

ANÁLISIS DE LA DIFICULTAD DE LOS ÍTEMES DE LA PRUEBA DE APTITUD ACADÉMICA Y SU EFECTO EN VALIDEZ PREDICTIVA, SEGUN DIFERENTES VARIABLES CRITERIO

Ma. Isabel González de Wong.
Olga de León Páez.

INTRODUCCION

La prueba de Aptitud Académica (P.A.A.) es una prueba específica del área académica que mediante evaluaciones de razonamiento en las áreas verbal y matemática, mide la aptitud del candidato para seguir, con buenas probabilidades de éxito, estudios universitarios. La Universidad de Costa Rica la emplea en su sistema de selección de candidatos a primer ingreso. Para esto combina los puntajes logrados en la P.A.A. con un promedio de calificaciones escolares y obtiene los puntajes de admisión.

Todo instrumento de evaluación que mida lo que pretende medir puede considerarse válido (Van Dalen, 1978). Un test debe tener varios tipos de validez de los cuales nos ocupa en este caso la validez predictiva o de pronóstico. Para evaluar la exactitud del instrumento se requiere saber qué rasgo se desea pronosticar después de cierto período. Se establece para su cálculo un coeficiente de validez que indica la relación entre los puntajes de la prueba y los puntajes en una variable de criterio.

Así, la validez predictiva de la P.A.A., es su capacidad de pronosticar el rendimiento académico después de una experiencia universitaria dada.

Los ítems de la P.A.A. (reactivos, problemas o preguntas) son continuamente sometidos, a la luz de sus objetivos, a análisis cualitativos en lo referente a contenido, forma, procesos cognoscitivos que se requieren para su solución. También se les aplican análisis cuantitativos en lo referente a sus aportes a la confiabilidad y validez de la prueba.

La confiabilidad de la prueba es fundamental para la interpretación de sus resultados. Un test con baja confiabilidad está cargado de errores y no ofrece seguridad en sus resultados.

La confiabilidad alude a la exactitud o consistencia de las puntuaciones. Indica hasta qué punto pueden atribuirse a errores de medida, las diferencias en las puntuaciones del test. Toda medida de confiabilidad denota la proporción de la variancia del test que no es variancia de error (Anastasi, 1972).

La confiabilidad es la proporción entre la variancia verdadera y la obtenida a partir de los datos, así, cuanto menor sea el error, mayor será la confiabilidad y viceversa (Kerlinger, 1979).

Para la P.A.A. la confiabilidad se calcula mediante un coeficiente α , (alfa de Cromback), porque no es posible, dada sus características y fines, hacerlo por métodos de test-re-test. La confiabilidad por el coeficiente α , parte del supuesto de la homogeneidad de los ítems. La participación de todos los ítems de la P.A.A. en la medición de un mismo rasgo (aptitud académica), produce intercorrelaciones entre ellos, que permiten obtener una buena confiabilidad de la prueba. El test debe ser homogéneo en su contenido; en consecuencia sus ítems se deben correlacionar sustancialmente entre sí (Nunnally, 1970).

Pero estas intercorrelaciones entre los ítems afectan las decisiones referentes a su dificultad, como se verá más adelante.

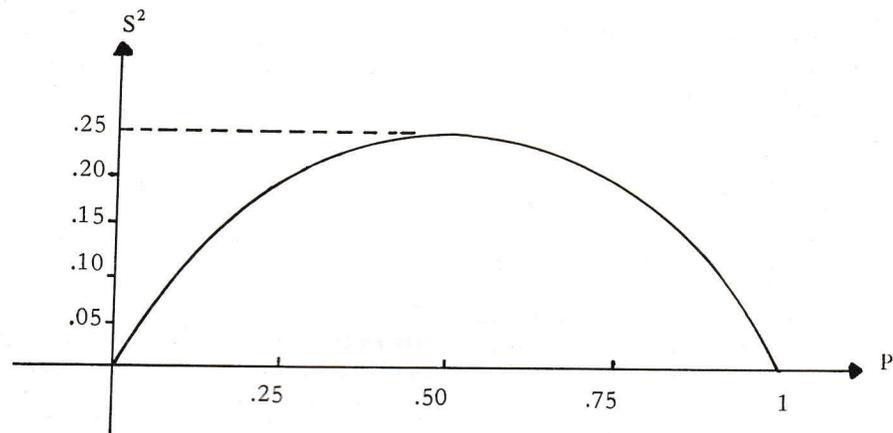
La dificultad de un ítem se basa en el porcentaje de candidatos que lo responden correctamente. Cuando un ítem resulta tan fácil que casi todos lo responden, o tan difícil que casi nadie lo acierta, no informa sobre las diferencias individuales y por eso no contribuye al proceso de selección para el que fue creada la prueba. Además, ítems de dificultad extrema no pueden tener una correlación perfecta con otros ítems de dificultad media, lo

que hace que aquéllos afecten el coeficiente de confiabilidad del test.

Si la dificultad de un ítem dicotómico viene dada por la proporción (p) de respuesta correcta, entonces el número $q = (1-p)$ da la proporción de

individuos que fallaron el ítem y por lo tanto $p+q = 1$.

La distribución hipotética de probabilidad de acierto de un ítem de dificultad p , tiene promedio p y varianza $p q$.



El gráfico adjunto ilustra la varianza ($p q$) de un ítem para varias frecuencias de respuesta correcta (p).

Allí vemos que cuando la dificultad del ítem es promedio (0,50) es cuando más alta es la variabilidad que produce, y es, por lo tanto, cuando mejor distingue a un individuo de otro. El número promedio de individuos de los que cada individuo es discriminado es entonces, un máximo (Magnusson, 1975).

En cambio, cuando un ítem es resuelto por todos o ninguno de los sujetos, no hay discriminación alguna y el ítem resulta impropio para el test. Los ítems que se agrupan alrededor de una dificultad de 0,50 permiten contar con máxima diferenciación entre los candidatos. Un ítem de dificultad $p = 0,25$ no puede tener correlación perfecta con otro con $p = 0,75$; así, en general, ítems con dificultad extrema correlacionan bajo con un test de dificultad promedio y es la discriminación de los ítems uno de los factores que determinan la variabilidad de los puntajes del test.

La forma de la distribución de los puntajes en un test depende de la dificultad de sus elementos. Los valores p determinan directamente el puntaje medio del test e influyen en su forma y su desviación estándar. La variancia de un test compuesto

de varios ítems depende totalmente de la variancia de cada ítem y de la covariancia.

La primera de ellas depende de la dificultad del ítem y la segunda de las correlaciones entre los ítems y por ende de las correlaciones de cada ítem con el resto de ellos (test). La variancia del test será entonces máxima cuando la variancia de los ítems sea máxima ($p = 0,50$) y las intercorrelaciones entre los ítems sean máximas.

Cuanto mayor sea la diferencia entre p y q tanto menor será la correlación ítem-test, de manera que los ítems muy fáciles o muy difíciles sistemáticamente tendrán bajas discriminaciones (Magnusson, 1975).

Estas bajas discriminaciones afectan tanto la variancia del test, la cual tiende a disminuir, como su confiabilidad.

Se dice que una población es normal con respecto a un rasgo si corresponde a una muestra aleatoria suficientemente grande y su distribución resulta superponible a la función de probabilidad que los matemáticos llaman normal. Mientras que un test con muchos ítems fáciles daría una distribución de puntaje con asimetría negativa, otro con muchos difíciles tendría asimetría positiva, pero en ambos extremos, la desviación estándar tenderá a ser pequeña.

La dispersión de la distribución de los puntajes depende parcialmente de la frecuencia de respuestas correctas para los ítemes y en parte de la magnitud de la correlación entre los ítemes individuales (Magnusson, 1975). Pero como ya dijimos, estas correlaciones a su vez dependen parcialmente de las dificultades de los ítemes.

Con la argumentación hecha hasta aquí, tanto para garantizarse una distribución modal con buena dispersión como para obtener una confiabilidad alta, pareciera que el test sólo debería componerse de ítemes de dificultad promedio.

Al estar los ítemes intercorrelacionados, los candidatos que aciertan un ítem, tenderán a acertar los demás, mientras que los que lo fallan tenderán a fallar los demás y esto haría que la distribución de puntajes fuera bimodal y no la de una curva modal. No se puede obtener una distribución normal del test, con ítemes que tengan máximo poder discriminativo e intercorrelaciones máximas (Magnusson, 1975).

Es mejor entonces escoger, para una prueba, ítemes cuya dificultad se disperse moderadamente, pero cuyo promedio corresponda a un porcentaje de acierto del 50,0% (Anastasi, 1972).

El problema sería entonces determinar esa "moderada dispersión" en la que se consideren los ítemes cercanos al 50,0% de acierto y se descarten los cercanos a los extremos (0,0% a 100,0%) para establecer así criterios que garanticen una distribución de los puntajes totales de la prueba que tenga una dispersión suficiente para efectuar la selección y una confiabilidad aceptable.

El problema debe, pues, resolverse por un compromiso entre las intercorrelaciones de los ítemes, su dificultad, la forma de la distribución y la confiabilidad.

La selección de candidatos a ingreso a la Universidad requiere conocer la posición exacta del alumno en la distribución de puntajes, por lo que se necesita que éstos se dispersen en una curva, y por ende, que las dificultades de los ítemes permitan tal distribución.

Por otra parte, la P.A.A. requiere un índice de confiabilidad alto que presupone altas correlaciones entre sus elementos.

Los índices de discriminación de los ítemes pueden obtenerse si se considera el test como uno sólo y cada ítem se correlaciona con el total de la prueba o si cada ítem es correlacionado únicamente con la parte de la prueba a la que pertenece (factor verbal o matemático).

Para establecer la validez predictiva, los puntajes en la P.A.A. y en admisión deben correlacionarse con una variable criterio evaluadora del rendimiento académico del alumno (variable por predecir). Esta puede definirse como un promedio ponderado de calificaciones universitarias o de cualquier otra forma que permita evaluar la bondad con la que la P.A.A. mide lo que debe medir.

En este estudio se han definido como criterios, diferentes variables. Por una parte, promedios ponderados por los créditos de las calificaciones en los cursos universitarios y por otra una variable acumulativa de éxito llamada "logro". En los promedios universitarios se separan aquellos cursos cuyas siglas son EG del resto de los cursos. Estos primeros corresponden al Sistema de Educación General del cual forma parte el Curso Integrado de Humanidades.

Es tradicional que la validez de las pruebas se calcule contra un promedio de calificaciones universitarias en el primer año. Este procedimiento tiene limitaciones a causa de que elimina el efecto del número de asignaturas que intervienen en su cálculo y no identifica cuáles de estos cursos fueron aprobados y cuáles reprobados. Por ejemplo, un promedio de 7,00 puede resultar de un solo curso aprobado con esa nota, de 5 cursos en las mismas condiciones o de un curso aprobado con 10 y tres reprobados con 6. Es evidente que el rendimiento es muy superior en el segundo caso.

El cálculo de correlaciones contra el promedio de calificaciones universitarias es una metodología práctica para validar los puntajes, pues no requiere un expediente de más de un año de estudios, pero supone que ese año es capaz de representar todo el quehacer universitario. Por otra parte, cálculos de validez con más de un año universitario se ven gravemente afectados por la pérdida de sujetos fundamentalmente en el primer año de estudios a causa de la deserción.

En un artículo anterior titulado "Validez predictiva de los puntajes de admisión versus éxito-fracaso en la Universidad de Costa Rica" (González de Wong y de León-Páez, 1986) se consideró como criterio por predecir la obtención de un título universitario o la suma de créditos acumulados por el alumno en un período de ocho años. Para el cálculo de la validez se usó el coeficiente de asociación T de Kendall entre las variables (medidas en escala ordinal) categoría de éxito-fracaso y admisión.

Resultó adecuado este criterio externo para medir el éxito relativo en la Universidad de Costa Rica para los alumnos que aún no se habían graduado. Esta metodología obligó a trabajar con una cohorte y requiere, por ende, de muchos años entre la medición de la variable y del criterio externo.

Estos antecedentes llevaron a la definición de una variable de criterio llamada "logro" que permitiera realizar el cálculo de la validez predictiva sin tener limitaciones típicas de un promedio, pero sí las ventajas del empleo de un número acumulado de créditos y la de no requerir de un período de tiempo muy extenso.

El "logro" tiene su fundamentación conceptual en el objetivo de la institución. La Universidad de Costa Rica es según su estatuto orgánico "Una institución autónoma de enseñanza superior... dedicada a la enseñanza... y a la difusión del conocimiento". Así, cuantos más cursos apruebe el alumno, mayor es el grado de éxito alcanzado, pero también, evidentemente, constituye mayor éxito el que un alumno gane el curso con la máxima nota, o que otro apenas lo apruebe.

Mientras que los promedios ponderados toman en cuenta los cursos no aprobados y el éxito se considera como una proporción con respecto al intento (matrícula), el logro considera el éxito acumulado, sin importar su proporcionalidad al intento. La magnitud del "intento" es más relevante en un análisis psicológico o en uno de costo-beneficio.

En este estudio se presentan los cálculos de correlación para variables criterio de promedios y logro en un primer año, un segundo año y un acumulado de los dos años.

Dentro de este marco conceptual, el presente estudio plantea los siguientes objetivos con base en la información del año académico 1982.

1. Analizar el efecto de diferentes grados de dificultad de los ítemes en la validez predictiva de los puntajes de la P.A.A. y de Admisión.
2. Estudiar el efecto de diferentes recorridos de los índices de dificultad en la validez predictiva, si se consideran diferentes criterios por predecir.
3. Estudiar el efecto en la validez predictiva de la utilización de ítemes cuya discriminación fue calculada contra el total de la P.A.A. o contra sus partes (verbal y matemática).

4. Analizar la incidencia de diferentes grados de dificultad en la confiabilidad del instrumento y en la distribución de los puntajes de admisión.

PROCEDIMIENTO METODOLOGICO.

La Prueba de Aptitud Académica (P.A.A.) se aplicó en 1981, en seis formularios paralelos en cada uno de los cuales había 70 ítemes para calificar al candidato. Cada alumno debía resolver sólo uno de estos formularios y disponía para ello de tres horas.

El puntaje obtenido por el alumno en la P.A.A. una vez trasladada a escala 1-100 se promedió con un promedio de calificaciones escolares para obtener así un puntaje de admisión.

La prueba se aplicó al finalizar 1981 a un total de 20.256 candidatos de los cuales quedaron 16.666, luego de separar a los egresados de secundaria y a los estudiantes universitarios que repetían la prueba. El número de personas que contestó cada formulario fue aproximadamente de 2.700 y con ese número se trabajó en el cálculo de la confiabilidad y la dispersión de la distribución.

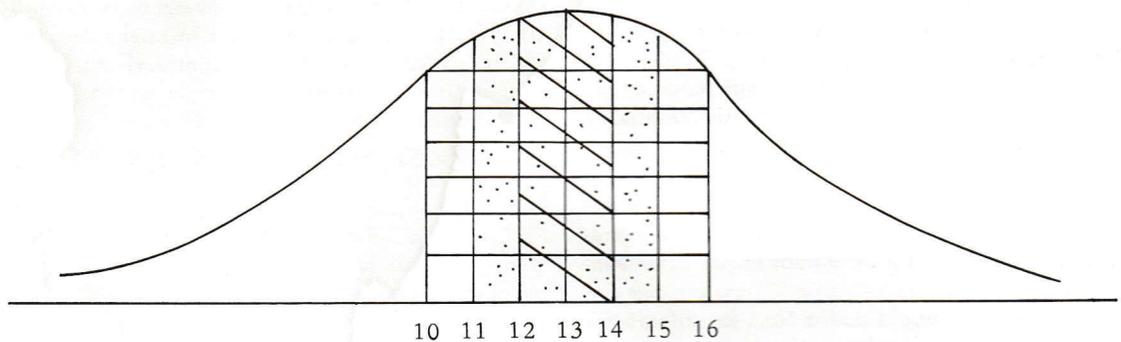
Para el cálculo de correlaciones se escogieron todos aquellos alumnos que hicieron efectiva su matrícula y que al finalizar los cursos de 1982 y 1983 tenían al menos una calificación numérica. Este número varía pues algunos alumnos sólo matriculan cursos EG, otros sólo adquieren créditos en el segundo año, etc. Pero en todos los cálculos se mantiene cercano a 3.000 sujetos.

Los ítemes que componían esa prueba y que habían sido aceptados de acuerdo con un análisis correspondiente a poblaciones anteriores, fueron analizados de nuevo en cuanto a su dificultad y discriminación para esta población.

Se fijaron así diferentes bandas o recorridos de dificultad alrededor del promedio en la distribución teórica de dificultades comúnmente empleada (promedio de 13 y desviación estándar de 4).

La banda 12-14 incluye sólo $\sigma/4$ a cada lado del promedio de dificultad. La 11-15 abarca $\sigma/2$ a cada lado y es la que actualmente se emplea para seleccionar los ítemes de la P.A.A. La banda 10-16 corresponde a $3\sigma/4$ a cada lado del promedio.

Para cada una de estas bandas de dificultad se hicieron análisis de discriminación de los ítemes tanto para el total de la prueba como para la parte a la que pertenece el ítem (parte verbal o matemática).



Con los ítemes aceptables según su discriminación ^{1/} y dificultad en cada banda se procedió a calificar a los alumnos para, con estos nuevos puntajes, proceder a calcular la validez predictiva.

Se calcularon coeficientes de correlación simple entre los puntajes en P.A.A. y Admisión y cada variable criterio. A saber:

Logro = suma de créditos por la calificación con la que se aprobaron (únicamente para cursos aprobados).

$$Cr_i N_i$$

Cr_i = créditos del curso aprobado i

N_i = calificación con la que se aprobó el curso i

$N_i = 0$ para cursos reprobados, suspendidos u otros

\overline{EG} = promedio ponderado de los cursos con siglas EG (únicamente para cursos con nota numérica o de equivalencia numérica).

$$\overline{EG} = Cr_i N_i$$

Cr_i = créditos del curso i de siglas EG.

N_i = calificación numérica en el curso i

$\overline{NO EG}$ = Promedio ponderado de los cursos con siglas diferentes a EG (únicamente para cursos con nota numérica o de equivalencia numérica).

$$\overline{NO EG} = \frac{Cr_i N_i}{Cr_i}$$

Cr_i = créditos del curso i de siglas diferentes a EG.

N_i = calificación numérica en el curso i .

\overline{UCR} = promedio ponderado de todos los cursos siglas EG y NO EG (únicamente para cursos con nota numérica o de equivalencia numérica).

$$\overline{UCR} = \frac{Cr_i N_i}{Cr_i}$$

Cr_i = créditos del curso i

N_i = calificación en el curso i

Todas estas variables criterio se utilizaron para 1982, para 1983 y para la suma de los dos años (2A).

RESULTADOS

En la banda más estrecha de dificultad de los ítemes (12-14) los alumnos fueron calificados con

(1) La Universidad de Costa Rica considera (para efectos de la P.A.A.) adecuados en lo que a discriminación se refiere, ítemes entre .35 y .70.

aproximadamente 30 ítemes (pues el número de preguntas aceptables en cada uno de los seis formularios de examen no fue el mismo).

En la banda 11-15 se usaron alrededor de 50 ítemes mientras que en la banda más amplia de 10-16 se usaron 58 ítemes o más.

La banda 12-14, que es la más cercana al promedio, no constituye la mejor opción para la selección de ítemes; más bien sus resultados son ligeramente inferiores a los obtenidos en las otras bandas más amplias.

De acuerdo con el Cuadro No.1 las diferencias entre las otras dos bandas resultaron mínimas, por lo que usar criterios más amplios en la aceptación de preguntas no altera la validez predictiva, pero sí permite ampliar los bancos de ítemes mediante la inclusión de preguntas más fáciles o más difíciles que las actuales. Este resultado es especialmente importante por cuanto el flexibilizar los criterios de aceptación de ítemes, permitirá contar con más preguntas de qué disponer para la estructuración de los formularios paralelos de examen.

Por otra parte, cuando el análisis de la discriminación de los ítemes se hace contra el sub-test al que pertenecen, en lugar de contra el total de la prueba, no se obtienen diferencias notorias en su efecto en la validez predictiva. Este resultado ratifica el procedimiento utilizado en la actualidad para el cálculo de la discriminación, a saber, contra las partes correspondientes de la prueba.

La predictibilidad de la variable admisión es superior a la de la P.A.A., lo cual es un resultado esperable por cuanto la admisión incluye a la P.A.A. además del IV ciclo y por ende mide además de aptitud académica, otros rasgos tales como conocimientos, constancia, intereses, etc. También es esperable que la medición realizada por un equipo de profesores durante un ciclo escolar aporte mucho a la realizada por una prueba por excelente que ésta resulte.

También para la variable admisión aceptar aquellos ítemes cuya dificultad se halle en una banda amplia alrededor del promedio resulta la decisión más acertada.

CUADRO 1

UNIVERSIDAD DE COSTA RICA

Correlaciones simples entre variables predictivas y variables criterio, según diferentes bandas de aceptabilidad de dificultad de los ítemes 1981-1982

Variable predictor	Banda de Dificultad	Tipo de analisis	Variable criterio											
			EG 82	NO EG 82	UCR 82	LOGRO 82	EG 83	NO EG 83	UCR 83	LOGRO 83	EG 2A	NO EG 2A	UCR 2A	LOGRO 2A
P.A.A.	12-14	Parcial	0,14	0,20	0,14	0,36	0,11	0,14	0,13	0,23	0,15	0,15	0,16	0,30
		Total	0,14	0,20	0,13	0,36	0,11	0,14	0,13	0,22	0,15	0,15	0,16	0,30
	11-15	Parcial	0,15	0,22	0,15	0,39	0,12	0,15	0,14	0,24	0,16	0,17	0,17	0,32
		Total	0,15	0,22	0,14	0,38	0,12	0,15	0,14	0,24	0,16	0,17	0,17	0,32
	10-16	Parcial	0,14	0,20	0,13	0,38	0,11	0,15	0,13	0,24	0,15	0,16	0,16	0,32
		Total	0,14	0,21	0,13	0,38	0,11	0,15	0,13	0,23	0,15	0,16	0,16	0,31
ADMISION	12-14	Parcial	0,24	0,34	0,24	0,50	0,19	0,28	0,26	0,36	0,26	0,30	0,27	0,45
		Total	0,24	0,34	0,24	0,49	0,19	0,27	0,26	0,35	0,25	0,30	0,26	0,45
	11-15	Parcial	0,26	0,36	0,25	0,52	0,20	0,30	0,28	0,37	0,27	0,33	0,28	0,47
		Total	0,25	0,37	0,25	0,52	0,20	0,30	0,28	0,37	0,27	0,33	0,28	0,47
	10-16	Parcial	0,25	0,36	0,25	0,52	0,20	0,29	0,28	0,37	0,27	0,32	0,28	0,47
		Total	0,25	0,36	0,25	0,52	0,20	0,29	0,27	0,37	0,26	0,32	0,28	0,47

FUENTE: IIP-UCR. Oficina de Informática U.C.R.

La variable criterio Logro es la que tiene mayor asociación con P.A.A. y Admisión, por lo que ofrece los índices de correlación más altos, especialmente el Logro en el primer año de estudios superiores, seguido por el Logro acumulado de dos años.

La pérdida de sujetos que se da entre el primer año universitario y los siguientes, afecta la variabilidad de los puntajes y por ende los coeficientes de correlación. Esto explica la disminución de los coeficientes del 83 respecto al 82.

Tanto para 1982 como para 1983 o la suma de ellos, siempre hay una mejor predicción para cursos universitarios de otras áreas diferentes a Estudios Generales. Como consecuencia de esto, es mejor la predicción de admisión con el promedio acumulado de dos años (UCR 2A) que con el primer año (UCR 82) donde los cursos EG tienen un peso muy fuerte.

Cuando se usa como variable criterio un promedio general resulta que es preferible una metodología de validación de los puntajes de P.A.A. o Ad-

CUADRO 2

UNIVERSIDAD DE COSTA RICA

Confiabilidad de la P.A.A., según diferentes bandas de aceptabilidad de dificultad de los ítemes. 1981-1982

Banda de dificultad	Tipo de Análisis	Formulario					
		1	2	3	4	5	6
12-14	Total	0,84	0,80	0,82	0,84	0,85	0,84
	Parcial	0,84	0,80	0,81	0,84	0,85	0,84
11-15	Total	0,90	0,87	0,88	0,90	0,89	0,88
	Parcial	0,90	0,87	0,88	0,90	0,90	0,88
10-16	Total	0,91	0,87	0,89	0,91	0,90	0,90
	Parcial	0,91	0,89	0,89	0,91	0,90	0,90

FUENTE: Instituto de Investigaciones Psicológicas - Universidad de Costa Rica, Oficina de Informática - Universidad de Costa Rica.

misión que considere un acumulado de dos años en lugar de sólo un año universitario. A pesar de que hay un efecto negativo en la correlación, originada por la pérdida de sujetos de primero a segundo año, es mayor la ganancia que se logra al considerar en el promedio, un historial académico más completo. En cambio, si se toma como variable criterio el logro o el promedio de cursos de siglas diferentes de EG, el primer año sería el más adecuado para validar los puntajes de P.A.A. o Admisión.

El hecho de que los ítemes de esa prueba habían sido anteriormente analizados, produce un efecto en el número de ítemes en cada banda que no debe ignorarse en el comentario de los resultados. La poca diferencia entre los resultados obtenidos para la banda 11-15 y la 10-16 bien podría deberse al bajo número de ítemes que diferencian éstas, aproximadamente 8.

Resulta entonces importante plantear otra investigación en la que se incluya un mayor número de ítemes con dificultades extremas ya que pareciera que la validez predictiva podría mejorar, o al menos ser igual a la actual, con criterios de aceptabilidad en una banda más permisible aún que las aquí tratadas.

En lo que se refiere a la confiabilidad de la prueba (Cuadro No.2) se obtienen índices adecuados para cualquier análisis que se haga, pero los mejores resultados se obtienen en la banda más amplia de dificultad, por lo que se concluye que es

CUADRO 3

UNIVERSIDAD DE COSTA RICA
Variabilidad de la distribución de los puntajes en P.A.A. y Admisión según diferentes bandas de dificultad de los ítemes. 1981-1982

Banda de Dificultad	Puntajes	Tipo de Análisis	Desviación Estandar
12-14	P.A.A.	Total	21,12
		Parcial	20,63
	Admisión	Total	12,25
		Parcial	12,02
11-15	P.A.A.	Total	19,57
		Parcial	19,21
	Admisión	Total	11,54
		Parcial	11,37
10-16	P.A.A.	Total	19,11
		Parcial	18,71
	Admisión	Total	11,32
		Parcial	11,13

FUENTE: Instituto de Investigaciones Psicológicas - Universidad de Costa Rica, Oficina de Informática Universidad de Costa Rica.

más importante el número de ítemes en su cálculo que las pérdidas en las intercorrelaciones de los ítemes originadas por la flexibilización en el criterio de dificultad.

Resta únicamente decidir si el efecto en la variabilidad de la distribución de los puntajes del test, es tanta que no permita considerar bandas más amplias de dificultad para los ítemes. A este respecto el Cuadro No.3 proporciona información de la que se concluye que las pérdidas en la dispersión de los datos no impide su empleo para fines de selección de candidatos a ingreso a la Universidad.

BIBLIOGRAFIA

- Anastasi, Anne, *Test Psicológicos*, Editorial Aguilar, España, 1977.
- Downe & Heath, *Métodos Estadísticos Aplicados*. Editorial Harla, México, 1979.
- González de Wong, María Isabel y de León-Páez Olga, Validez Predictiva de los Puntajes de Admisión vs. Exito-Fracaso en la Universidad de Costa Rica. *Revista de Educación*, Universidad de Costa Rica, Volumen 10, Número 1. 1986.
- Gullon, A. *Introducción a la Estadística Aplicada*. Editorial Alhambra, España, 1970.
- Kerlinger, Fred. *Investigación del Comportamiento*. Editorial Interamericana, México, 1979.
- Magnusson David, *Teoría de los Test*. Editorial Trillas, México, 1975.
- Nunnally, Jum, *Introducción a la Medición Psicológica*. Editorial Paidós, Buenos Aires, 1973.
- Van Dalen y Mejer, *Manual de Técnica de la Investigación Educativa*. Editorial Paidós, Argentina, 1978.