

## Desarrollo de la Paep

### [I. INTRODUCCIÓN](#)

### [II. MARCO TEÓRICO](#)

### [III. METODOLOGÍA](#)

### [IV. COMITES Y SUS FUNCIONES](#)

### [V. ÁREAS Y CONTENIDOS DE LA PAEP](#)

### [VI. RESULTADOS OBTENIDOS](#)

### [VII. CONCLUSIONES](#)

## 1. INTRODUCCIÓN

### Antecedentes

El desarrollo de la Prueba de Admisión a Estudios de Posgrado (PAEP) se inició en Marzo de 1991 como un proyecto del Sistema Tecnológico de Monterrey, realizado en el Campus Eugenio Garza Sada. Participaron profesores de todas las rectorías del Sistema y fue apoyado desde la perspectiva psicométrica por el director técnico del College Board en Puerto Rico. La necesidad de que el Sistema contara con una prueba actualizada diseñada específicamente para nuestros alumnos fue el motivo que ha impulsado este proyecto lográndose un desarrollo de recursos humanos internos en este trabajo. Este instrumento se realizó con una exigencia técnica obteniendo un alto nivel de validez y confiabilidad. La prueba fué diseñada para evaluar aptitudes, habilidades y conocimientos del candidato a ingresar a estudios de posgrado, en cinco áreas.

### Objetivos

Los objetivos que se tuvieron fueron los siguientes:

- Elaborar una prueba que facilitara el proceso de admisión y que tuviera la capacidad de observar las diferencias individuales entre los candidatos a los diferentes programas.
- Capacitar a un equipo de trabajo en las técnicas del campo de la evaluación con el fin de desarrollar la prueba y posteriormente apoyar proyectos de evaluación.

### Importancia

La prueba PAEP fue diseñada con una estructura de tres componentes para evaluar aptitudes, habilidades y conocimientos del candidato a ingresar a estudios de posgrado; estos tres componentes se dividieron en cinco áreas. Un aspecto importante de esta prueba es que el instrumento mide una nueva área de habilidad cognitiva, en donde el candidato debe demostrar su habilidad para resolver problemas desde un punto de vista de razonamiento cognitivo. Otro aspecto importante es la evaluación de comprensión de lectura con un análisis crítico y no sólo análisis tradicional. Para el análisis crítico de la lectura se le presentan al candidato dos lecturas de donde tendrá que contestar a reactivos que impliquen comparar ambos textos. Los reactivos que miden este tipo de comprensión se dan en las secciones Verbal y de Inglés. Se incluyó un tercer aspecto en el diseño de la prueba, que fue el de valorar una composición que el candidato hace sobre un tema específico; dicha composición se valora con una metodología objetiva (holística) en donde participan dos o tres lectores por composición. Actualmente se tienen tres formas paralelas de la prueba: A, B y C. La equivalencia de las tres formas se logró diseñando la segunda y tercera forma con reactivos comunes a la primera forma.

## II. MARCO TEÓRICO

El propósito de la evaluación es conocer los niveles de conocimiento, habilidad, o cualquier otro aspecto de la persona. Una vez que se determina lo que se desea evaluar, se toman decisiones sobre el tipo de información que se necesita y la manera en cómo se estructura. Existen diversas maneras para medir comportamiento, siendo una de ellas las pruebas. Asimismo, existen muchos tipos de pruebas, como son las de conocimientos, las de aptitudes o las de personalidad.

En cualquier modalidad, es necesario considerar que las pruebas son un instrumento que debe utilizar el evaluador como parte de un conjunto de medidas para tomar decisiones. Y en el caso de pruebas de admisión, es importante considerar lo anterior, ya que para tomar decisiones sobre la continuación del desarrollo académico de una persona, un sólo factor limitaría el criterio de aceptación o rechazo.

Las pruebas se pueden desarrollar en forma de opción múltiple, abiertas, orales, escritas, etc. La forma que se decida adoptar dependerá de las necesidades para la que una prueba es diseñada. Todas las pruebas de esta índole tienen limitaciones, ya que este tipo de medición nunca será tan exacta como una medición en ciencias exactas. Se debe también considerar el error estándar de medición, que indica el rango en donde el examinado puede variar su puntuación.

Los resultados de una prueba se nos dan en forma numérica, por lo que deben ser interpretados de alguna forma. El diseño de pruebas puede ser de referencia normativa o de referencia de criterio. En las pruebas de referencia normativa podemos referir los puntajes a una norma, a una población específica para poder comparar los resultados. Las pruebas de referencia de criterio tienen que ver con diferentes grados de un tipo de comportamiento del sujeto para compararse consigo mismo; en este caso, una actividad se divide en componentes para observar los logros individuales. En el caso del desarrollo de la PAEP se optó por el tipo de prueba normalizada con el fin de poder tomar decisiones a nivel institucional.

Existe una metodología respecto a la forma de desarrollar nuevas pruebas, como son las consideraciones que se deben tener en cuenta al elaborar los reactivos, así como los diferentes distractores en el caso de preguntas de opción múltiple. De la misma manera, hay una gran variedad de conceptos y análisis estadísticos que nos permiten observar el comportamiento de los reactivos que se incluirán en una prueba. Todos estos aspectos son importantes para elaborar una prueba formal con altos índices de confiabilidad y validez.

La PAEP fue elaborada tomando en consideración el Modelo IRT (Item Response Theory), teoría estadística que consiste en varios modelos que expresan la probabilidad de observar una respuesta particular a un reactivo en función de ciertas características del mismo; también nos indica el nivel de habilidad del sujeto. El modelo utilizado fue uno de tres parámetros que se aplica a preguntas de opción múltiple. Los parámetros representan: a) la capacidad del reactivo para discriminar entre los niveles de habilidad, es decir el rango de probabilidad de que una respuesta correcta de un reactivo cambie como función del nivel de habilidad del sujeto; b) el índice de la dificultad del reactivo; y c) la probabilidad de que un sujeto de muy baja habilidad pueda contestar un reactivo correctamente, es decir adivinando.

Con el fin de que las tres versiones elaboradas de la PAEP fueran paralelas o equivalentes, se utilizó el método de Tucker que consiste en utilizar reactivos comunes a las tres pruebas. Este modelo permite que los puntajes en las diferentes versiones sean comparables para que ningún candidato se vea afectado por la forma que presente. Este procedimiento de conversión se llama 'calibración de puntajes'.

### **III. METODOLOGÍA**

La PAEP está diseñada con una estructura tridimensional con los siguientes componentes: un primer aspecto de aptitud académica (habilidades de razonamiento verbal y matemático); un segundo aspecto de habilidad cognitiva y una tercera sección de conocimiento académico (redacción e inglés como herramienta de trabajo). Estos tres componentes están divididos en la prueba en las siguientes cinco áreas: Razonamiento Verbal, Razonamiento Matemático, Habilidad Cognitiva, Redacción e Inglés. En total, la PAEP tiene 170 reactivos divididos de la siguiente manera: Razonamiento Verbal: 52, Razonamiento Matemático: 40, Habilidad Cognitiva: 28, Redacción: 25 e Inglés : 25. El diseño del número de reactivos para cada sección fue dado de acuerdo a los tres componentes de la prueba, teniendo como parte principal el componente de aptitud (verbal y matemático).

Además, la PAEP incluye el desarrollo de una composición que se valora objetivamente por un Comité de Lectores. Existen diferentes temas para cada aplicación. Para evaluar objetivamente las composiciones, se utilizó una metodología especial en donde cada composición es leída por al menos dos lectores independientes; la valoración se da en una escala de 6 puntos, sumándose las puntuaciones de los dos lectores a una escala de 12 puntos. El lector deberá evaluar básicamente en tres rangos (excelente, regular y malo), teniendo dos puntos en cada uno de los rangos. En caso de que los dos lectores no coincidan al dar la evaluación de rango, la composición es evaluada por un tercer o cuarto lector.

Los puntajes finales de la prueba se ofrecen en una escala de 200-800. Al mismo tiempo la prueba arroja cinco puntajes parciales, uno para cada una de las áreas; éstos se dan en una escala de 20-80, pudiendo de esta manera observar el comportamiento de los candidatos en las diferentes secciones. El desarrollo de la prueba se realizó de acuerdo al siguiente procedimiento:

- Detección de necesidades a evaluar en los candidatos
- Formación de comités, integrando a profesionistas de los diferentes Campus
- Entrenamiento en redacción de reactivos
- Diseño, elaboración y revisión de reactivos
- Entrenamiento a los examinadores para la revisión de reactivos
- Elaboración y aplicación de 5 pre-pruebas para probar 500 reactivos
- Elaboración de la Prueba Oficial, dividida en 5 áreas: verbal, matemáticas, habilidad cognitiva, redacción e inglés
- Elaboración de una guía de examen para el candidato
- Elaboración de un instructivo de aplicación y revisión de la prueba para los examinadores
- Entrenamiento de lectores para valorar composiciones del área de redacción
- Aplicación de la primera prueba oficial (Forma A) a los candidatos a estudios de posgrado
- Normalización y estandarización de la Forma A
- Entrega de resultados a Directores de Escolar y Directores de Programas de Graduados
- Aplicación de la segunda prueba oficial (Forma B) a los candidatos a estudios de posgrado
- Normalización y estandarización de la Forma B
- Estudios de calibración entre las dos pruebas (Formas A y B)
- Elaboración de la tercera prueba oficial (Forma C)
- Normalización y estandarización de la Forma C
- Estudios de calibración entre las tres versiones (Formas A, B y C)

#### **IV. COMITES Y SUS FUNCIONES**

- **Comité Ejecutivo:**  
Responsables directos del proyecto: (1 consultor técnico, 1 coordinador de estadística y 1 coordinador del proyecto).
- **Comité Consultivo:**  
Directivos responsables de tomar decisiones, representando los tres programas de graduados y los Campus con mayor número de alumnos
- **Comité Examinador:**  
Cinco expertos en cada una de las cinco áreas que evalúa la prueba: verbal, ciencias-matemáticas, habilidad cognitiva, redacción e inglés.  
Responsables de elaborar las especificaciones de cada área y revisar los reactivos elaborados por los redactores.
- **Comité de Redactores:**  
Cinco profesores de cada una de las áreas de la prueba, responsables de diseñar los reactivos

- Comité de Lectores:  
Profesores capacitados para valorar las composiciones que cada candidato debe elaborar en el área de redacción.

## V. ÁREAS Y CONTENIDOS DE LA PAEP

- Verbal  
Antónimos, Completar Oraciones, Analogías y Comprensión de Textos (análisis tradicional y análisis crítico). Comité Consultivo:  
Directivos responsables de tomar decisiones, representando los tres programas de graduados y los Campus con mayor número de alumnos
- Matemáticas-Ciencias  
Aritmética, Geometría, Álgebra, Cálculo, Probabilidad, Física, Química y Biología.
- Habilidad Cognitiva  
Secuencias y Relaciones Lógicas, Transformaciones y Consideraciones Lógicas.
- Redacción  
Estructura, Léxico, Madurez Sintáctica y Composición
- Inglés (como herramienta de trabajo)  
Estructura, Palabras Funcionales y Comprensión de Texto.



## VI. RESULTADOS OBTENIDOS

Se realizaron análisis estadísticos para obtener la normalización y estandarización de la prueba. La primera prueba (Forma A) se aplicó a una muestra de 459 candidatos en el mes de julio de 1992. La segunda prueba (Forma B) se aplicó a una muestra de 514 candidatos en el mes de diciembre del mismo año y la tercera prueba (Forma C) se aplicó a una muestra de 587 candidatos en el mes de julio de 1993.

Los análisis estadísticos mostraron el índice de dificultad de la prueba así como los índices de correlación biserial con el fin de determinar la capacidad de la prueba para diferenciar a los candidatos. Las siguientes tablas muestran los resúmenes descriptivo e instrumental de las tres versiones que se han desarrollado de la prueba.

### RESUMEN DESCRIPTIVO

Forma	Muestra	Reactivos	Mínimo	Máximo	Media Arit.	Mediana	Des. Est.
A	459	170	34	134	84.264	83	18.937
B	514	170	27	146	89.696	90	22.463
C	587	170	18	140	86.165	85	19.992

En esta tabla podemos observar el número de reactivos mínimos y máximos para cada una de las versiones de la prueba, en donde la Forma A resulta ser la más difícil y la B la más fácil. Las desviaciones estándares muestran similitud entre las tres, así como el resto de los parámetros.

## RESUMEN INSTRUMENTAL

Forma	Confiabilidad	Error Est. Med.	Dific. Observ.	Correl Biserial	Des.Est.rb
A	0.912	5.629	0.496	0.335	0.124
B	0.936	5.678	0.528	0.386	0.126
C	0.920	5.654	0.507	0.348	0.100

En esta tabla observamos el alto índice de confiabilidad de .9 en las tres versiones con iguales errores estándares de medición. La dificultad observada nos indica un 50% de dificultad en las tres formas y un índice de discriminación alrededor de .3, indicando una capacidad para discriminar entre los alumnos de alta habilidad y baja habilidad.

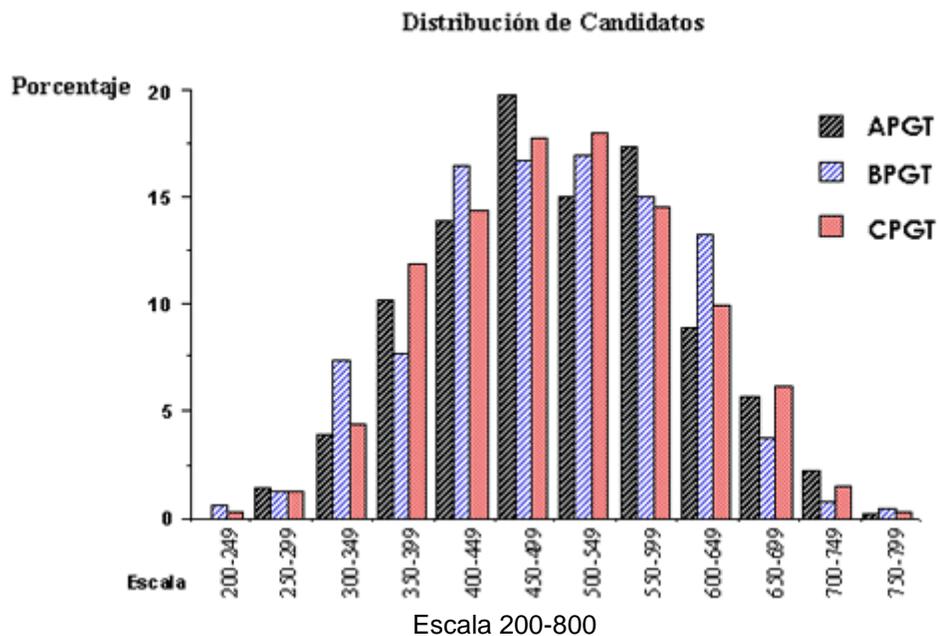
Con el propósito de establecer la equivalencia entre las versiones, se realizó un procedimiento de equivalencia lineal mediante el enfoque de Tucker con un conjunto de reactivos comunes a ambas pruebas. El proceso de igualamiento de los puntajes burdos de la Forma A a los puntajes burdos de la Forma B se logra a través de estimaciones de las medias y varianzas de los puntajes burdos de ambas formas. Posteriormente se estimó la conversión de los puntajes de la Forma C a la Forma A. Estos parámetros permitieron anclar los puntajes de las formas B y C a la Forma A, misma que queda como la prueba normativa. Por lo tanto, los puntajes obtenidos mediante este procedimiento representan los puntajes que un sujeto al que se le administra la Forma B o C obtendría si se le hubiese administrado la Forma A. De esta manera se pueden comparar los puntajes en las tres versiones.

### VII. CONCLUSIONES

Los resultados generales obtenidos hasta hoy cumplen con los objetivos planeados y nos indican una alta confiabilidad de la prueba. La confiabilidad de las tres formas de la prueba está entre .91 y .94, indicando un alto nivel. La dificultad de la prueba se encuentra en 13 puntos promedio en la escala delta de 1-20.

En cuanto al paralelismo de las tres formas se ha encontrado que éstas tienen el mismo nivel de dificultad así como un buen índice de correlación biserial. Ésto permite que se puedan usar indistintamente sin afectar al candidato.

DISTRIBUCIÓN DE PUNTAJES DE LA PAEP  
(Porcentaje de candidatos)



Los puntajes de los candidatos mostraron una distribución normal en las tres formas de la prueba. Con los resultados de la PAEP se han hecho otro tipo de investigaciones, tanto en la estructura interna de la prueba como en análisis predictivos sobre la actuación del alumno en sus estudios de posgrado.