be used to monitor the academic performance of students throughout their stay at the institution of higher education. The theoretical part provides the basic principles of an adaptive test, the psychometry of the items including their difficulty, discrimination and predictability; the construction of performance levels from the cognitive processes based on Bloom's Taxonomy, as well as the description and design of an intelligent algorithm for the selection of items. For the experimental part, items calibrated and tested in written tests applied to a sample of more than 10,000 students were used. In addition, the development of the intelligent evaluation system using a three-level client-server architecture, implemented in the java programming language, using MySQL as a database management system and HTML for the user interface is explained and finally, some of the results obtained when implementing the system in a group of students are provided.

**Keywords:** Cognitive processes; Adaptive computerized test; Algorithm of intelligent evaluation; Adaptive algorithm.

---

\*Contacto: roberto.contreras@correo.buap.mx

## 1. Introducción

### 1.1. Test Adaptativo Informatizado

En el campo educativo, la evaluación es el medio utilizado por los docentes para estimar las aptitudes y conocimientos de los alumnos. Generalmente la evaluación está asociada a una prueba escrita que contiene ejercicios sobre temas desarrollados en clase. Como es sabido, el proceso de construcción y selección de los ejercicios depende directamente del docente pues se basan en su criterio, provocando que la prueba no se adapte al estudiante y en muchos casos, se termine ignorando el nivel de habilidades de los alumnos.

En las últimas décadas, los avances computacionales han generado un importante cambio de enfoque, estrategias y objetivos en el campo de la evaluación educativa (Gil et al., 2000). En este ámbito, los test informatizados son instrumentos de evaluación que se responden en un ordenador para estimar el nivel de rendimiento de un estudiante (Hontangas Beltran, 2000). Sin embargo, este tipo de instrumento evalúa a todos los estudiantes con el mismo conjunto de reactivos, sin la posibilidad de presentarle a un alumno un reactivo con la complejidad adecuada a sus conocimientos.

Por ello surgen los test adaptativos informatizados (TAI), los cuales se basan en la presentación de los ítems en grupos que pueden ir de uno en uno hasta de tres en tres, donde la selección de un siguiente ítem está condicionada a la respuesta del ítem anterior. Los primeros intentos de aplicación de los TAI se realizaron a partir de los años setenta gracias a los desarrollos de modelos teóricos de la Teoría de Respuesta al ítem (TRI) y de la computación. El primer sistema integrado para aplicar bancos de ítems de forma adaptativa fue editado por la Assessment Systems Corporation en 1984 (Sierra-Matamoros et al., 2007). Los primeros TAI elaborados se usaron en la milicia y fueron desarrollados por ASVAB en los Estados Unidos y MicroPAT en Europa. En la actualidad, el uso de los TAI se ha extendido a varios test estandarizados como el Test of English as a Foreign Language (TOEFL).

Un Test Adaptativo Informatizado (TAI) se define como un instrumento conformado por un banco de ítems, calibrado según principios de la Teoría de Respuesta al Ítem, que implica un procedimiento para estimar el nivel de habilidad del examinado, y otro procedimiento para la selección del ítem más adecuado de acuerdo con dicho nivel (Sierra-Matamoros et al., 2007); cuya elaboración, aplicación, clasificación y actualización se realiza por medio de un soporte computacional (Hambleton et al., 1991). Algunas de las pruebas adaptativas tienen su antecedente en su versión escrita y en ocasiones la versión escrita permite alimentar el banco de reactivos de las versiones adaptativas (Jodoin et al., 2006).

Para la implementación de un TAI es necesario un algoritmo adaptativo que describa el comportamiento que debe tener la evaluación, y el cual se basa en la respuesta al ítem presentado en cada iteración durante la prueba. Para selección y presentación de los ítems la TRI y la Teoría de Decisión Bayesiana (TDB) se han considerado como alternativas de solución. La TRI ofrece modelos matemáticos y métodos estadísticos para la justificación en la elección de los ítems, así como, la estimación del nivel de habilidad para el estudiante (Jodoin et al., 2006). Por su parte, la TDB se basa en razonamiento probabilista, el cual considera el vector de respuestas que provee el estudiante para decidir si se domina o no, un determinado tópico (Luecht et al., 2006). Aunque estas teorías han guiado el desarrollo

de los TAI, aún no consideran las habilidades cognitivas en el estudiante, las cuales pueden afectar o mejorar el desempeño durante una evaluación.

Para el desarrollo de este proyecto de investigación, también se analizaron los fundamentos de la psicometría para la construcción de los perfiles cognitivos del estudiante, los cuales permitirán a los docentes tener un mejor panorama de las habilidades en sus estudiantes y en el caso de las IES una mejor percepción de los estudiantes que está admitiendo, proporcionando de manera más precisa una proyección del buen desempeño del estudiante.

La psicometría a diferencia de la evaluación psicológica hace énfasis en la medición de instrumentos y test, no en el diagnóstico de una persona. Entre 1904 y 1910 Charles Spearman funda la Teoría Clásica de los Test (TCT), casi a la par con el desarrollo de instrumentos que permiten medir la inteligencia (García, 2016). A partir de este trabajo crece el desarrollo de las teorías de medición y de métodos estadísticos para asegurar que un instrumento satisface una serie de indicadores que permita afirmar que su construcción está correctamente medida. Sin embargo, desde los comienzos de la TCT se observó que un instrumento de medición está estrechamente ligado con la población que lo contesta y el nivel de conocimientos está relacionado con la prueba.

En 1980 Rasch plantea la Teoría de la Respuesta al Ítem (TRI), estableciendo dentro de sus principios que su unidad básica de estudio es el ítem y no la prueba completa (Fernández, 2010). Igualmente, la aptitud del examinado es la que manifiesta al momento de responder y no depende de la prueba en particular o de la población considerada.

El desarrollo de la TCT se centra en el análisis de la prueba como una unidad. En comparación con la TRI, no hay tanto énfasis en el estudio de los reactivos individuales ni se han desarrollado técnicas que correspondan a ello. Sin embargo, se han creado una gran cantidad de indicadores para asegurar que determinado instrumento de evaluación es robusto, como son el índice de dificultad y el índice de discriminación (Chávez y Saade, 2010).

Los modelos de la TRI tienen como propósito la estimación de los parámetros de los ítems, tomando en cuenta la invarianza de la medida respecto del instrumento y del grupo de examinados. Los supuestos que la fundamentan a la TRI son la curva característica, la unidimensionalidad y la independencia local (Hambleton et al., 1991; Lord, 1980; Muñoz, 1997).

Los supuestos de unidimensionalidad e independencia local, resultan imprescindibles en un TAI, pues al presentar un número reducido de ítems, la estimación final del nivel de habilidad no puede depender de que los ítems se refieran a uno u otro rasgo (Gil et al., 2000).

En relación con el procedimiento de estimación del nivel de habilidad, en los TAI se toman en cuenta las respuestas que el examinado da a los ítems presentados y se lleva a cabo utilizando las características psicométricas de los ítems, principalmente el índice de dificultad. Así mismo, la forma en la que dicha estimación se realiza varía de acuerdo con el tipo de estrategias de funcionamiento de un TAI

Un test adaptativo informatizado está compuesto por los siguientes elementos.

- Un banco de Ítems con características psicométricas establecidas previamente por el modelo de la Teoría de la Respuesta al ítem.

- Un procedimiento para implementar el inicio y el final de la prueba, además, una estrategia para establecer los ítems a presentar, de manera que sean los más deseables.
- Estadística de estimación de los rasgos de las personas de un mismo nivel.

Los componentes principales de un TAI integran un proceso iterativo de 3 pasos que se describe a continuación y se muestra gráficamente en la figura 1.



Figura 1. Proceso Iterativo de un TAI  
Fuente: Elaboración propia.

1. *Evaluación del criterio de paro.* Se debe establecer cierto número de ítems para administrar y presentar.
2. *Selección de reactivos.* Se deben establecer los criterios para seleccionar el ítem que debe presentarse después de la respuesta ofrecida por el estudiante. Este es uno de los aspectos con mayor investigación en el campo de los TAI.
3. *Respuesta.* Es el único campo en el que se involucra el examinado, ya que después de automatizar la prueba, debe hacerlo adaptativo, proceso que se deriva de la respuesta proporcionada por él.

El objetivo de un TAI es optimizar tiempo y seguridad durante su aplicación. En el proceso de generación de un TAI se debe considerar los siguientes aspectos (Barrada et al., 2010).

- Las propiedades psicométricas, donde se mide con exactitud el diseño de la prueba y su validez de contenido y de respuestas.
- El banco de reactivos, que almacena los rasgos y habilidades más sobresalientes en la prueba aplicada, de donde se extraen las características más sobresalientes.
- La estimación de los parámetros de los ítems, que proporcionan los modelos de elaboración para las escalas de medición de los diferentes parámetros que componen el banco de ítems.
- La comparación de las estimaciones de los parámetros en las pruebas, de forma que al confrontarse diferentes test puedan ser equivalentes los puntajes deseados.
- Un programa que convierta las pruebas aplicadas a un TAI, el objetivo, ofrecer una visión actualizada de la aplicación de los test.

Con esto, se presenta la estructura básica de un TAI, se muestra en la figura 2.

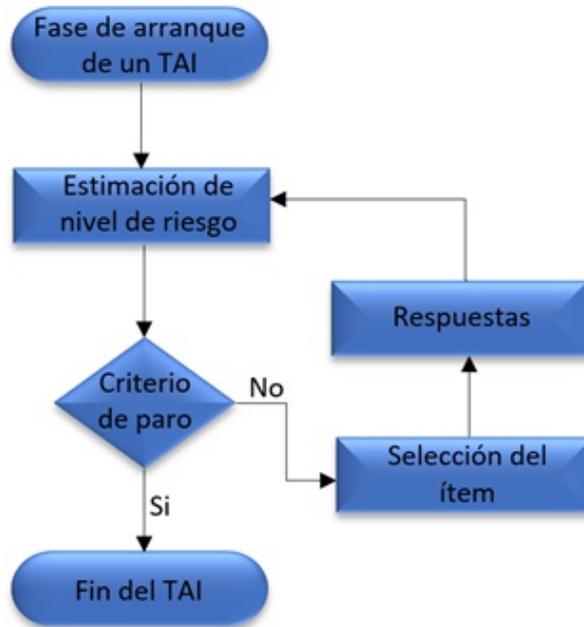


Figura 2. Diagrama de Flujo de un TAI  
Fuente: Elaboración propia.

El supuesto que soporta a los TAI es la construcción de pruebas hechas a la medida, es decir, las pruebas se configuran a partir de la respuesta que el examinado da a un ítem y su correspondencia con un nivel estimado de habilidad que permite la elección y presentación de uno de los ítems restantes del banco (Hambleton et al., 1985).

Este procedimiento implica, la posibilidad de presentar distintos ítems y pruebas a diferentes examinados, pues se supone que poseen diferentes niveles de habilidad, aunado al uso de pocos ítems, ya que se aplican aquellos que más informan acerca del nivel de habilidad (Muñiz, 1997). Así, el uso de los TAI proporciona una evaluación de habilidades o conocimientos precisos y eficientes incrementando las posibilidades de mejorar los procesos de enseñanza y aprendizaje (Olea y Ponsoda, 1998).

Actualmente el desarrollo de los TAI ha pasado al área de la inteligencia artificial, incorporando los Sistemas de Tutor Inteligente (STI), aparte de los propios TAI, la TRI y la TCT. Un sistema de tutor inteligente es un sistema de software que utiliza técnicas de inteligencia artificial (IA) para representar el conocimiento e interactuar con los estudiantes para enseñárselo (Simanca Herrera y Abuchar Porras, 2014). Tal es el caso de la plataforma Algorithm of Intelligent Evaluation (AEI), donde se propone un prototipo que utiliza como algoritmo adaptativo una red neuronal de tipo *backpropagation* tomando como datos de entrada los resultados de las evaluaciones parciales disgregados en dos formas.

- a) Tomando el caso de resolución de los ejercicios.
- b) Tomando ejercicios en función de los logros cognitivos.

Esto usando datos provenientes de las evaluaciones parciales de los estudiantes a fin de poder predecir un futuro rendimiento (Barrada et al., 2010; Backhoff et al., 1996; Chang y

Ying, 2004), sin embargo, el modelo solo se propone de forma teórica, considerando un sistema completo de calibración de reactivos y la parte adaptativa.

Otro modelo encontrado en la literatura hasta este momento es el modelo de evaluación adaptativa y personalizada mediante razonamiento probabilístico, MODEVA. Este modelo utiliza redes bayesianas e integra factores del perfil del estudiante tales como: gustos, preferencias, idioma, localización, estilos de aprendizaje etc. (Toledo et al., 2013). De la misma manera en la literatura se identifican alternativas para la construcción de TAI utilizando algoritmos genéticos y árboles de decisión (Backhoff et al., 1996; Toledo et al., 2013), lógica difusa (Chang y Ying, 2004), pero ninguno de ellos considera el perfil cognitivo del estudiante.

### **1.2. Objetivo**

El objetivo principal de este trabajo fue *diseñar y desarrollar un prototipo de prueba adaptativa basada en un algoritmo de selección inteligente que permita medir las habilidades de razonamiento matemático, generando así, un perfil cognitivo que sirva para dar seguimiento al desempeño académico del estudiante a lo largo de su estadía en la institución de educación superior (IES).*

Aportar un perfil cognitivo al resultado de evaluaciones en procesos de admisión para alumnos que hayan concluido su instrucción media superior, es fundamental para poder realizar un seguimiento académico eficaz. Lograr esto, requiere de realizar un análisis de los requerimientos del sistema, definir procesos cognitivos a partir de la Taxonomía de Bloom, diseñar e implementar un métodos de selección de reactivos basada en una técnica de aprendizaje estadístico automático, desarrollar y aplicar una metodología para la implementación de la prueba adaptativa, obtener un perfil cognitivo del alumno a partir de sus respuestas a la prueba adaptativa diseñada y obtener puntajes consistentes entre la prueba adaptativa y la escrita.

### **1.3. Población**

La población considerada en este proyecto consta de estudiantes que han concluido su instrucción del nivel medio superior (nueve años de escolaridad) y aspiran a cursar algún programa educativo de nivel superior en la Benemérita Universidad Autónoma de Puebla, México.

El tamaño de la muestra para la validación de los ítems fue de 10.000 estudiantes quienes tomaron el Seminario de Orientación de la Benemérita Universidad Autónoma de Puebla, con una edad promedio de 18 años y provenientes de diferentes sistemas educativos de nivel medio superior que se ofrecen en la República Mexicana y que aspiran a ingresar a la máxima casa de estudios del Estado de Puebla en México.

## **2. Metodo**

### **2.1. Niveles Cognitivos**

Lo cognitivo es aquello que pertenece o está relacionado al conocimiento, el desarrollo cognitivo enfoca los procedimientos intelectuales y las conductas que emanan de estos procesos. Un perfil cognitivo se define como el conjunto de las propiedades intelectuales que rigen la conducta de una persona. Enfocando estos conceptos a la educación, diremos que la capacidad que permite desarrollar conocimientos recibe el nombre de cognición, es

decir, se trata de la habilidad para asimilar y procesar datos, valorando y sistematizando la información a la que se accede a partir de la experiencia, la percepción u otras vías.

Los procesos cognitivos, por tanto, son los procedimientos (operaciones mentales) que llevará cabo el estudiante para incorporar nuevos conocimientos. En dichos procesos intervienen facultades muy diversas, como la inteligencia, la atención, la memoria y el lenguaje.

Como es sabido, la Taxonomía de Bloom o Taxonomía de Objetivos de la Educación es una clasificación de los objetivos de una acción educativa y sirve como punto de partida para el diseño de objetivos de aprendizaje. Los objetivos se clasifican en función de la dimensión a la que corresponden: dominio psicomotor, dominio afectivo y dominio cognitivo, siendo este último de nuestro interés (Bloom, 1956). En el dominio cognitivo se distinguen 6 niveles en orden ascendente, que se pueden ver gráficamente en la figura 3.

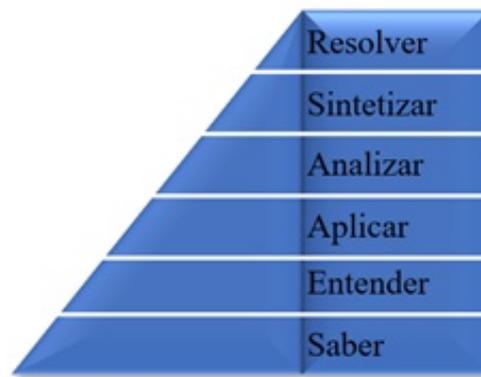


Figura 3. Taxonomía de Bloom  
Fuente: Elaboración propia.

La Taxonomía de Bloom es jerárquica, esto es, se asume que el aprendizaje a niveles superiores depende de la adquisición del conocimiento y habilidades de ciertos niveles inferiores. Para este proyecto de investigación se propuso integrar los seis objetivos de la Taxonomía de Bloom en 3 grupos cognitivos y jerárquicos que se describen en el cuadro 1.

Cuadro 1. Grupos Cognitivos

GRUPO COGNITIVO	TAXONOMÍA DE BLOOM
Entender	Saber, Entender
Aplicar	Aplicar, Analizar
Solución de Problemas	Sintetizar, Resolver

Fuente: Elaboración propia.

La taxonomía propuesta se puede visualizar en la figura 4.

### 2.2. Formato y Estructura de los Ítems

Los reactivos que se utilizaron para el desarrollo de este sistema son reactivos calibrados con la TCT y la TRI, los parámetros que se consideran son el índice de dificultad (fácil, mediano y difícil), el índice de discriminación, el contenido específico y la categoría cognitiva al que pertenecen. Además, cada reactivo es relacionado con un puntaje en una

escala calibrada de 200 a 800 que nos indica el puntaje mínimo requerido por un estudiante para contestarlo correctamente.

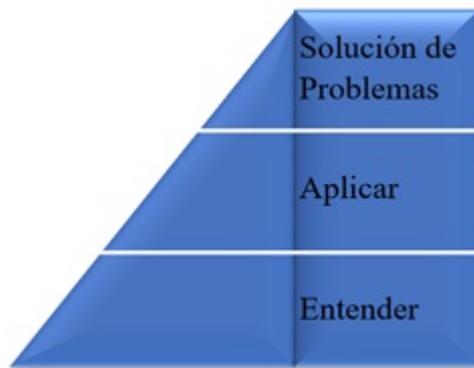


Figura 4. Taxonomía Propuesta  
Fuente: Elaboración propia.

El tipo de reactivos manejados por el test son los de tipo BIC (Banco de Ítems Calibrados), unidimensionales y de opción múltiple que para ser calificados se vuelven dicotómicos, si la respuesta es incorrecta se asigna un 0 y si es correcta se asigna un 1.

Cada reactivo se integra con los siguientes elementos:

1. Enunciado del problema o ítem
2. Cinco opciones de respuesta
3. Una clave o respuesta correcta
4. Contenido específico
5. Proceso Cognitivo
6. Nivel de dificultad
7. Nivel de desempeño
8. Puntaje

Cada uno de los procesos cognitivos propuestos en la figura 4 se describen a continuación y, a partir del análisis psicométrico realizado a los ítems de las pruebas escritas, presentamos algunos de los ítems del área de álgebra clasificados según esta categoría cognoscitiva.

*Entender.* Un estudiante en esta categoría muestra su capacidad para comprender los conceptos y propiedades algebraicas. El cuadro 2 muestra un reactivo con este tipo de características.

*Aplicar.* Un estudiante en esta categoría muestra su capacidad para utilizar operaciones o fórmulas que lo llevan a obtener la solución. El cuadro 2 muestra un ejemplo.

*Resolver problemas.* Un estudiante en esta categoría tiene la capacidad de plantear soluciones a problemas reales. En algunos casos se requiere construir un algoritmo (serie de pasos) que lleven a la solución. Los cuadros 4 y 5 muestran dos ejemplos de este proceso cognitivo.

Cuadro 2. Ítem de Entender

La expresión que permite obtener la distancia del origen al punto $(a, b)$ es	
$\frac{a-b}{a+b}$ $\sqrt{a^2 - b^2}$ $\sqrt{a^2 + b^2}$	
<b>Proceso Cognitivo</b>	Entender
<b>Respuesta Correcta</b>	D
<b>Dificultad</b>	Difícil
<b>Discriminación</b>	0,80
<b>Área</b>	Álgebra
<b>Contenido Específico</b>	Distancia entre dos puntos
<b>Descripción</b>	Conoce el concepto de distancia entre dos puntos.
<b>Conocimiento Básico</b>	Identificar puntos en el plano cartesiano. Sumar y multiplicar números. Potencia cuadrada de un número.
<b>Conocimiento Específico</b>	Identificar la expresión para calcular la distancia entre dos puntos.

Fuente: Elaboración propia.

Cuadro 3. Ítem de Aplicar

Si $x = -1$ , ¿cuál es el valor de $f(x) = x^2 + 1$ ?	
$3$ $2$ $0$ $-1$	
<b>Proceso Cognitivo</b>	Aplicar
<b>Respuesta Correcta</b>	B
<b>Dificultad</b>	Mediano
<b>Discriminación</b>	0,72
<b>Área</b>	Álgebra
<b>Contenido Específico</b>	Evaluación de Funciones
<b>Descripción</b>	Realizar operaciones sustituyendo el valor de $x$ para llegar al resultado.
<b>Conocimiento Básico</b>	Sumar y multiplicar. Función cuadrática.
<b>Conocimiento Específico</b>	Evaluar funciones sustituyendo la variable por el valor dado.

Fuente: Elaboración propia.

Cuadro 4. Ítem de Resolver Problemas

Judith desea subir al techo de su casa utilizando una escalera. Si la altura de su casa es $4m$ y la escalera se coloca en una pared de tal forma que su base está a $2m$ de la pared. ¿Cuál es la longitud de la escalera, si desea que se alcance el techo de la casa?	
$20$ $6$ $2\sqrt{5}$ $\sqrt{6}$	
<b>Proceso Cognitivo</b>	Resolver Problemas
<b>Respuesta Correcta</b>	C
<b>Dificultad</b>	Difícil
<b>Discriminación</b>	0,67
<b>Área</b>	Álgebra
<b>Contenido Específico</b>	Teorema de Pitágoras
<b>Descripción</b>	Requiere realizar un dibujo con la escalera y la pared para formar un triángulo rectángulo y después aplicar el Teorema de Pitágoras en la solución del problema.
<b>Conocimiento Básico</b>	Sumar y multiplicar. Potencia cuadrada de un número. Raíz cuadrada.
<b>Conocimiento Específico</b>	Triángulo Rectángulo. Teorema de Pitágoras.

Fuente: Elaboración propia.

Cuadro 5. Ítem de Resolver Problemas

Por pintar una casa dos pintores cobraron \$480.00 y lo hicieron en 6 horas. Si uno de ellos ganó por hora \$42.00, ¿cuánto ganó por hora el segundo pintor?	
	38
	80
	228
	252
<b>Proceso Cognitivo</b>	Resolver Problemas
<b>Respuesta Correcta</b>	A
<b>Dificultad</b>	Fácil
<b>Discriminación</b>	0,51
<b>Área</b>	Álgebra
<b>Contenido Específico</b>	Operaciones básicas: suma, resta, división y multiplicación
<b>Descripción</b>	Problemas verbales de álgebra que involucra operaciones básicas: suma, resta, división y multiplicación, donde se requiere construir una serie de pasos para obtener la solución.
<b>Conocimiento Básico</b>	<i>Sumar, restar, multiplicar y dividir. Modelar ecuaciones lineales en una variable.</i>
<b>Conocimiento Específico</b>	<i>Resolver ecuaciones lineales en una variable.</i>

Fuente: Elaboración propia.

Por cuestiones prácticas y de simplicidad, se consideró únicamente el área de álgebra y se ha tenido que delimitar al siguiente contenido específico.

1. Operaciones básicas de números reales.
2. Propiedades de los números enteros.
3. Razones y proporciones.
4. Propiedades de los exponentes enteros.
5. Operaciones con radicales.
6. Ecuaciones de primer orden en una variable.
7. Ecuaciones de segundo orden en una variable.
8. Factorización y Factor común.
9. Diferencia de cuadrados.
10. Desigualdades de primer orden en una variable.
11. Desigualdades de segundo orden en una variable.
12. Evaluación de expresiones algebraicas.
13. Funciones algebraicas.
14. Sistemas de ecuaciones lineales en dos variables.
15. Problemas de álgebra.

Utilizando un análisis de clúster (Sokal, 1958), a partir de la categorización de los procesos cognitivos y del contenido específico de cada reactivo se construyen los niveles cognitivos, que nos dan una radiografía de los estudiantes según los puntajes obtenidos y las respuestas ofrecidas a cada uno de los reactivos del área. El cuadro 6 muestra los niveles que se obtienen.

Cada nivel cognitivo se describe de acuerdo con el contenido (básico y específico) y la categoría cognitiva a la que pertenece el reactivo. Por ejemplo, en el Nivel I, *el alumno entiende las propiedades de los números enteros positivos. Entiende el concepto de factor común en*

una expresión algebraica. Conforme su puntaje se aproxima a la frontera superior, aplica operaciones básicas como la suma, la resta, la multiplicación y división de números reales.

Cuadro 6. Niveles Cognitivos

NIVEL COGNITIVO	PUNTAJE
I	200 – 360
II	361 – 500
III	501 – 670
IV	671 – 800

Fuente: Elaboración propia.

### 2.3. Implementación del Sistema de Evaluación

La metodología funciona como el soporte conceptual que rige la manera en que se aplican procedimientos en una investigación o en este caso, el desarrollo e implementación de una prueba adaptativa. Para cubrir este objetivo, se utilizará un proceso que comprende las siguientes fases para la construcción de un TAI (Pinsach y Diéguez, 1999).

- Planificación y prospección de un TAI.
- Producción y calibración del Banco de ítems.
- Implementación y ejecución de un TAI.
- Explotación y gestión de un TAI.
- Fiabilidad y validez del TAI.

Para poder llevar a cabo la implementación de un sistema es indispensable recurrir a una de las metodologías de desarrollo de software diseñadas para este fin. El sistema de evaluación fue desarrollado bajo el modelo de desarrollo de software en cascada con algunas modificaciones, ya que la implementación del software debe ser rigurosa para poder determinar cierto nivel de calidad en el producto final.

Un aspecto específico a tratar es la construcción de un algoritmo que permita seleccionar el siguiente mejor reactivo a mostrar, parte central del desarrollo de este sistema.

Debido a que el desarrollo de software se hace por módulos y cada módulo forma siempre un sistema mayor, el trabajo comienza estableciendo los requisitos de todos los elementos del sistema y luego asignando algún subconjunto de estos requisitos al sistema completo. En esta etapa son detallados los requerimientos básicos para el funcionamiento del sistema.

1. Se requiere el diseño de una interfaz intuitiva para el usuario.
2. Para acceder al sistema se debe solicitar un *usuario* y una *contraseña*, donde se observan las acciones que podrán realizar el usuario administrador y el usuario estudiante.
3. Los usuarios deben poder registrarse para realizar un examen.
4. Antes de realizar la prueba, el usuario podrá visualizar las instrucciones del examen para una mayor información al momento de realizarlo.
5. Los reactivos serán presentados de manera individual, mostrando sus opciones de respuesta y un botón que permita ir al siguiente reactivo.

6. El sistema debe proporcionar como dato de salida una tabla que contenga el perfil cognitivo del sustentante y el puntaje obtenido en la prueba.
7. Se requiere diseñar e implementar una base de datos que administre el banco de reactivos calibrados.
8. Se necesita diseñar un módulo que administre la comunicación cliente servidor de la base de datos con la interfaz para el usuario, así como, procesar un algoritmo de selección de reactivos.
9. Diseñar e implementar un algoritmo que administre los reactivos y seleccione el siguiente mejor reactivo.

La interfaz debe ser intuitiva para el usuario, debe contener información que sirva como guía al usuario al momento de utilizar el sistema, contener colores que proyecten seguridad y tranquilidad al usuario, en el caso de la interfaz principal deberá pedir el usuario y contraseña para acceder al sistema, en cada uno de sus roles.

*Rol Administrador.* El administrador será aquel que podrá ingresar y modificar los datos de entrada del sistema.

- Registrar a los usuarios
- Administrar la base de datos
- Visualizar resultados

*Rol Estudiante.* El estudiante tendrá un acceso restringido.

- Registrar
- Realizar test
- Visualizar resultados

En la figura 5 se visualiza el diagrama de caso de uso: Roles del sistema

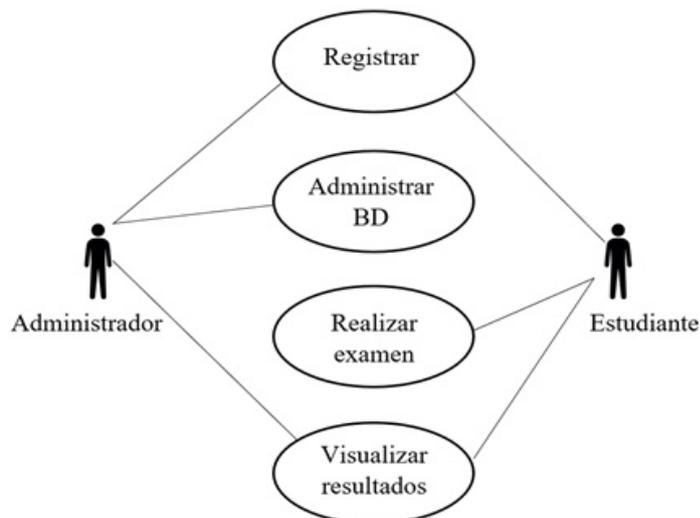


Figura 5. Diagrama de uso: Roles del Sistema  
Fuente: Elaboración propia.

#### 2.4. Diseño del Sistema de Evaluación

El diseño de software es el proceso mediante el cual se traducen los requisitos del sistema en modelos que permitan visualizar de manera detallada y gráfica las funcionalidades del sistema. El proceso como ya se ha mencionado consiste en dividir en componentes del software denominados módulos. En esta fase, el análisis de los requerimientos proporciona la facilidad de especificar los módulos que conforman el sistema.

*Módulo 1:* Requisitos funcionales y no funcionales.

*Módulo 2:* Base de Datos.

*Módulo 3:* Arquitectura de software.

*Módulo 4:* Diseño del algoritmo.

Los requerimientos, requisitos, propiedades o restricciones determinadas de forma precisa de un sistema describen los servicios que ha de ofrecer y las restricciones asociadas a su funcionamiento. Los requerimientos funcionales expresan como interacciona el sistema con su entorno y cuáles serán sus estados y funcionamiento del sistema, los no funcionales son las restricciones sobre el espacio de las posibles soluciones, en otras palabras, los requisitos funciones definen que debe hacer un sistema y los no funcionales definen como debe ser el sistema.

Los casos de uso describen el modo en que un actor interactúa con el sistema, narran el comportamiento dinámico del sistema desde el punto de vista del actor y pueden expresar tanto requerimientos funcionales como no funcionales.

El cuadro 7 muestra el caso de uso para registrar usuarios, el cuadro 8 muestra el proceso de administrar una base de datos, el cuadro 9 muestra el caso de uso de realizar el examen y en el cuadro 10 se puede visualizar el proceso para ver los resultados de un examen realizado.

Cuadro 7. Caso de Uso: Registrar Usuario

<b>Caso de Uso</b>	Registrar Usuario
<b>Actores</b>	Administrador, Estudiante
<b>Propósito</b>	Permite el registro de los usuarios que quieran utilizar el sistema
<b>Descripción</b>	Este caso de uso se ejecuta cuando un usuario quiere registrarse en la plataforma
<b>Acción del Actor</b>	Respuesta al sistema. El sistema valida la información
<b>Caso Alternativo</b>	Datos incorrectos, el correo no es válido

Fuente: Elaboración propia.

Cuadro 8. Caso de Uso: Administrador de BD

<b>Caso de Uso</b>	Administrador BD
<b>Actores</b>	Administrador
<b>Propósito</b>	Permite visualizar, actualizar, guardar o resguardar la BD
<b>Descripción</b>	Este caso de uso se ejecuta cuando el administrador desea modificar o consultar la BD
<b>Acción del Actor</b>	Respuesta al sistema. El sistema muestra en pantalla la BD
<b>Caso Alternativo</b>	Si el usuario y/o la contraseña no son correctos, el sistema no muestra esta opción

Fuente: Elaboración propia.

Cuadro 9. Caso de Uso: Realizar Test

<b>Caso de Uso</b>	Realizar Test
<b>Actores</b>	Estudiante
<b>Propósito</b>	Permite al usuario el acceso a la aplicación para realizar la prueba
<b>Descripción</b>	Este caso de uso se ejecuta cuando un usuario quiere acceder a realizar la prueba en plataforma
<b>Acción del Actor</b>	Respuesta al sistema. El sistema muestra las preguntas para que el usuario responda la prueba

Fuente: Elaboración propia.

Cuadro 10. Caso de Uso: Visualizar Resultados

<b>Caso de Uso</b>	Visualizar resultados
<b>Actores</b>	Administrador, Estudiante
<b>Propósito</b>	Permite ver los resultados de los Test
<b>Descripción</b>	Este caso de uso se ejecuta cuando el estudiante o el administrador desean ver los resultados en pantalla.
<b>Acción del Actor</b>	Respuesta al sistema. El sistema retorna los resultados en la pantalla
<b>Caso Alternativo</b>	Si no existe registro de la prueba, el sistema no muestra nada

Fuente: Elaboración propia.

### 2.5. Base de Datos y Arquitectura de Software

Una base de datos es el conjunto de datos informativos organizados en un mismo contexto para su uso y inoculación. Para la implementación de este módulo se requiere tener el modelado de datos que permita representar las entidades relevantes de un sistema de información, así como, sus interrelaciones y propiedades.

Se requiere una base de datos que almacene los siguientes elementos

1. Una tabla que almacene los datos de los usuarios
2. Una tabla para asignar los roles de los usuarios
3. Una tabla para almacenar los reactivos
4. Una tabla para cada atributo del reactivo
5. Una tabla que almacene los exámenes realizados

De acuerdo con el Software Engineering Institute (SEI), la arquitectura de software se refiere a las estructuras de un sistema, compuestas de elementos con propiedades visibles de forma externa y las relaciones que existen entre ellos (Bass et al., 2003). Los elementos pueden ser entidades que existen en tiempo de ejecución (objetos, hilos), entidades lógicas que existen en tiempo de desarrollo (clases, componentes) y entidades físicas (nodos, directorios).

Por otro lado, las relaciones entre elementos dependen de propiedades visibles de los elementos quedando ocultos los detalles de implementación. Finalmente, cada conjunto de elementos relacionados de un tipo particular corresponde a una estructura distinta, de ahí que la arquitectura está compuesta por distintas estructuras.

En la arquitectura de tres capas en general (arquitectura multicapa), el cliente implementa la lógica de presentación (cliente fino), el servidor o servidores de aplicación implementan la lógica de negocio y los datos residen en uno o varios servidores de bases de datos, ver figura 6.

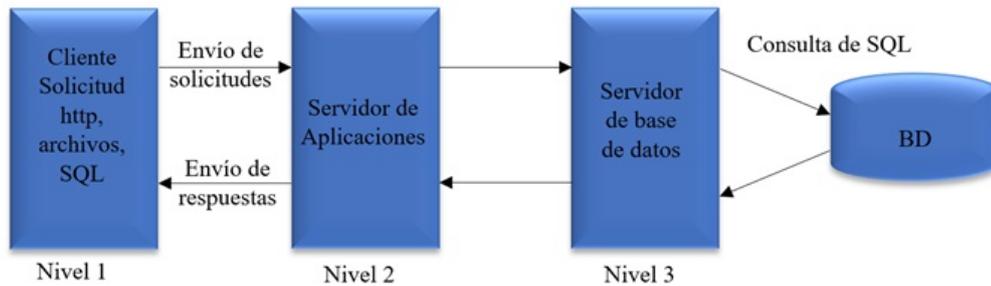


Figura 6. Arquitectura Cliente Servidor de tres niveles  
Fuente: Elaboración propia.

El componente Front-end proporciona la presentación del sistema, trabaja del lado del cliente principalmente se ocupa de los componentes externos de un sitio web o de una aplicación web. Como requisito obligatorio el desarrollo Front-end involucra el conocimiento de lenguajes como HTML, CSS que proporciona estilo a HTML y JavaScript para las páginas web interactivas.

La figura 7 muestra la interfaz principal del sistema de evaluación que utiliza todos los recursos mencionados anteriormente.



Figura 7. Interfaz Principal  
Fuente: Elaboración propia.

El componente Back-end proporciona el acceso a servicios dedicados, trabaja del lado del servidor, para desarrollar este componente son numerosos los lenguajes y los frameworks entre los que se debe elegir, todo depende de los requisitos del sistema. Los más comunes están ASP.NET, PHP, Python, Node.js, Spring.

Sin embargo, no es suficiente con dominar un lenguaje y un framework. Toda aplicación web debe almacenar datos de alguna manera. Por lo tanto, el desarrollo back-end también debe estar familiarizado con las bases de datos. Entre las más comunes destacan SQL Server, MySQL, PostgreSQL, MongoDB.

Para la implementación de la base de datos del sistema de evaluación se eligió MySQL 5.5 como sistema gestor de bases de datos. El modelo relacional de la base de datos se visualiza en la figura 8.

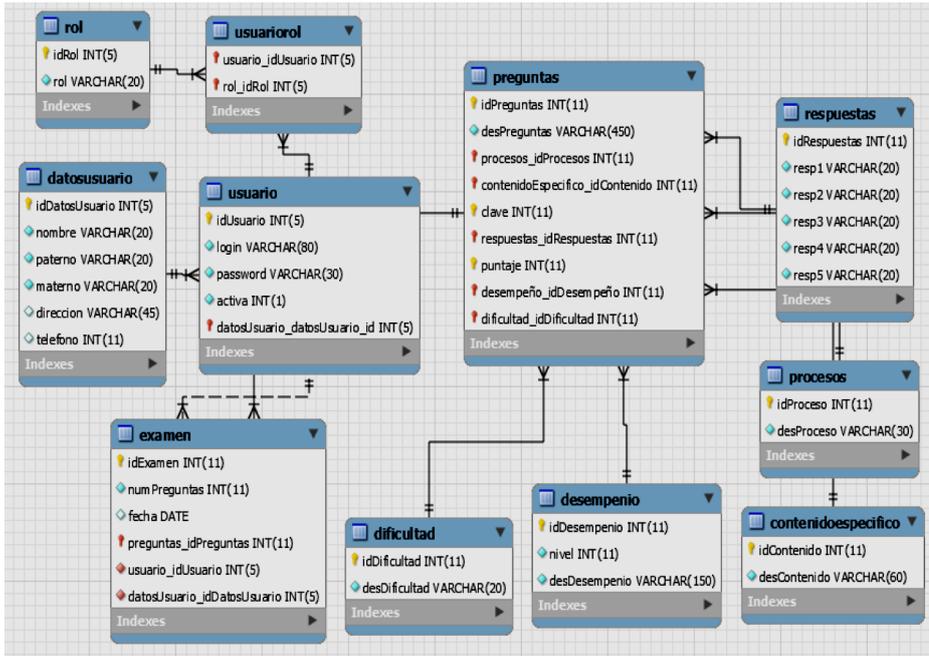


Figura 8. Modelo Relacional  
Fuente: Elaboración propia.

El Modelo vista controlador (MVC) es un estilo de arquitectura de software que separa los datos de una aplicación, la interfaz de usuario y la lógica de control en tres componentes distintos.

1. Modelo
2. Vista
3. Controlador

La ventaja de utilizar MVS es que al realizar un cambio de base de datos, programación o interfaz de usuario solo tocaremos uno de los componentes. Se puede modificar uno de los componentes sin conocer cómo funcionan los otros. Clara separación entre interfaz, lógica de negocio y de presentación. Su funcionamiento lo podemos observar de manera gráfica en la figura 9.

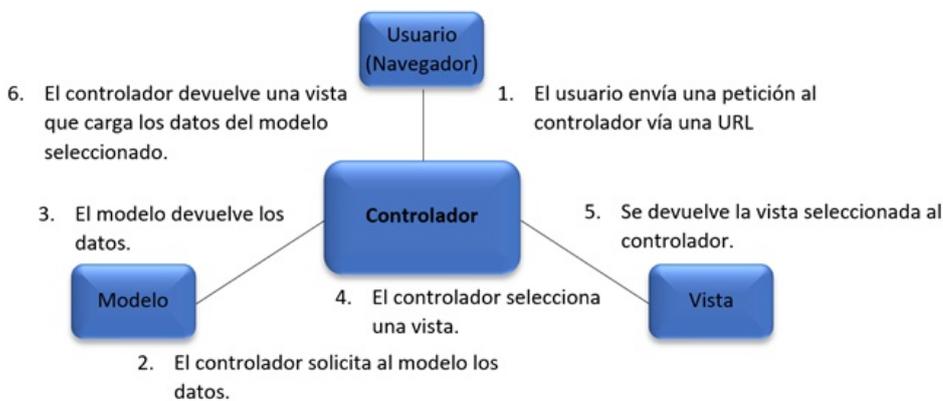


Figura 9. Modelo Vista Controlador  
Fuente: Elaboración propia.

Las figuras 10 y 11 muestran el diagrama de flujo del funcionamiento del sistema de evaluación y se visualiza el modelo que ilustra la arquitectura del sistema de evaluación, quedando como único punto faltante a desarrollar, el algoritmo de selección de reactivos.

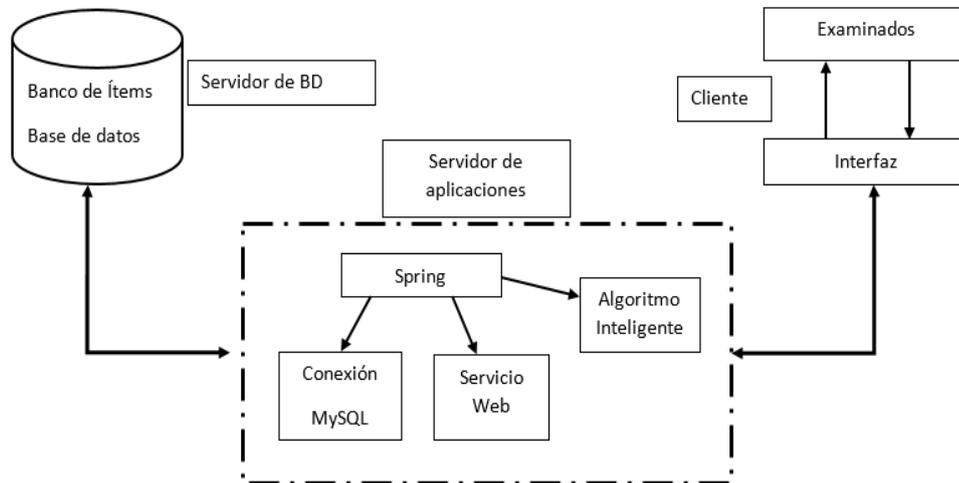


Figura 10. Arquitectura del Sistema de Evaluación  
Fuente: Elaboración propia.

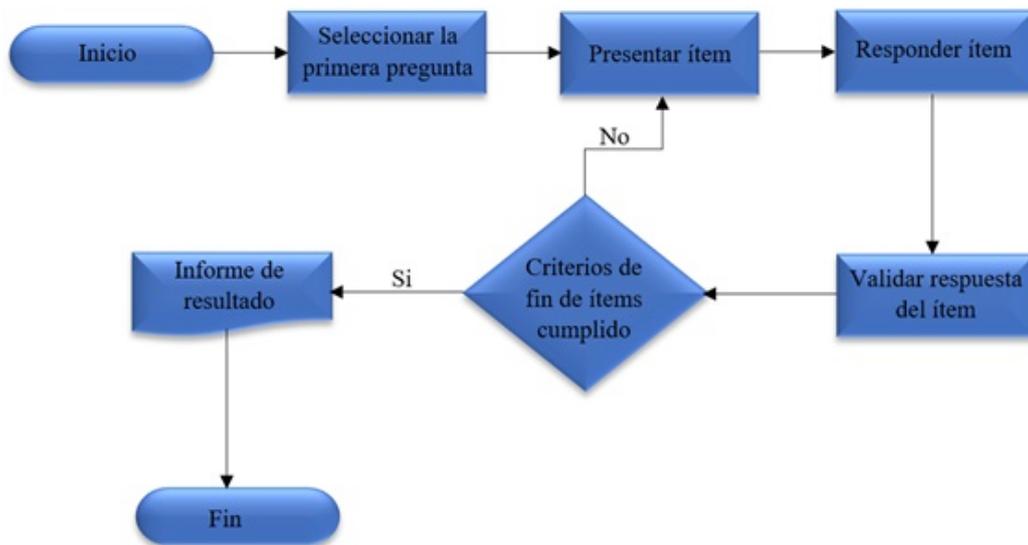


Figura 11. Diagrama de flujo del Sistema de Evaluación  
Fuente: Elaboración propia.

### 2.6. Selección del Algoritmo Adaptativo

Parte importante y fundamental del mecanismo de un TAI es la elección de un algoritmo adaptativo que administre el inicio, la continuación y el final del test. Para el desarrollo del algoritmo es importante mencionar que existen diferentes maneras de combinar estos momentos y cada una de las combinaciones significa una u otra forma de establecer la evaluación (Díaz et al., 1999; Gil et al., 2000).

Con el paso del tiempo han evolucionado los sistemas que proporcionan estrategias adaptativas para un test, de mecanismos rígidos a otros mucho más flexibles. Existen tres procedimientos característicos para realizar una prueba.

1. Binivel: Se pasa al sujeto una serie de ítems empezando por un reactivo fácil y acabando con uno de mayor dificultad. Esta serie de ítems formarán lo que se denomina primer nivel y deben aclarar la zona delimitada entre el último acierto y el primer error. Después se pasa otra serie de ítems que corresponden a esa zona delimitada para establecer así unos ítems de nivel medio.
2. Piramidal y ramificado: lo primero que se hace es administrar un ítem de dificultad media, después dependiendo de si se ha acertado o no, se pasa otro ítem siguiendo la estructura rígida de un árbol.
3. Ramificación variable: aunque la estimación a la que lleguemos con este proceso sea muy parecida a la de los dos anteriores, el camino seguido para ello es menos determinístico y tolera mejor el hecho de que haya aciertos o errores que no sean consistentes.

Es necesario dejar claro que los primeros dos procedimientos se consideran demasiado rígidos por lo que se busca otro procedimiento que resuelva este problema. Además, cuentan con el inconveniente de la excesiva derivación que se le da tanto a los ítems de nivel uno como a los de las primeras ramificaciones, ya que al divulgarse sus resultados pueden dar a conocer su respuesta acertadas a los nuevos sujetos que vayan a examinarse (Belloch, 2004).

En términos generales, los algoritmos adaptativos no dependen exclusivamente de la forma de seleccionar el siguiente ítem. Cada momento tiene sus variantes.

*Inicio.* La manera en que se comience un test influye en el primer ítem, pero además en las instrucciones y ejemplos que se exponen antes de comenzar el test. El primer ítem elegido puede ser de una u otra forma en función del procedimiento a seguir posteriormente. Un ítem de dificultad media si es piramidal, uno sencillo si es binivel o uno al azar si es flexible.

*Continuación.* Aquí se aprecia si es o no posible omitir y rectificar las respuestas. Si se pueden realizar omisiones, podrá optarse por ignorarlas, penalizarlas o considerarlas como un error más. El dejar omitir conlleva riesgos como el hecho de que el sujeto espere a adquirir confianza antes de empezar a responder o que se produzca una rápida divulgación del BIC.

Por otro lado, puede permitirse la rectificación de forma dinámica, bien durante la sesión o después de acabar el test. Otro de los aspectos a decidir es el tiempo máximo que estará un ítem y el control de exposición de este. En los TAIs suele ponerse límite de respuesta para cada ítem. En lo referente al control de exposición, lo que se hace es condicionar el método de selección del siguiente ítem, según su frecuencia de aparición acumulada. También es necesario decidir si se informará a los sujetos del resultado obtenido en la prueba o no. Esta decisión influirá notoriamente en el itinerario de presentación. Por norma general los TAIs no permiten ni omitir ni rectificar, y esta se considera una de las causas para que estos tipos de test no terminen de aceptarse públicamente.

*Final.* Puede haber diversos motivos por los que se termina un TAI, los cuales dan lugar a reflexionar. Después de un número de presentaciones: se verán aumentadas las diferencias en el error típico de medida (ETM) de los sujetos. Tras agotar un tiempo límite

impuesto: la puntuación del sujeto dependerá de las diferencias individuales debidas al tiempo de respuesta.

La *heurística*, como disciplina científica, y en su sentido amplio puede ser aplicada a cualquier ciencia con la finalidad de elaborar medios, principios, reglas, estrategias como ayuda para lograr encontrar la solución más eficaz y eficiente al problema que analiza el individuo. Los procedimientos heurísticos, se dividen en:

1. Principio heurístico, establece sugerencias para encontrar la solución idónea al problema.
2. Reglas heurísticas, señalan los medios para resolver el problema.
3. Estrategias heurísticas, son aquellas que permiten organizar los materiales o recursos compilados que contribuyen a la búsqueda de la solución del problema.

Para la computación, la heurística consiste en encontrar o construir algoritmos con buena velocidad para ser ejecutados. Ejemplos claros de la definición proporcionada son, los juegos informáticos o programas que detectan si un correo electrónico es un spam o no. Los objetivos fundamentales de la programación son encontrar algoritmos con buenos tiempos de ejecución y buenas soluciones, usualmente óptimas.

Las heurísticas proporcionan la herramienta adecuada para el diseño de un algoritmo propio que permita aportar un método de selección, basado en las variantes que el especialista y el desarrollador consideran apropiadas para el TAI.

Por tal motivo, se implementó una heurística que administra el banco de reactivos, tomando como pieza fundamental un reactivo, 5 opciones de respuesta, el nivel de dificultad, el proceso cognitivo y su contenido.

Considerando los tres momentos que debe contener la prueba se determinan las siguientes características

*Inicio.* Se establece como primer reactivo, un reactivo que tenga asignado un nivel de dificultad fácil, un proceso cognitivo entender y un contenido indistinto.

*Continuación.* El algoritmo no permite la omisión ni el rectificar una respuesta, por lo que tendrá un tiempo específico de 60 segundos para contestar un reactivo y no permite regresar a preguntas anteriores.

*Final.* Como criterio de paro se consideran dos situaciones:

- El examen consta de 15 reactivos, por lo que, si el alumno contesta de manera satisfactoria, el examen terminará cuando conteste el reactivo número 15.
- Si el sustentante falla a tres reactivos de manera consecutiva el examen termina.

La figura 12 muestra de forma gráfica las características del algoritmo.

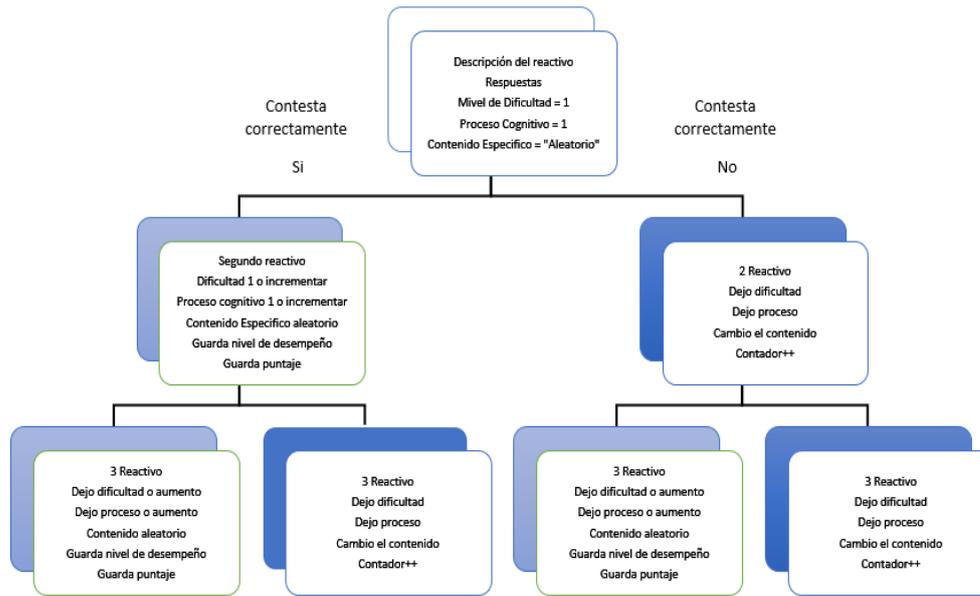


Figura 12. Diseño del Algoritmo  
Fuente: Elaboración propia.

### 2.7. Pseudocódigo

El cuadro 11 muestra el algoritmo propuesto en forma descriptiva.

Cuadro 11. Algoritmo de Selección

ALGORITMO DE SELECCIÓN DE ÍTEMS
Inicio
Nivel de Dificultad = 1;
Proceso Cognitivo=1;
Contenido Temático= $x$ ;
$X:1:16$ ;
Cont=0;
Mientras (ítem $<15$ y Cont $\leq 3$ )
... Presentar el primer ítem
... Leer la respuesta op
... Si op=true entonces
..... Guarda nivel, Guarda puntaje
..... Guarda dificultad o incrementa
Guarda proceso o incrementa
Contenido específico= $x$
... Sino
..... Mantiene dificultad
..... Cambia forzosamente contenido específico
Cont ++
Fin

Fuente: Elaboración propia.

### 2.8. Codificación

Una vez diseñado el algoritmo para la aplicación se inició la fase de codificación. En esta etapa se traduce el algoritmo a un lenguaje de programación de alto nivel, como se utiliza Spring, el lenguaje de programación es JAVA.

El proyecto se divide en 2 partes principales la carpeta "WEB-INF" que contienen las paqueterías necesarias para implementar las interfaces al usuario y la carpeta "Source Packages" que contiene toda la estructura cliente servidor.

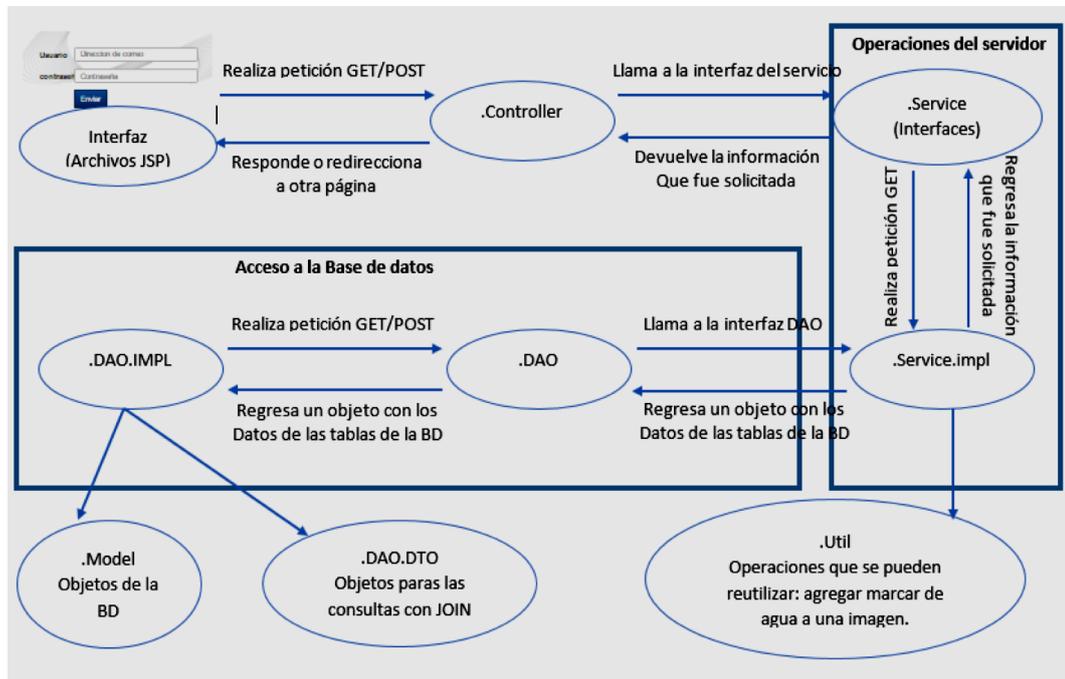


Figura 13. Funcionamiento del Sistema  
Fuente: Elaboración propia.

Para la codificación del algoritmo se ha diseñado una query que implementa el uso de funciones aleatorias y permitirán elegir los parámetros para la selección del siguiente reactivo, estas funciones están en un rango del valor máximo al valor mínimo de cada intervalo.

### 3. Resultados

#### 3.1. Pruebas

El desarrollo de sistemas implica una etapa de pruebas en la que se diseñan situaciones en las que el sistema puede producir fallos. Las pruebas del sistema son un elemento crítico para la garantía de la calidad del sistema y representan una revisión final de las especificaciones, del diseño. El objetivo de la etapa de pruebas es garantizar la calidad del producto desarrollado. La prueba es un proceso que se enfoca sobre la lógica interna del software y las funciones externas, es un proceso de ejecución de un programa con la intención de descubrir algún error (Pressman y Troya, 1988).

En la figura 13, se visualiza la interfaz principal, la cual pide el usuario y contraseña para acceder al sistema, también permite registrarse o recuperar la contraseña utilizando el correo electrónico. Una vez dentro del sistema, se muestra el bloque de instrucciones que el usuario debe tener presente para realizar el test, como se puede ver en la figura 14.

Es importante mencionar que, durante la presentación de los ítems al estudiante, se muestra el tiempo del que dispone para contestar cada uno de los ítems (60 segundos por

cada uno). El tiempo total otorgado a cada estudiante es de 15 minutos, durante el cual podrá responder hasta 15 ítems.

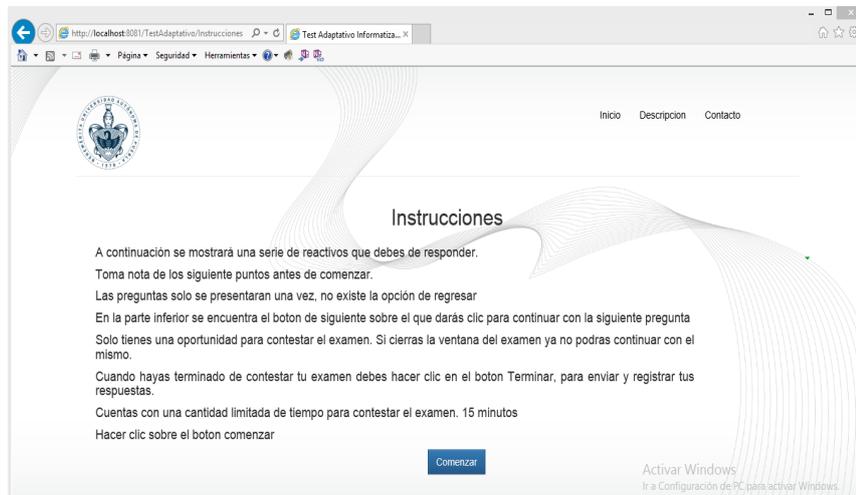


Figura 14. Interfaz de Instrucciones  
Fuente: Elaboración propia.

Existen varios escenarios en los que se puede verificar el buen funcionamiento del software. Para este proyecto se describen tres casos.

- *Caso 1:* Los alumnos contestan los 15 reactivos en no más 15 minutos.
- *Caso 2:* Los alumnos contestan de manera incorrecta 3 reactivos consecutivamente.
- *Caso 3:* Los alumnos contestan de manera correcta e incorrecta los 15 reactivos alternadamente.

Estos 3 casos pueden encontrarse en el cuadro 12, en la que las filas representan los alumnos evaluados y las columnas las preguntas correspondientes al examen. El valor 1 corresponde a las respuestas correctas y el valor 0 a las incorrectas. Es importante aclarar que las preguntas presentadas por el sistema a cada estudiante no tienen por qué ser las mismas, aquí solo se presenta la respuesta recibida a la pregunta presentada en el lugar 1, 2, y así sucesivamente hasta la 15, los espacios en blanco representan las preguntas que los estudiantes no contestaron.

Cuadro 12. Consulta Básica

CLAVE	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14	P15
201603062	1	0	1	1	1	1	1	1	0	1	1	1	0	0	0
201601902	1	1	0	1	1	1	1	1	0	1	1	1	1	1	0
201601250	1	1	1	0	0	1	1	0	1	0	1	1	0	0	1
201602255	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1
201601642	0	0	1	1	1	0	0	0							
201602139	1	0	1	1	0	1	1	0	0	1	1	1	0	1	0
201602744	1	0	1	1	1	1	1	1	1	1	0	1	1	0	0
201600485	0	1	0	0	0										
201601696	1	0	1	0	1	1	0	1	0	0	1	0	0	0	
201600337	1	1	0	1	1	1	1	1	0	0	1	1	0	0	0

Fuente: Elaboración propia.

En el cuadro 13 se muestra un ejemplo de un ítem trabajado en el sistema.

Cuadro 13. Ejemplo de Ítem del Sistema

La empresa GASE utiliza una mezcla de combustible para sus vehículos, que consiste en 2/3 partes de gasolina y 1/3 parte de etanol. Si ponen 30 litros de gasolina dentro del tanque de un vehículo, ¿cuántos litros de etanol deberán poner en el tanque para obtener la mezcla deseada?			
10			
15			
40			
45			
60			
<b>Área</b>	Álgebra		
<b>Contenido Temático</b>	Razones y Proporciones		
<b>Proceso Cognitivo</b>	Resolver Problemas		
<b>Descripción</b>	Solución de un problema verbal que conlleva operaciones con número reales		
<b>Dificultad</b>	Difícil		
<b>Índice de Discriminación</b>	0,74		
<b>Respuesta Correcta</b>	(B)		
Dificultad	Mediana (12,17)	Dificultad en el Sistema	Mediana (12,35)
Escala	494	Escala en el Sistema	516

Fuente: Elaboración propia.

### 3.2. Resultados

Los resultados obtenidos en el sistema desarrollado son los datos más importantes de un proyecto, e interpretarlos de manera adecuada es primordial para llegar a la meta definida. Uno de los requisitos importantes de este proyecto es obtener puntajes consistentes entre la prueba adaptativa desarrollada y su versión escrita.

Como se mencionó anteriormente, los reactivos cargados en el sistema de evaluación han sido probados y validados en un evento real a través de tres pruebas escritas aplicadas a una población de 10.000 estudiantes del Seminario de Orientación de la Benemérita Universidad Autónoma de Puebla. Cada prueba consta de 145 ítems, de los cuales 50 corresponden al área de Razonamiento Matemático, y fueron clasificados de acuerdo con su contenido temático y con sus resultados psicométricos.

En la figura 13 se muestra la distribución de la población y de los reactivos al utilizar un análisis de Rasch (TRI) de la prueba escrita (Prieto y Delgado, 2003).

La columna de la izquierda muestra la distribución de los estudiantes en la prueba escrita de acuerdo con el puntaje obtenido, mientras que la columna de la derecha muestra la distribución de los reactivos de acuerdo con el puntaje mínimo requerido por un estudiante para contestar correctamente el ítem. El puntaje promedio de los estudiantes fue de 540 puntos con una desviación estándar de 122 puntos.

Las pruebas se ubicaron con una dificultad de 0,62 (mediana) con una confianza del 79%. El INFIT, *estadístico de ajuste con información ponderada que se enfoca al comportamiento general de un ítem*, para estas pruebas es de 0,9, el ideal es de 1,0, no obstante, para este tipo de pruebas es aceptable un INFIT entre 0,8 y 1,2. El INFIT se calcula con el promedio ponderado de las desviaciones (o diferencias) cuadráticas estandarizadas entre el desempeño observado y el esperado.

MAP OF PERSONS AND ITEMS				
N=10000	MEAN=540	D.S.=122	INFIT=0.9	RMS=.79
MEASURE				MEASURE
800	.			800
	.#			
750	.#		XX	750
	##		X	
700	####		X	700
	#####		XX	
650	#####		XXXX	650
	#####		XXXXXXX	
600	#####		XXXXXXXXXX	600
	#####		XXXXXXXXXX	
550	#####		XXXXXXXXXX	550
	#####		XXXXXXXXXXXXXX	
	##			
500	#####		XXXXXXXXXXXXXXXXXX	500
			XX	
	#####		XXXXXXXXXXXXXXXXXX	
450	#####		XXXXXXXXXXXXXXXXXX	450
			X	
	#####		XXXXXXXXXXXXXXXXXX	
400	#####		XXXXXXXXXXXXXXXXXX	400
	####		XXXXXXXXXXXX	
350	#####		XXXXXX	350
	##		XXX	
300	.#		XX	300
	.#		X	
250	.		X	250
200	.			200

EACH '#' IN THE PERSON COLUMN IS 3 PERSON; EACH '.' IS 1 TO 2 PERSON

Figura 13. Resultados Prueba Escrita  
Fuente: Elaboración propia.

Considerando las 3 versiones aplicadas, la base de datos para el sistema de la prueba adaptativa se conformó con 150 ítems de Razonamiento Matemático validados a través del análisis de Rasch y con la finalidad de comprobar la eficiencia del sistema propuesto, se aplicó el TAI a una muestra determinística de 200 estudiantes de diferentes áreas geográficas. Se cuidó que las edades de los estudiantes estuviesen entre los 17 y los 20 años, buscando que las características fueran lo más parecidas a la población de estudiantes considerada en el estudio.

Después de superar algunos errores técnicos en cuanto al uso y manejo de las tecnologías por parte de los estudiantes, se aplicó exitosamente el TAI utilizando el sistema desarrollado.

Inmediatamente se realizaron los análisis estadísticos usando Rasch (TRI) con la finalidad de comparar los resultados arrojados por el sistema con los resultados obtenidos en la prueba escrita. La figura 14 muestra este análisis.

La columna de la izquierda muestra que la distribución de los estudiantes en el TAI es muy similar al mostrado en la prueba escrita con una media de 525 y una desviación estándar de 129 con una confianza de 80%. La columna de la derecha muestra la distribución de los reactivos de acuerdo con el puntaje mínimo requerido por un estudiante para contestar correctamente el ítem.

MAP OF PERSONS AND ITEMS				
N=200	MEAN=525	D.S.=129	INFIT=1.1	RMS=.80
MEASURE				MEASURE
800	.			800
750	.		XX	750
700	.#		X	700
650	.#		XX	650
600	.#		XXXX	600
550	.##		XXXXXXX	550
500	.###		XXXXXXXXXX	500
450	.####		XXXXXXXXXX	450
400	.#####		XXXXXXXXXXXXXX	400
350	.#####		XXXXXXXXXXXXXX	350
300	.#####		XXXXXXXXXXXXXX	300
250	.#####		XXXXXXXXXXXXXX	250
200	.#####		XXXXXXXXXXXXXX	200

EACH '#' IN THE PERSON COLUMN IS 3 PERSON; EACH '.' IS 1 TO 2 PERSON

Figura 14. Resultados del TAI  
Fuente: Elaboración propia.

Se puede observar una ligera tendencia de los estudiantes a una distribución tipo Chi-cuadrada, esto es lógico toda vez que el sistema inicia con los ítems más fáciles y conforme el estudiante contesta correctamente, se va aumentando la dificultad. El INFIT para estas pruebas es de 1,1, el aumento respecto a la prueba escrita deriva en el hecho del nulo uso de la tecnología en los estudiantes para realizar evaluaciones.

La figura 15 presenta la interfaz del sistema cuando un estudiante concluye la prueba.

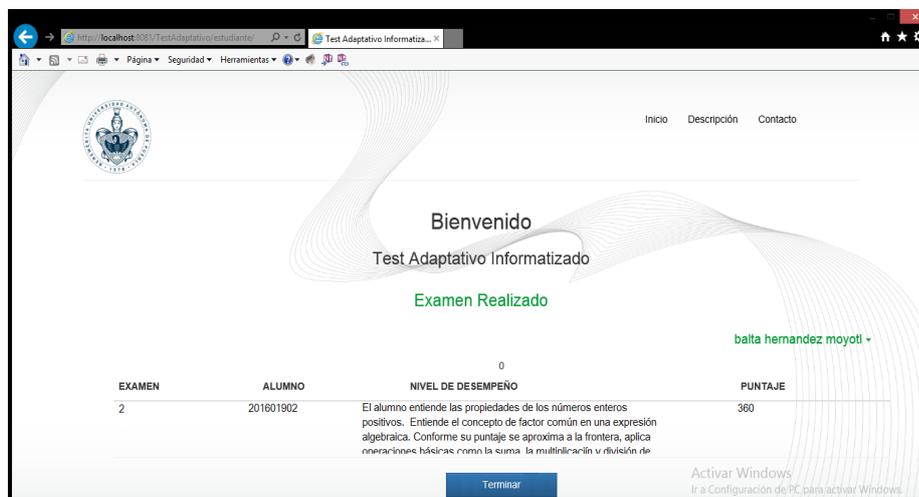


Figura 15. Interfaz de Resultados  
Fuente: Elaboración propia.

El sistema, si bien es susceptible de mejoras, ha mostrado un excelente desempeño permitiendo un ahorro considerable en la utilización de papel e insumos de impresión. Además, cada estudiante utilizó un tiempo que osciló entre 2.5 y 15 minutos, siendo el promedio 95 minutos por alumno, reduciendo considerablemente el tiempo promedio respecto a la prueba escrita que es de 19,2 minutos.

Los resultados obtenidos muestran que la utilización de un TAI para la selección de estudiantes a ingresar a una universidad puede ser aplicado, pero antes se debe dar la suficiente capacitación a los estudiantes con el uso y manejo de la tecnología para realizar evaluaciones, ya que algunos estudiantes no se percatan del tiempo, aún cuando lo tienen visible en pantalla, y terminan perdiendo el reactivo por no contestar.

En la figura 16 se muestra un reporte contrastado con la prueba escrita de algunos estudiantes a los que se les aplicó la prueba escrita y el TAI con una diferencia de 3 meses entre una y otra.

La prueba escrita aplicada a la población contiene 15 ejercicios de álgebra, misma cantidad utilizada en la prueba aplicada a través del sistema. Como se ve en la figura 15, los resultados obtenidos son similares. La coincidencia entre la prueba escrita y el TAI propuesto es de aproximadamente 95%, sin embargo, el TAI propuesto requiere de mucho menos tiempo que la prueba escrita.

Por otra parte, como se observa en la figura 16, todo estudiante que presenta el TAI propuesto recibe, al finalizar su prueba, el puntaje obtenido y una descripción del nivel de desempeño, sin necesidad de tener que esperar entre 15 y 20 días como es el caso de las pruebas escritas.

Núm.	Clave	# Plantel	Prueba Escrita		TAI	
			Puntaje	Nivel	Puntaje	Nivel
1	201603052	1	392	Nivel 2	361	Nivel 2
2	201601902	1	348	Nivel 1	360	Nivel 1
3	201601250	1	552	Nivel 3	556	Nivel 3
4	201602255	1	800	Nivel 4	750	Nivel 4
5	201601642	1	520	Nivel 2	560	Nivel 2
6	201602139	1	392	Nivel 1	456	Nivel 1
7	201602744	1	500	Nivel 2	520	Nivel 3
8	201600485	1	520	Nivel 3	552	Nivel 3
9	201601696	1	288	Nivel 1	224	Nivel 1
10	201600337	1	740	Nivel 4	780	Nivel 4

Figura 16. Comparación Prueba Escrita y TAI

Fuente: Elaboración propia.

## 4. Conclusiones

Este trabajo contribuye en la investigación y desarrollo de pruebas, ya que no solo proporciona los puntajes obtenidos por los aspirantes, sino que además permite llenar el vacío existente sobre indicadores del perfil cognitivo del examinado, al menos los que mide la prueba. En un futuro, se podría considerar, tanto el puntaje como el perfil cognitivo, en la selección de estudiantes que deseen ingresar a algún programa de la universidad.

El trabajo desarrollado sienta las bases para que Sistema de Evaluación Inteligente propuesto pueda extenderse a otras áreas (español, inglés, ciencias naturales, ciencias sociales, etc.), basta definir de forma explícita y precisa los niveles cognitivos, así como los contenidos temáticos específicos y contar con un banco de ítems robusto.

Gracias a la arquitectura trabajada, el Sistema de Evaluación Inteligente propuesto puede mejorarse en cada uno de sus módulos, esto es, proporciona escalabilidad.

Uno de los retos más importantes en el que se está trabajando, es utilizar este sistema como base para el diseño y construcción de perfiles vocacionales, pero desde una perspectiva académica. La Benemérita Universidad Autónoma de Puebla cuenta con más de 85 programas educativos de nivel superior y que se agrupan, de acuerdo con los perfiles de ingreso y egreso, en cuatro Divisiones de Estudios Superiores (DES).

Para ingresar a los programas educativos agrupados en una DES, se aplica una prueba por área de conocimiento que es complementario al examen de admisión y cuyo objetivo es medir el nivel de conocimiento y competencias que posee el estudiante. Por ejemplo, la prueba del área de ciencias naturales y de la salud contiene ítems que corresponden a las asignaturas de química, biología, física y matemáticas trabajadas en el nivel medio superior.

Nuestro objetivo es construir los procesos cognitivos para cada una de las asignaturas evaluadas en todas las DES y diseñar las pruebas, utilizando el Sistema de Evaluación Inteligente desarrollado, para aplicarlas a los estudiantes y poder determinar, con base a estos perfiles, si sus habilidades, destrezas y conocimientos los hacen candidatos idóneos al programa de la DES demandado, o bien, indicarles cuales serías sus posibles opciones.

El proyecto es muy ambicioso y actualmente se encuentra en la fase de construcción de los contenidos temáticos para cada una de las asignaturas.

## 5. Referencias

- Backhoff, E., Ibarra, M. y Rosas, M. (1996). Desarrollo y validación del sistema computarizado de exámenes (sicodex). *Revista de la Educación Superior*, 25(1), 41-54.
- Barrada, J. R., Olea, J., Ponsoda, V. y Abad, F. J. (2010). A method for the comparison of item selection rules in computerized adaptive testing. *Applied Psychological Measurement*, 34(6), 438-452. <https://doi.org/10.1177/0146621610370152>
- Bass, L., Clements, P. y Kazman, R. (2003). *Software architecture in practice*. Reading, MA: Addison-Wesley.
- Belloch, C. (2004). *Recursos tecnológicos para la evaluación psicoeducativa*. Valencia: Universidad de Valencia.
- Bloom, B. S. (1956). *Taxonomy of Educational Objectives, Handbook: The Cognitive Domain*. Nueva York, NY: David McKay.
- Chang, Y. I. y Ying, Z. (2004). Sequential estimation in variable length computerized adaptive testing. *Journal of Statistical Planning and Inference*, 121(2) 249-264. [https://doi.org/10.1016/S0378-3758\(03\)00119-8](https://doi.org/10.1016/S0378-3758(03)00119-8)
- Chávez, C. y Saade, A. (2010). *Procedimientos básicos para el análisis de reactivos*. Ciudad de Mexico: Centro Nacional de Evaluación para la Educación Superior.

- Díaz, J. O., Gil, V. P. y Prieto, G. (1999). *Tests informatizados: fundamentos y aplicaciones*. Madrid: Universitas.
- Fernández, J. M. (2010). Las teorías de los tests: teoría clásica y teoría de respuesta a los ítems. *Papeles del Psicólogo*, 31(1), 57-66.
- Garca, J. E. (2016). La introducción de la escala de inteligencia de stanford-binet en el Paraguay/the introduction of the stanford-binet intelligence scale in Paraguay. *Interacciones. Revista de Avances en Psicología*, 2(1), 65-83.  
<https://doi.org/10.24016/2016.v2n1.28>
- Gil, V. P., Adanes, G. P. y Díaz, J. O. (2000). Test informatizados. Fundamentos y aplicaciones. *Psicothema*, 12(2), 321-322.
- Hambleton, R., Swaminathan, H. y Rogers, H. (1985). *Principles and applications of item response theory*. Boston, MA: Kluwer Academic Publisher.
- Hambleton, R. K., Zeal, J. N. y Pieters, J. P. (1991). Computerized adaptive testing: Theory, applications, and standards. *Advances in Educational and Psychological Testing: Theory and Applications*, 2, 341-366.
- Hontangas Beltran, P. M. (2000). Los tests adaptativos informatizados en la frontera del siglo XXI: una revisión. *Metodología de las Ciencias del Comportamiento*, 2(2), 183-216.
- Jodoin, M. G., Zenisky, A. y Hambleton, R. K. (2006). Comparison of the psychometric properties of several computer-based test designs for credentialing exams with multiple purposes. *Applied Measurement in Education*, 19(3), 203-220.  
[https://doi.org/10.1207/s15324818ame1903\\_3](https://doi.org/10.1207/s15324818ame1903_3)
- Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Nueva York, NY: Routledge.
- Luecht, R., Brumeld, T. y Breithaupt, K. (2006). A test let assembly design for adaptive multistage tests. *Applied Measurement in Education*, 19(3), 189-202.  
[https://doi.org/10.1207/s15324818ame1903\\_2](https://doi.org/10.1207/s15324818ame1903_2)
- Muñiz, J. (1997). *Introducción a la teoría de respuesta a los ítems*. Madrid: Pirámide.
- Olea, J. y Ponsoda, V. (1998). Evaluación informatizada en contextos de aprendizaje. *Nuevas Tecnologías para el Aprendizaje*, 3(1), 161-175.
- Pinsach, J. R. y Dieguez, E. D. (1999). Tests adaptativos informatizados: Estructura y desarrollo. En J. Olea, V. Ponsoda y G. Prieto (Eds.), *Tests informatizados: Fundamentos y aplicaciones*, (pp. 127-162). Madrid: Piramide.
- Pressman, R. S. y Troya, J. M. (1988). *Ingeniería del software*. Madrid: McGraw-Hill.
- Prieto, G. y Delgado, A. R. (2003). Análisis de un test mediante el modelo de Rasch. *Psicothema*, 15(1), 94-100.
- Sierra-Matamoros, F. A., Valdelamar-Jiménez, F., Hernandez-Tamayo, F. A. y Sarmiento-García, L. M. (2007). Test adaptativos informatizados. *Avances en Medición*, 5, 157-162.
- Simanca Herrera, F. y Abuchar Porras, A. (2014). *AEI - Algoritmo de Evaluación Inteligente. Virtual Educa Innovación, competitividad y desarrollo*. Lima: Feijóo.
- Sokal, R. R. y Michener C. D. (1958). A statistical method for evaluating systematic relationships [J]. *Univ. Kans. Sci. Bull*, 28, 1409-1438.
- Toledo, G., Mezura-Godoy, C., Cruz-Ramírez, N. y Benítez-Guerrero E. (2013): *Modelo de Evaluación Adaptativa y Personalizada Mediante Razonamiento Probabilista. 8va Conferencia Latinoamericana de Objetos y Tecnologías de Aprendizaje*. Valdivia: Ediciones Chile.

## **Breve Cv de los autores**

### **Roberto Contreras Juárez**

Profesor Investigador en la Facultad de Ciencias de la Computación de la Benemérita Universidad Autónoma de Puebla. Doctor en Ciencias: Matemáticas Aplicadas. Doctorado en Ciencias: Matemáticas Aplicadas. Profesor Investigador de la Facultad de Ciencias de la Computación, impartición de cursos a nivel Licenciatura y Posgrado. Miembro del Cuerpo Académico "Entornos Colaborativos Digitales para el Desarrollo de la Ciencias y la Tecnología". Miembro del Comité Internacional de Evaluadores de las pruebas PIENSE de the College Board Puerto Rico. Miembro del Programa para el Desarrollo Profesional Docente en México. Publicación de varios artículos y capítulos de libro en las áreas de matemáticas, evaluación educativa y matemáticas aplicadas. ORCID ID: <https://orcid.org/0000-0001-3271-6754>. Email: [roberto.contreras@correo.buap.mx](mailto:roberto.contreras@correo.buap.mx)

### **Etelvina Archundia Sierra**

Profesor Investigador en la Facultad de Ciencias de la Computación de la Benemérita Universidad Autónoma de Puebla. Doctor en Tecnologías de la Información y Análisis de Decisiones. Doctorado Tecnologías de la Información y Análisis de Decisiones. Profesor Investigador de la Facultad de Ciencias de la Computación, impartición de cursos a nivel Licenciatura y Posgrado. Miembro del Cuerpo Académico "Entornos Colaborativos Digitales para el Desarrollo de la Ciencias y la Tecnología". Miembro del Programa para el Desarrollo Profesional Docente en México. Publicación de varios artículos arbitrados e indexados, y capítulos de libro en las áreas de evaluación educativa y Tecnología. ORCID ID: <https://orcid.org/0000-0001-9686-5305>. Email: [etearchunn@gmail.com](mailto:etearchunn@gmail.com)

### **Héctor David Ramírez Hernández**

Profesor Investigador en la Facultad de Ciencias de la Computación de la Benemérita Universidad Autónoma de Puebla. Doctor en Ciencias: Matemáticas abstracta y aplicadas. Doctorado en Ciencias: Matemáticas Puras y Aplicadas. Profesor Investigador de la Facultad de Ciencias de la Computación, impartición de cursos a nivel Licenciatura y Posgrado. Miembro del Programa para el Desarrollo Profesional Docente en México. Publicación de varios artículos y capítulos de libro en las áreas de matemáticas, evaluación educativa y matemáticas aplicadas. ORCID ID: <https://orcid.org/0000-0003-3741-4285>. Email: [hector.ramirez@correo.buap.mx](mailto:hector.ramirez@correo.buap.mx)

### **Nelva Betzabel Espinoza Hernández**

Profesor Investigador en la Facultad de Ciencias de la Computación de la Benemérita Universidad Autónoma de Puebla. Maestra en Ciencias: Matemáticas abstracta y aplicadas. Profesor Investigador de la Facultad de Ciencias de la Computación, impartición de cursos a nivel Licenciatura. Miembro del Programa para el Desarrollo Profesional Docente en México. Publicación de varios artículos y capítulos de libro en las áreas de matemáticas, evaluación educativa y matemáticas aplicadas. ORCID ID: <https://orcid.org/0000-0002-5620-2336>. Email: [nespinoza@cs.buap.mx](mailto:nespinoza@cs.buap.mx)

*R. Contreras-Juárez, E. Archundia-Sierra, H. D. Ramírez-Hernández, N. B. Espinoza-Hernández y V. J. Hernández-Moyotl*

**Verónica Judith Hernández Moyotl**

Egresada de la Maestría en Ciencias de la Computación de la Facultad de Ciencias de la Computación de la Benemérita Universidad Autónoma de Puebla. Becaria CONACyT. Dos artículos publicados en el área de Tecnologías. Email: judith\_h21@hotmail.com