

VALIDACION DE LA PRUEBA DE J. C. RAVEN MATRICES PROGRESIVAS Y DE LA PRUEBA ACE PARA ESTUDIANTES DE PRIMER AÑO UNIVERSITARIO

FORMA '47

Fanny Gómez de Pedraza

Colaboradoras:

Diva M. de Rincón

Gladys Montealegre

EL PROBLEMA

El presente estudio es apenas una parte del plan para elaborar un equipo de pruebas psicológicas que permita una selección eficiente de estudiantes que aspiren al primer año de las distintas Facultades y Departamentos de la Universidad Nacional.

Es de todos conocido el grave problema que se le presenta cada año a esta Universidad por el exceso de demanda a sus diferentes facultades, y la falta de cupos en algunas de ellas.

Otro problema anual es el referente al número de fracasos entre los estudiantes admitidos. No exageramos al decir que aproximadamente un 50 por ciento de los estudiantes, no obtiene éxito en los estudios.

La razón, no la sabemos. Pero podemos inferir que si se usara otro sistema de selección diferente del actualmente empleado, exámenes de conocimientos y entrevistas personales posiblemente no sería tan elevado el número de estudiantes que no puede terminar sus estudios.

No es nuestra idea presentar un remedio infalible, sino más bien informar sobre algunas de las pruebas que se han usado en otros países con fines de selección universitaria, y estudiar la manera de aumentar el índice de éxitos en los estudiantes de esta universidad.

PROPOSITO:

El propósito de este estudio es:

- a) La validación de la Prueba del Consejo Americano de Educación para Estudiantes de Primer Año Universitario -Forma '47, parte cuantitativa.
- b) La validación de la prueba de J. C. Raven-Matrices Progresivas.

En ambas pruebas se usarán como criterio de eficiencia las calificaciones finales correspondientes al primer semestre en matemáticas, en las facultades de Ingeniería y Arquitectura de la Universidad Nacional, Año de 1960.

REVISION DE ALGUNOS CONCEPTOS SOBRE VALIDEZ

Qué es Validez?

La validez de una prueba indica lo que sus puntajes miden y predicen. Generalmente el grado de validez se expresa en forma de coeficiente de correlación. Cuando se usa para predecir la ejecución en cierta situación vital, la validez se describe en términos de correlación entre la prueba predictora y cierta media de eficiencia en esa situación.

Guilford (4, pg 398) dice refiriéndose a la validez:

"... Un puntaje es válido para predecir algo con lo que se relaciona, donde "Algo" no incluye el puntaje en sí, ya que la auto-predicción tiene que ver con la confiabilidad. Cuando dos pruebas indican una intercorrelación mayor que cero, lo "que" uno mide es idéntico, a lo menos en parte, a lo "que" el otro mide... la variación del factor común, es pues, la base de la validez".

Evaluación del Criterio de Eficiencia. Medida Criterio o Calificación.

Según Nunnally (5), una calificación es una medida en sí. Si un profesor en determinada clase da a un estudiante una nota de 4 o 5, se puede decir que el estudiante ha sido evaluado, calificado o medido en su conocimiento sobre el curso. Cuando un patrón asciende a su empleo, este ha sido calificado o evaluado en cuanto a su trabajo. Estos ejemplos y muchos otros, son considerados criterios de eficiencia, medida criterio o calificación.

Un examen de clase es bueno si es fiel y válido. Cualquier cosa que se haga para aumentar la confiabilidad de una prueba vale la pena intentarla. Algunas de las técnicas que se usan para mejorar la confiabilidad de los exámenes son los siguientes:

- a) Calificación del examen dado por dos o más profesores.
- b) Calificación doble dada por el

mismo profesor, obtenida la segunda de ellas sin tener en cuenta la primera.

- c) Presentación de los exámenes objetivos cuya valoración se hace con una clave pre-establecida.

La validez de un examen de clase se refiere a la representación del contexto en otras palabras, que la prueba sea una muestra completa de las preguntas importantes y pertinentes a la materia.

Thorndike (6, pg. 214) afirma:

"La Pertinencia es un requisito fundamental de la medida criterio. Tanto como sea posible, es importante que toda la varianza sistemática de la medida criterio sea varianza pertinente. Si la medida criterio posee una cantidad considerable de varianza no casual y que es impertinente o, todavía peor, negativamente relacionada con último fin, es posible que aparentemente los exitosos procedimientos de selección se basen enteramente sobre varianza impertinente y no tengan validez real para el fin del trabajo. Esto es lo que pasa en situaciones de entrenamiento altamente académico, cuando se usan como criterio de éxito en algunas situaciones no académicas".

La discusión anterior concerniente a la validez de la medida de criterio está muy relacionada con el problema de este estudio. La medida criterio que usamos o sea las calificaciones de matemáticas, sin duda contiene mucha varianza impertinente. No sabemos si esta varianza es en algún modo sistemática o no.

A pesar de que las calificaciones de materias tomadas como medidas criterio pueden contener varianza impertinente o no relacionadas al asunto, se usan en este estudio porque:

1. fueron fáciles,
2. son más o menos objetivas,
3. son pre-requisitos para el último fin de éxito académico.

Evaluación de las Pruebas Predictoras o Predictores.

La mayor parte de las pruebas que usan los psicólogos son predictoras. Todos los exámenes de aptitudes y los de personalidad son de este tipo. Su utilidad se extiende a campos tan diversos como la selección y clasificación de individuos en las escuelas, universidades, fuerzas armadas, agencias del gobierno y en la industria.

Las pruebas predictoras tienen como fin pronosticar o ver de antemano cómo obrará un individuo ante una situación, o ejecutará una tarea; o qué calificación recibirá en situaciones importantes.

De acuerdo con Nunnally (5, pg. 63):

“...el término predicción se usa en el sentido amplio de expresar relación funcional de toda clase. Un test se puede usar para “predecir” lo que la gente hizo en el pasado, de la misma manera que un detective reúne pruebas para “predecir” quien cometió un crimen. También los tests se usan para “predecir” cómo se comportó la gente siendo niño. Los tests a menudo se usan para predecir una condición actual más que para predecir, en su sentido estricto, o para ver de antemano el comportamiento futuro. Un ejemplo típico es una prueba para diagnosticar lesión cerebral. El propósito es “predecir” la lesión cerebral y la prueba de la “predicción” saldrá de minuciosas investigaciones fisiológicas”.

La Utilidad de las Pruebas Predictoras

Una pregunta muy común en esta clase de problemas es “¿Qué tan buena es la prueba x?”; y la respuesta es que todo depende de las circunstancias. Tomemos este ejemplo:

El número de estudiantes que solicita admisión a un establecimiento

generalmente es mayor al número que va a ser seleccionado.

$$\frac{\text{N}^\circ \text{ de seleccionados}}{\text{N}^\circ \text{ de solicitantes}} = \text{razón de selección.}$$

De los seleccionados, algunos triunfarán y los otros fracasarán.

$$\frac{\text{N}^\circ \text{ de éxitos}}{\text{N}^\circ \text{ de seleccionados}} = \text{razón de éxito}$$

La relación entre estas dos razones es muy importante.

Si la razón de éxito es muy grande no hay necesidad de mejorar los procedimientos de selección. Si la razón de selección es grande, no hay nada que pueda hacer una prueba para mejorar la selección.

Una prueba predictoras es válida para el proceso de selección en la proporción en que disminuyen los fracasos académicos, o en la medida en que permita una enseñanza superior o más especializada.

Es muy importante tener presente que un instrumento es válido para un uso particular y no en sentido general. El fin es tratar de determinar qué tan bueno es un instrumento para predecir cierto criterio o medidas criterio, más que tratar de probar que el instrumento es en general bueno.

MÉTODOS Y PROCEDIMIENTOS

Descripción de los Instrumentos

La Prueba J. C. Raven -Matricens Progresivas (De ahora en adelante llamada RAVEN).

Esta prueba mide el “factor g” o factor general inteligencia. Contiene 60 problemas seriados en orden de dificultad. Las primeras series plantean variados problemas de percepción estructurada y de educación de relaciones y las últimas de educación de correlatos. No tiene tiempo límite.

Prueba del Consejo Psicológico de Educación para estudiantes de Primer Año Universitario Forma 47.

(De ahora en adelante llamada ACE) Esta prueba contiene 80 puntos divididos en tres partes:

Problemas de aritmética

Analogías de figuras

Series Numéricas

Cada parte tiene tiempo límite y algunos problemas de práctica.

Guilford (1, pg. 297) refiriéndose al ACE, afirma:

“La parte cuantitativa es un conglomerado factorial. Además de medir la habilidad numérica mide, tal vez, tres clases de razonamiento, como también otros factores en menor grado. Probablemente selecciona estudiantes con buenas habilidades de razonamiento”.

Examen Semestral en Matemáticas de las Facultades de Ingeniería y Arquitectura.

(De ahora en adelante llamado MAT). La nota final de los exámenes se basó sobre las previas semestrales y un examen final tipo ensayo con 5 problemas por resolver. Las notas van de 0.00 a 5.00. Para facilitar los cálculos redujimos los puntajes de 00 a 50 y suprimimos el punto.

Sujetos y Procedimientos

Los sujetos de experimentación fueron estudiantes que cursan su primer año en las facultades de Ingeniería y Arquitectura de la Universidad Nacional y en 1960, quienes fueron citados durante horas de clases en grupos de 30 a 40 y a quienes se les dieron las instrucciones correspondientes a cada una de las pruebas, ACE y RAVEN, de acuerdo con lo expuesto en los respectivos manuales. El número de sujetos se redujo a 74 en Arquitectura y a 84 en Ingeniería, porque solamente estos tenían todos los datos completos. El RAVEN fue administrado a los estudiantes

de Ingeniería como parte de los exámenes de ingreso, aunque los resultados no se tuvieron en cuenta para este fin.

Valoración de las pruebas

Se hizo de acuerdo con lo expuesto en los respectivos manuales: Puntaje para ACE = N° de respuestas correctas. Puntaje para el RAVEN = N° de respuestas correctas.

PROCEDIMIENTOS

ESTADÍSTICOS

Las tres variables, ACE, RAVEN, y MAT., fueron sometidas al siguiente análisis estadístico: Promedios, medianas, desviaciones estandar y medidas de asimetría. Para el cálculo de la asimetría usamos la fórmula (2, pg. 99):

$$As = \frac{(P90 + P10) - P50}{2}$$

A continuación calculamos el error estandar de As y comprobamos si las distribuciones eran asimétricas o no, por medio de la razón crítica.

Calculamos el coeficiente de correlación de Pearson entre las siguientes variables: MAT-RAVEN, MAT-ACE, ACE-RAVEN, para Arquitectura y MAT-RAVEN, MAT-ACE-RAVEN para Ingeniería.

Posteriormente sometimos las correlaciones entre las anteriores variables a una correlación múltiple, separadamente para Ingeniería y Arquitectura.

Creímos importante el cálculo de la correlación múltiple, ya que nuestro propósito no es solamente validar prueba por prueba, sino buscar el valor predictivo de un equipo de pruebas de selección universitaria. En las correlaciones múltiples se hizo necesaria una corrección por ser el número de sujetos inferior a 100, en ambos grupos.

ANALISIS DE LOS RESULTADOS

Descripción de los grupos examinados

La descripción de los grupos —Arquitectura e Ingeniería— en las tres pruebas: ACE, RAVEN y MAT, se presenta en las tablas I y II. En cada una de ellas se informa el puntaje máximo en cada prueba, N° de sujetos, promedios (M), medianas (Mdn), y desviaciones estandar (DS)

TABLA I

Promedios, Medianas, Desviaciones Estandar, obtenidos de las distribuciones del ACE, RAVEN y MAT en 74 estudiantes de Arquitectura en la U. N., 1960.

PRUEBA	Nº	Puntaje	M	Mdn	Ds
RAVEN	74	60	50.54	51.50	4.68
ACE	74	80	32.60	34.05	10.80
MAT	74	50	21.32	19.16	9.50

TABLA II

Promedios, Medianas, Desviaciones Estandar, obtenidos de las distribuciones del ACE, RAVEN y MAT en 84 estudiantes de Ingeniería, U. N. 1960.

PRUEBA	Nº	Puntaje	M	Mdn	Ds
RAVEN	84	60	50.31	50.86	5.76
ACE	84	80	39.38	39.77	9.20
MAT	84	50	29.91	30.29	8.65

Los niveles de importancia para probar la normalidad de las distribuciones en nuestros ejemplos fueron verificadas, pero de ellos no se da informe ya que no fueron significativas, a excepción de la distribución de MAT en Arquitectura, la cual resultó algo positivamente asimétrica, Razón Crítica = 199 que es significativa al nivel .05 de confianza. Estos resultados nos garantizan el uso apropiado que hicimos del coeficiente

de correlación de Pearson como técnica de validez ya que según comenta Guilford (3 pg. 170):

“Algunos escritores sugieren que solamente cuando ambas distribuciones son normales o casi normales se satisfacen por completo los requisitos para computar el Pearson r. En la práctica probablemente nadie insiste en que las distribuciones sean normales”.

Muchos autores discuten la normalidad de las distribuciones de calificaciones y de la continuidad de la escala, pero en el sistema de notas que se usa en la U. N., en el que se califica de 0.00 a 5.00, es propio asumir que generalmente las distribuciones son normales y que la variable es continuamente medible. Esta es otra razón que nos permite computar el Pearson r.

Datos Estadísticos sobre validez del ACE y RAVEN.

La Tabla III contiene los coeficientes de correlación Pearson (r), obtenidos entre ACE-MAT y RAVEN-MAT en Arquitectura e Ingeniería. La Tabla III incluye además, N° de sujetos en cada grupo, los grados de libertad y los valores correspondientes a los niveles .05 y .01 de confianza.

TABLA III

Coefficientes de Correlación Pearson obtenidos entre el ACE y MAT y RAVEN y MAT, en 74 estudiantes de Arquitectura y 84 estudiantes de Ingeniería U. N., 1960.

PRUEBAS	Nº	r	dl	Nivel	Nivel
				.05	.01
Arquitectura					
ACE-MAT	74	.44	74.2	r.23	r.30
RAVEN-MAT	74	.33	74.2	r.23	r.30
Ingeniería					
ACE-MAT	84	.19	84.2	r.22	r.28
RAVEN-MAT	84	.25	84.2	r.22	r.28

PRUEBAS	Nº	r	R	cR
Arquitectura				
ACE-RAVEN	74	.37		
MAT-ACE-RAV.	74		.48	.45
Ingeniería				
ACE-RAVEN	84	.19		
MAT-ACE-RAV.	84		.29	.25

El análisis de la Tabla III nos indica que hay una correlación positiva y muy significativa entre ACE-MAT en Arquitectura. Un coeficiente de validez de .44 es muy aceptable ya sea que se use en selección u orientación vocacional. (Coeficientes de validez entre .40 y .60 son muy confiables). La correlación entre estas dos variables en Ingeniería, es $r = .19$; correlación positiva pero no significativa. Una razón que se nos ocurre para explicar esta discrepancia entre $r = .44$ y $r = .19$ es que la medida-criterio en Arquitectura es más fiel y válida que en Ingeniería. En cuanto a los coeficientes de validez del RAVEN $r = .33$ en Arquitectura y $r = .25$ en Ingeniería, aunque positivos y significativos son un poco bajos como índices de validez. En general, los coeficientes de correlación son más bajos en Ingeniería que en Arquitectura. Esta diferencia nos hace presente la necesidad de una investigación adicional en donde nos sea posible investigar más íntimamente la confiabilidad de la medida-criterio.

Datos sobre las correlaciones múltiples.

En la Tabla IV presentamos los coeficientes de Correlación de Pearson entre ACE-RAVEN para ambos grupos, dato necesario para el cálculo de las correlaciones múltiples. También aparecen los coeficientes de correlación múltiple (R) para Ingeniería y Arquitectura entre las variables: MAT-ACE, RAVEN y la correlación de las correlaciones múltiples (CR).

TABLA IV

Coeficientes de correlación Pearson r , Coeficientes de Correlación Múltiple y corrección de Correlación Múltiples Obtenidos de las Variables MAT-RAVEN, ACE, en 74 estudiantes de Arquitectura y 84 Estudiantes

Analizando los resultados en Arquitectura vemos que $cR = .45$; por lo tanto deducimos que el RAVEN no aporta nada al poder predictivo del ACE, que como ya vimos es de $r = .44$. En Ingeniería, $cR = .35$, por lo tanto el ACE no aumenta el coeficiente de validez del RAVEN, $r = .25$.

Revisando las correlaciones entre ACE-RAVEN en Ingeniería y Arquitectura encontramos $r = .19$ y $r = .37$ respectivamente. Estos resultados indican que hasta en un tanto por ciento, el ACE y el RAVEN están midiendo un mismo factor, que aparece también en MAT. El resto de la varianza común en MAT tiene que medirse con otras pruebas que contengan otros factores no mesurables en el ACE o en el RAVEN.

En lo concerniente a los cursos de matemáticas parece que usar juntos el ACE y el RAVEN no aumenta el poder de predicción del conjunto, de tal suerte que el problema resta en decidir cual de los dos, ACE o RAVEN, es mejor predictor de las calificaciones en matemáticas.

Con el fin de resolver el interrogante anterior, buscamos si la diferencia entre ACE y el RAVEN era significativa o no. En Arquitectura la diferencia fue muy significativa al nivel .01 de confianza con una razón crítica de 4.64, resultando favorecido el ACE. En Ingeniería la diferencia fue significativa al nivel .05 con una razón crítica de 2.40, siendo el favorecido en esta ocasión el RA-

VEN. Por la magnitud de los resultados en Arquitectura tal vez no se deba incluir en el índice de los cursos de matemáticas de otras facultades.

RESUMEN Y CONCLUSIONES

El propósito de este estudio fue el verificar la prueba ACE y para ello se utilizaron los datos de la prueba de Matemáticas de la Facultad de Ingeniería y Arquitectura de la Universidad Nacional año 1950. Como resultado de los análisis se concluye que la prueba ACE es un buen predictor de las calificaciones en Matemáticas de los estudiantes de Ingeniería y Arquitectura de la Universidad Nacional año 1950. Como resultado de los análisis se concluye que la prueba ACE es un buen predictor de las calificaciones en Matemáticas de los estudiantes de Ingeniería y Arquitectura de la Universidad Nacional año 1950. Como resultado de los análisis se concluye que la prueba ACE es un buen predictor de las calificaciones en Matemáticas de los estudiantes de Ingeniería y Arquitectura de la Universidad Nacional año 1950.

REFERENCIAS

1. Burso, O. K. (ED). **The third mental measurements yearbook.** New Bruswick, N. J.: Rutgers University Press, 1949.
2. Garret, H. E. **Statistics in Psychology and Education.** New York: Longmans Green, 1953.
3. Guilford, J. P. **Fundamental Statistics in Psychology and Education.** New York: Mc Graw Hill, 1950.
4. Guilford, J. P. **Psychometric methods.** New York: Mc Graw Hill, 1959
5. Nunnally, J. C. **Tests and Measurements.** New York: Mc Graw Hill, 1959
6. Thorndike, R. L. **Personnel selection.** New York: Wiley 1949.