



CUADERNOS DE LA ESCUELA DE SALUD PÚBLICA

Volumen 3 Número 89 Caracas ENERO - JUNIO 2015

ISSN 0798-0388. Depósito legal pp. 196502 dj 714

http://saber.ucv.ve/ojs/index.php/rev_edsp

Tercer Volumen Digital

Fundada en agosto de 1965



ANÁLISIS DE LA PRUEBA DE RAZONAMIENTO NUMÉRICO Y LÓGICO DE LA EVALUACIÓN DIAGNÓSTICO EN EL ÁREA DE CIENCIAS DE LA SALUD DE LA UNIVERSIDAD CENTRAL DE VENEZUELA

Amalio Sarco Lira Barreto, Audy Salcedo, Gilberto Graffe ()*

RESUMEN

Hasta el Proceso de Admisión del año 2010, el mayor volumen de estudiantes admitidos en la UCV, ingreso por las llamadas pruebas internas que se aplicaban en cada una de las facultades. En 2010, la UCV decidió la realización de una prueba que permitiera evaluar las condiciones de ingreso de los aspirantes a estudiar en la institución, esta evaluación debía realizarse por carreras afines y se denominó Evaluación Diagnóstica por Áreas del Conocimiento (EDAC). En este trabajo se analizan los resultados de la prueba Razonamiento Numérico y Lógico, de la EDAC de Ciencias de la Salud (versión 2011). Se parte de las respuestas ofrecidas por los 9268 bachilleres que solicitaron ingreso a las carreras de Ciencias de la Salud. Primero se describe el rendimiento alcanzado por los aspirantes, luego se analiza la validez y confiabilidad de la prueba, por último, se evalúan el nivel de dificultad y el poder discriminativo de las preguntas. Los resultados indican que los aspirantes mostraron un bajo rendimiento en la prueba por lo que la mayoría de los ítems fueron calificados como Dificiles o Muy Dificiles, pero el instrumento se considera con suficiente validez y confiabilidad, por lo que resulta apto para estimar el posible rendimiento de los aspirantes en la Educación Superior. Los resultados estadísticos alcanzados por los participantes en la prueba sugieren que la Universidad debiera ofrecer alguna actividad académica remedial que permitiera mejorar las condiciones de entrada, al menos de los estudiantes admitidos a las diferentes carreras del Área de Conocimiento.

Palabras clave: Evaluación Diagnóstica. Condiciones de Ingreso. Educación Superior. Construcción de Prueba, Ciencias de la salud.

()Profesores de la Escuela de Educación, Facultad de Humanidades y Educación,
Universidad Central de Venezuela, Caracas, Venezuela. asarcolira@gmail.com*

Introducción

En julio de 2007, el Consejo Nacional de Universidades (CNU) decide la eliminación de la Prueba de Aptitud Académica que con carácter nacional se venía aplicando en el país desde el año 1984. Este acto administrativo obligó a la búsqueda de nuevos procedimientos para efectuar la selección y asignación de aspirantes a ingresar a la Educación Universitaria a través del Proceso Nacional de Ingreso (PNI). A tal efecto, el CNU designa una comisión técnica integrada por representantes de la Universidad Central de Venezuela (UCV), Universidad de Carabobo (UC), Universidad Simón Bolívar (USB), Universidad Centro Occidental Lisandro Alvarado (UCLA), la Universidad Pedagógica Experimental Libertador (UPEL), el Director de la Oficina de Planificación del Sector Universitario (OPSU) y un representante del Ministerio del Poder Popular para la Educación Universitaria (MPPEU).

La comisión realizó un ciclo de reuniones durante las cuales recibió y discutió las propuestas que le hicieron llegar los Núcleos de Secretarios y de Vicerrectores Académicos del CNU, las Universidades Autónomas y Experimentales, igualmente se recibió los aportes de algunos especialistas en admisión e ingreso. El lugar común de las participaciones de los núcleos, las universidades y de los especialistas estuvo signado por la necesidad de evaluar las condiciones de ingreso de los aspirantes, mientras que los representantes del MPPEU y la OPSU proponían el libre ingreso de los candidatos.

En mayo de 2008, la OPSU eleva a la consideración del CNU para su aprobación un procedimiento que denomino *Indicé Multivariado*, como mecanismo para la selección y asignación de aspirantes a través del PNI. Ese mecanismo basado en la lógica difusa incluye variables académicas y otros aspectos que son considerados como prioridades del Estado para juzgar los expedientes de los usuarios y decidir sobre la selección y asignación. Como ventaja comparativa con anteriores mecanismos, este último ofrece eliminar la desigualdad en el trato y la exclusión de diferentes sectores de la población. Sus propulsores indican que adoptar el mecanismo garantiza la Equidad y la Justicia Social en el Ingreso a la Universidad.

Durante los años 2008 y 2009, en la UCV se continúan aplicando las Pruebas Internas para la admisión de aspirantes a las diferentes facultades, 11 en total. González ⁽¹⁾, al analizar los resultados de la aplicación del Índice Multivariado para la asignación de aspirantes a la UCV a través del PNI, demuestra que su aplicación no garantiza equidad en el trato de los sectores de la población del nivel socioeconómico IV y V definidos mediante la metodología Graffar – Méndez Castellano ⁽²⁾, utilizada por OPSU y que apenas logra una mediana efectividad; ya que solo se inscribe el 55% de los aspirantes asignados a la UCV, lo cual deja en entredicho el cumplimiento de la oferta sobre la inclusión.

En setiembre de 2010, los Decanos de las 11 Facultades se reúnen con el Secretario y el Coordinador de la Secretaría de la UCV para evaluar la situación de la admisión a la Educación Superior tanto en el país, como al interior de la UCV. Luego de una amplia discusión se acordó la realización una prueba que permitiera evaluar las condiciones de ingreso de los aspirantes a estudiar en la UCV. Esta evaluación debía realizarse por carreras afines y se denominó *Evaluación Diagnóstica por Áreas del Conocimiento* (EDAC). Se definieron cuatro áreas: (a) *Ciencias de la Salud*, que agrupa las carreras de las facultades de Medicina, Odontología y Farmacia; (b) *Ciencia y Tecnología* que incluye las facultades de Ciencias, Ingeniería y Arquitectura; (c) *Humanidades y Ciencia Sociales y Jurídicas*, para atender la demanda por las facultades de Humanidades y Educación, Ciencias Económicas y Sociales y Ciencias Jurídicas y Políticas y (d) *Ciencias del Agro y el Mar*, para atender la demanda de las facultades de Agronomía y Ciencias Veterinarias.

Se establecieron comisiones técnicas de admisión en cada facultad e Interfacultades, por área de conocimiento, para el diseño y construcción de las distintas EDAC, las cuales se aplican en la UCV a partir del año 2011. A diferencia de las Pruebas Internas, a través de la EDAC los aspirantes pueden postularse a tres de las carreras ofrecidas por las facultades que componen el área de conocimiento.

Luego de tres años consecutivos de utilizar este nuevo mecanismo de selección de los aspirantes a ingresar a la UCV, parece lógico comenzar a realizar su evaluación. Son varias las perspectivas de evaluación que se pueden (por ejemplo: validez, confiabilidad, potencia de la prueba, resultados), ya sea para cada una de las subpruebas que conforman la EDAC o

de los resultados en general. Además también se puede evaluar la validez predictiva de los resultados o de cada una de las subpruebas. Cualquier evaluación que se haga debe ayudar a orientar a las personas que toman decisiones en la EDAC y en el ingreso de los estudiantes a la UCV. En esa línea se encuentra este trabajo.

Este trabajo forma parte de un conjunto de investigaciones que se realizan sobre las pruebas que conforman la *Evaluación Diagnóstica por Áreas del Conocimiento*. En este caso se trabaja con el área de Ciencias para la Salud y se describe la versión aplicada en el año 2011. Específicamente se analizan en detalle los resultados la prueba denominada Razonamiento Numérico y Lógico, primero desde la perspectiva del rendimiento logrado por los estudiantes y luego considerando dos indicadores fundamentales de la calidad de la prueba su validez y confiabilidad ⁽³⁾ y de los ítems que la integran, en los cuales se evalúa el nivel de dificultad que ofrecen al participante y el poder discriminativo de cada uno de ellos.

La Evaluación Diagnóstica para el Área de Ciencias de la Salud

La Evaluación Diagnóstica para el Área de Ciencias de la Salud, en su versión 2011, consta de siete (7) partes o subpruebas, cada una de ellas evalúa aspectos específicos. En total se dispone de 3 horas con 30 minutos para responder toda la evaluación. Este tiempo aparece distribuido entre las secciones componentes de manera que cada subprueba tiene un tiempo preciso estipulado. Al establecer el tiempo al usuario se le solicita también rapidez en la producción de la respuesta, la velocidad requerida es variable según el tiempo disponible y la extensión de las partes lo que nos lleva a concluir que además de diagnóstico hay subpruebas que se tipifican como de potencia por lo rápido que se deben dar las respuestas.

Las primeras cuatro subpruebas constituyen la evaluación de los conocimientos específicos que abarcan las áreas de: *Razonamiento Numérico y Lógico*, *Química*, *Biología* y *Habilidad Verbal*, el cuestionario en cada una de ellas consta de 30 preguntas. En los componentes de Química y Habilidad Verbal se ofrecen 5 alternativas de respuesta, en tanto que para Biología y Matemática las alternativas de respuesta son 4. En ninguna parte de los documentos revisados se explica esa diferencia entre el número de alternativas de las

diferentes pruebas. El tiempo disponible para cada una de estas cuatro subpruebas es de 30 minutos, de manera que el aspirante dispone, en promedio, de 1 minuto para responder cada pregunta, lo cual incluye la lectura y comprensión de la pregunta como tal, conocer primero el pie del ítem es decir la base de la pregunta, para luego pasar al análisis de las alternativas de respuesta que se le presentan y evaluarlas, para determinar la alternativa correcta.

Suponiendo que la comprensión de la base de la pregunta ofrezca la misma dificultad que la evaluación de cada alternativa de respuesta, cuando se le presentó cuatro alternativas el aspirante dispuso de aproximadamente 12 segundos para cada una (casos de las pruebas de *Biología* y *Razonamiento Numérico y Lógico*), en tanto que para las preguntas de cinco alternativas dispuso de 10 segundos (*Química* y *Habilidad Verbal*). Esta estimación de los tiempos disponibles es menor en el caso de la prueba de Habilidad Verbal, en la cual 15 de las 30 preguntas estaban referidas a dos textos que debían ser leídos y comprendidos para luego iniciar el trabajo de responder a cada una de ellas.

La segunda parte de la prueba es la evaluación de Aptitudes Específicas, y está constituida por las subpruebas de: *Aptitud Espacial*, *Intereses y Preferencias* hacia el área de Ciencias de la Salud y *Tendencias Conductuales*. La subprueba de *Aptitud Espacial* consta de 30 preguntas de 5 alternativas y se disponía de 25 minutos para responder, en promedio 50 segundos para cada pregunta y 8,3 segundos para la evaluación de cada alternativa.

El *Inventario de Preferencias hacia el área de Ciencias de la Salud* consta de 80 frases que describen actividades típicas de las diferentes profesiones del área y se solicita del aspirante indicar si la actividad es de su agrado o no, teniendo además la opción de decir que le es indiferente. Para responder a esta prueba se dispone de 25 minutos, en promedio 18,75 segundos para emitir cada respuesta. La última prueba es el *Inventario de Tendencias Conductuales* su propósito es evaluar los sentimientos y actitudes a través de una serie de 180 frases que usualmente son usadas por las personas cuando se refieren a sí mismas, se requiere que el aspirante indique si es verdadera o falsa. Para responder a esta prueba se dispone de 40 minutos, en promedio para cada ítem se dispone de 13,3 segundos.

La Evaluación de los Conocimientos Específicos

El conjunto de pruebas que permite la evaluación de los Conocimientos Específicos está constituido por dos pruebas de Razonamiento: Numérico Lógico y Habilidad Verbal y dos de conocimientos específicos: Química y Biología. El primer problema que se plantea en relación con la evaluación de pruebas es la validez de los instrumentos ⁽⁴⁾. Este constituye el punto de partida de toda la investigación, tener validez constituye el rasgo básico más importante en términos de la calidad de los instrumentos. La validez, en términos generales, se refiere al grado en que un instrumento realmente mide la variable que pretende medir. Aunque parece sencillo, no es tan fácil cuando se trata de variables complejas como motivación, inteligencia, habilidad verbal.

En la literatura especializada ⁽⁵⁾ se distinguen tres tipos de procedimientos para establecer la validez de un instrumento: (1) de contenido, 2) de criterio y 3) de constructo. Es importante señalar cada vez se tiende más a concebir la validez como un proceso unitario que tiene como objetivo aportar pruebas sobre las inferencias que se pueden realizar con un instrumento.

La validez de contenido, es la mínima necesaria para instrumentos como la evaluación de los Conocimientos Específicos ⁽⁶⁾. Ésta se refiere al grado en que un instrumento refleja un dominio específico de contenido de lo que se mide, el grado en que la medición representa al concepto medido.

De la revisión de los procedimientos utilizados por los equipos técnicos en el diseño y construcción de las pruebas y de los documentos consultados se evidencia que las pruebas tienen validez de contenido, ya que los instrumentos en general evalúan conocimientos específicos en cada área y en dos de ellas se evalúa, además, el razonamiento en el uso de los lenguajes: numérico lógico y verbal. Aun cuando no se dispone de una Tabla de Especificaciones para cada prueba, la revisión que se hiciera de las mismas por parte de los especialistas indica que el muestreo de contenidos sobre los cuales se elaboraron las preguntas se realizó en base a los programas de Educación Media de las diferentes asignaturas ⁽⁷⁾.

La prueba de Habilidad Verbal consta de tres partes, la primera de ellas de diez preguntas en las cuales se le suministra una definición y el aspirante debe seleccionar el concepto a que corresponde, dominio de vocabulario. La segunda parte consta de cinco preguntas, para cada una de ellas se le da una lista de cinco palabras, una de las cuales esta transcrita con un error ortográfico y se le solicita al aspirante identificar la palabra con el error, uso de reglas de ortografía. En la tercera parte se le suministran dos textos sobre cuyos contenidos se formulan quince preguntas, 8 y 7 respectivamente, que deben ser respondidas con base en la lectura realizada, comprensión de lectura.

Las pruebas de Razonamiento Numérico y Lógico, Química y Biología, están construidas sobre un muestreo de objetivos y contenidos enunciados en los Programas de Educación Media que corresponden a las asignaturas de Matemática, Química y Biología. En la prueba de Razonamiento Numérico y Lógico se consideran tópicos de Aritmética, Algebra, Geometría y Estadística. En la prueba de Química se consideran: nomenclatura química que incluye el manejo de conceptos fundamentales, materia, materiales químicos y soluciones, propiedades estructura y clasificación de los materiales, así como reacciones químicas con elementos básicos de interacción de los elementos y el manejo de fórmulas. Los contenidos de la prueba de Biología están referidos a: biología celular, estructura y funcionamiento, anatomía, estructura y funciones de sistemas, aparatos, nutrición, fases y metabolismo, y genética conceptos leyes y aplicaciones. Las preguntas en estas subpruebas son separadas, no se distinguen en ellas secciones o bloques.

En las subpruebas se pueden identificar grupos de preguntas por tópico o tema tratado en los programas de Educación Media, correspondiéndose el número de preguntas con la extensión del tema en dichos programas. En términos generales se puede afirmar que cada subprueba tiene validez de contenido.

La Información Recolectada y el Procesamiento Realizado

Para realizar este trabajo se contó con la colaboración de las facultades de Medicina, Odontología y Farmacia que integran en la UCV el área de Ciencias de la Salud; proporcionando la siguiente información: cuadernillo de prueba utilizado en 2011 (contiene todos las subpruebas aplicadas en esa oportunidad), las bases de datos producto de la

lectura digital de las planillas de inscripción de los aspirantes y las hojas de respuesta. Estos archivos contienen los datos personales, académicos y socioeconómicos y las respuestas seleccionadas por cada candidato en cada una de las preguntas de cada subprueba. También se nos suministró: las claves de respuestas utilizadas en la corrección, el instructivo para la aplicación de las pruebas, y la lista de los aspirantes seleccionados. Se contó además con una permanente atención del personal técnico de las tres facultades para conocer las etapas de trabajo y los procedimientos utilizados en cada una de ellas, con lo cual se logró una visión de conjunto de todo el proceso, además de aclarar las dudas que surgieron en el desarrollo del trabajo.

Las bases de datos fueron procesadas utilizando dos paquetes estadísticos, el Paquete Estadístico para las Ciencias Sociales versión 12 SPSS ⁽⁸⁾, por sus siglas en inglés, para generar estadísticas básicas que evalúan: tendencia central, dispersión, medidas de posición, forma de la distribución como asimetría y Curtosis sobre las pruebas y distribuciones de frecuencia sobre la demanda y los asignados para variables personales, académicos y socioeconómicos. Para evaluar la calidad de las pruebas y sus preguntas se utilizó el Paquete Lertap ⁽⁹⁾, mediante del cual se analizan los ítems, se calculan los niveles de dificultad y de discriminación de cada pregunta y se determinan la confiabilidad y el error de estimación de cada prueba.

La dificultad de un ítem se define como la proporción de personas que responden correctamente un ítem de una prueba. Entonces, para calcular la dificultad de un ítem, se divide el número de personas que contestó correctamente el ítem entre el número total de personas que lo respondió, no se consideran las personas que dejaron la pregunta en blanco. Para facilitar la comprensión, el nivel de dificultad se expresa en porcentaje ⁽¹⁰⁾. Entre menor sea el porcentaje, mayor será la dificultad del ítem evaluado, ya que significa que un menor número de estudiantes respondieron en forma correcta el ítem. Para clasificar la dificultad del ítem se utilizó la siguiente tabla:

Tabla 1: Nivel de dificultad

Nivel de dificultad	Porcentaje de respuestas correctas
Muy fácil	100 – 81
Fácil	80 – 61
Promedio	60 – 41
Difícil	40 – 21
Muy difícil	20 – 0

El índice de discriminación indica si existe relación responder correctamente o no una pregunta y la puntuación alcanzada por el aspirante en la prueba. El poder discriminativo alto en una pregunta indica que responder correctamente a esa pregunta está asociado con la obtención de altos puntajes en la prueba lo cual es evidencia de un alto desarrollo en el rasgo que se evalúa y las respuestas erradas aparecerán asociadas con un bajo puntaje en la prueba y un escaso desarrollo del rasgo medido. Esta debe ser una condición necesaria en cualquier pregunta que se incorpore a una prueba ya que el índice de discriminación también expresa la importancia de la pregunta dentro de la prueba y su contribución a la formación del puntaje total; si por el contrario el índice de discriminación es muy bajo debemos prescindir de ella pues su contenido será irrelevante en la evaluación del rasgo y lo único que garantiza es que el aspirante le dedique parte del tiempo que tiene para responder a la prueba sin garantía de retribución o compensación dado que estas preguntas aportan muy poco en las calificaciones altas.

Para medir el poder de discriminación de las preguntas se utilizó el coeficiente de correlación punto biserial ⁽¹¹⁾. Este coeficiente permite saber si las personas que muestran un buen desarrollo en el rasgo evaluado son las que emiten las respuestas correctas, qué tanto poder predictivo tiene el ítem sobre el puntaje total, y cómo puede contribuir a las predicciones sobre el comportamiento futuro. Para su interpretación se utiliza la información de la siguiente tabla ⁽¹²⁾:

Tabla 2. Índice de discriminación de los ítems según valor r_{pb}

Nivel de dificultad	r_{pb}	Recomendación
Bajo	< 0,2	Eliminar o revisar a profundidad
Medio	0,2 - 0,25	Revisar
Alto	0,25 - 0,35	Se puede mejorar
Muy Alto	> 0,35	Excelente

El coeficiente de correlación punto biserial (r_{pb}) es una estimación del coeficiente de correlación producto-momento de Pearson, considerando la puntuación total de la prueba y la pregunta, dicotomizada en respuestas correctas e incorrectas. Es por ello que se considera que el coeficiente proporciona información sobre el grado de asociación entre las competencias o rasgo que mide el test y la pregunta considerada.

En el año 2011 presentaron la Evaluación Diagnóstica para el Área de Ciencias de la Salud 9271 bachilleres que solicitaron ingreso en alguna de las escuelas de las facultades que conforman el área, la información procesada se identifica como DEMANDA.

Una vez procesada la demanda se seleccionó los expedientes de los admitidos de acuerdo a la lista proporcionada por las facultades y se aplicaron de nuevo los paquetes estadísticos SPSS y Lertap para generar las estadísticas de esta muestra sesgada de la demanda constituida por el grupo de los ADMITIDOS. En este sentido los resultados estadísticos que se exponen son indicadores de las características que exhiben tanto los aspirantes a ingreso como los candidatos seleccionados, ellos reflejan la conducta de los grupos ante el instrumento y esta información permite clasificar las preguntas según niveles de dificultad y discriminación así como establecer la confiabilidad de la prueba en su totalidad; por esta razón los estadísticos calculados al considerar la demanda ofrecen diferencias con respecto al grupo de los asignados pudiendo por ejemplo una pregunta resultar muy difícil en la demanda y de dificultad media entre los asignados. El número total de admitidos fue de 392, lo cual representa el 4,23% de la demanda, dicho de otra manera por este mecanismo se selecciona aproximadamente un candidato por cada veinticinco aspirantes evaluados.

No es objeto de este trabajo evaluar la selección ni el algoritmo a través del cual se decide la admisión de algún candidato a alguna de las escuelas, sin embargo, para una

mejor comprensión de la información estadística que se suministra es conveniente aclarar que la puntuación asignada a cada aspirante en cada prueba es igual al número de preguntas respondidas correctamente, que cada prueba tiene una ponderación específica y que el ordenamiento de los candidatos en la demanda se realiza en función de los resultados alcanzados en el conjunto de pruebas y no de alguna en particular, el índice mediante el cual se realiza el ordenamiento constituye una compilación de los resultados parciales obtenidos en cada prueba. Esto hace posible que alguno(s) de los seleccionados pudiera no haber obtenido resultados satisfactorios en alguna(s) de las pruebas y que aun cuando los seleccionados constituyen el 4.23% de la demanda, los alumnos seleccionados pudieran no formar parte del 5% superior en todas las pruebas.

Resultados alcanzados en la Prueba de Razonamiento Numérico y Lógico

El rendimiento de los grupos ante la prueba

El promedio aritmético de la demanda fue de 7,46 puntos de un total de 30 posibles lo cual representa un rendimiento aproximado de 25% de lo solicitado en la prueba. El grupo de los seleccionados promedió 12,05 puntos, algo más del 40% de lo solicitado. La diferencia en el rendimiento de los grupos es de 4,59 puntos, equivalente al 15,3% de la extensión de la prueba. El grupo seleccionado en promedio contestó correctamente cinco preguntas más que los demandantes.

Tabla 3: Comportamiento de la Demanda y los Asignados EDAC Ciencias de la Salud 2011. Prueba de: Razonamiento Numérico – Lógico

Índices	Media Arit.	Mediana	Modo	Xi menor	Xi Mayor	Desv. Tipica
Demanda	7,46	7	7	0	29	3,876
Asignados	12,05	11	9	2	29	5,421
Dif. Asig.- Dem	4,59	4	2	2	0	1,545
Diferencia %	15,3	13,33	6,67	6,67	0	5,15

Un comportamiento similar se observa en las otras medidas de centralización. La Mediana que divide la distribución de frecuencias en dos partes iguales pasa de 7 puntos en la Demanda a 11 puntos en los Asignados y el Modo que es la puntuación de más alta

frecuencia pasa de 7 a 9 puntos. Aunque en el grupo Asignados presenta un mejor rendimiento que el grupo Demanda, esa diferencia no es sustancial.

La desviación típica brinda información sobre la homogeneidad del grupo analizado. Los resultados indican que el grupo Demanda, es homogéneo, tiene una desviación típica de 3,876 puntos, sobre un recorrido de 30 puntos. Esto sugiere que se está en presencia de un grupo con poca variabilidad que se agrupa en los valores bajos de la variable; lo cual denota un bajo rendimiento. La variabilidad en el caso de los Asignados es de 5,421 puntos, ahora en un recorrido de 28 puntos, lo que sugiere que es un grupo heterogéneo. No obstante, el coeficiente de variación, la relación porcentual entre la desviación típica y la media, indica que el grupo de Asignados (44,99%) es más homogéneo que el grupo Demanda (51,96%), esto se debe a que el incremento en la medida de tendencia central es de cinco puntos y el aumento en la medida de dispersión es de un punto con respecto a la demanda.

Las medidas de posición y la forma de la distribución de puntajes

Los índices de asimetría tanto de la Demanda como de los Asignados son positivos, 0,8 y 0,68 respectivamente, lo cual sugiere que en ambas distribuciones los alumnos tienden a concentrarse hacia los valores bajos de la variable. Esta concentración es más acentuada en el caso de la Demanda por cuanto la medida alcanza un valor mayor. Este resultado se corresponde con lo indicado en la sección anterior, donde las medidas de tendencia central sugerían un menor rendimiento en el grupo Demanda en comparación con el grupo Asignado.

Tabla 4: Comportamiento de la Demanda y los Asignados EDAC Ciencias de la Salud 2011. Prueba de: Razonamiento Numérico – Lógico.

Índices	Asimetría	Centil 25	Centil 50	Centil 75	Kurtosis
Demanda	0,8	5	7	10	1,372
% Prueba		16,67	23,33	33,33	
Asignados	0,677	8	11	16	-0,208
% Prueba		26,67	36,67	53,33	
Dif.Puntuac.		3	4	6	
Diferencia %		10	13,33	20	

En relación con la altura relativa que alcanza la distribución, en la demanda la curva es leptocúrtica el índice es de 1.372 y la concentración de casos es superior a la del Modelo de Distribución Normal, en el caso de los seleccionados la curva es ligeramente platicúrtica el índice es -0.208 y hay una mayor dispersión de los puntajes alcanzados de la que se considera Normal.

Las medidas de posición indican que el 25% de la demanda solo alcanza a responder correctamente un máximo de 5 preguntas, lo que representa aproximadamente el 17% de la prueba, mientras que en el caso de los admitidos es 8 preguntas, que representa el 27%. La mitad de los aspirantes responde un máximo de 7 preguntas correctamente y los seleccionados un máximo de 9, lo cual representa un 23% y un 37% respectivamente. El centil 75 de la demanda se ubica en 11 puntos, lo cual corresponde al 37% de la prueba, los admitidos se ubican en 13 que representa el 43%. Esta dispersión de valores en el 25% superior de la distribución se corresponde con la cola de la derecha que se indicó al analizar los índices de asimetría que resultaron positivos. Obsérvese que en la Demanda el 75% de los aspirantes obtienen puntuaciones entre cero y diez puntos en tanto el 25% superior se distribuye entre los 11 y los 29 puntos.

De acuerdo con las medidas de tendencia central, el grupo Asignados presenta siempre un rendimiento superior al grupo Demanda, y aunque internamente es un grupo heterogéneo, cuando se le compara, mediante el coeficiente de variación, con el grupo Demanda, resulta ser un grupo más homogéneo. Estos son resultados lógicos, por cuando se espera que el grupo de estudiantes seleccionados presente un mejor rendimiento que el grupo total. El índice de asimetría calculado indica la existencia de una asimetría positiva en ambos grupos: concentración de puntuaciones en los valores bajos de la variable. El grupo Asignados presenta una asimetría menor que el grupo Demanda, lo cual indica que aunque los dos tienen concentración hacia valores bajos de la variable, el del grupo Asignados es menor, por lo tanto, se puede suponer un mejor rendimiento en este grupo. Esta observación coincide con la información que aportan las medidas de posición. En general, los resultados indican que la prueba de Razonamiento Numérico y Lógico, resultó

difícil para el conjunto de los aspirantes y de dificultad media para los candidatos seleccionados.

La Confiabilidad de la Prueba y los Niveles de Dificultad y de discriminación de las preguntas

Evaluación de la consistencia interna

Para analizar la Confiabilidad del instrumento se utilizó el índice ideado por C.J. Hoyt en 1941 ⁽⁴⁾ que evalúa la consistencia interna de la prueba. La evaluación de la Confiabilidad permite conocer con que precisión se logra la medición del rasgo, una vez establecida la validez, la confiabilidad es la segunda condición en orden de importancia para evaluar la calidad de una prueba. La consistencia interna nos informa el grado de homogeneidad con que el evaluado se comporta a través del instrumento, esta característica se traduce al instrumento en el sentido de que todas las preguntas evalúan el mismo rasgo y el instrumento es homogéneo internamente. El índice de Hoyt es probablemente uno de los más potentes por el uso intensivo que hace de la información, está basado en el Análisis de la Varianza (ANOVA) y los cálculos se realizan a partir de los resultados obtenidos por los participantes en la prueba completa y las respuestas emitidas por todos los alumnos a cada pregunta de la prueba, los resultados se ofrecen en un rango de cero a uno.

La consistencia interna es la evaluación de confiabilidad más adecuada en este tipo de pruebas en las cuales se evalúan conocimientos, el uso de otras técnicas como la de Pruebas Paralelas resulta impráctico desde el punto de vista logístico, significaría al menos duplicar el tiempo de aplicación de la prueba que en este caso pasaría de 3 horas y 30 minutos a 7 horas, la del test retest haría muy compleja la interpretación de resultados ya que el conocimiento es un rasgo modificable en el corto tiempo y las diferencias obtenidas en las mediciones antes-después habría que explicarlas en función del esfuerzo individual de los participantes por mejorar su ejecución, esto último resulta muy difícil de conocer y más de cuantificar.

El Hoyt calculado para la demanda es de 0,55 y para los admitidos es de 0,81, en relación con los resultados podemos decir que el uso de las puntuaciones obtenidas en la

prueba se hará con mayor seguridad y confianza en la medida que el índice Hoyt se aproxime a uno, se puede afirmar que el comportamiento de los admitidos ante la prueba es mucho más homogéneo que el observado en la demanda, por ello hay mayor consistencia interna en los resultados, obsérvese que el error de estimación en los admitidos 2,32 es medio punto menor que el calculado para la demanda 2,78 lo cual se traduce en una mayor precisión en la medición y esto explica en buena parte el aumento de la consistencia interna.

Los niveles de dificultad y de discriminación de las preguntas de la prueba

La siguiente tabla muestra la distribución de las preguntas clasificadas según el grado de dificultad que ofrecen al aspirante o al seleccionado,

Tabla 5: Nivel de dificultad

Nivel de dificultad	Demanda	Asignados
Muy fácil	0	0
Fácil	0	4
Promedio	3	10
Difícil	11	13
Muy difícil	16	3

Para la Demanda, 27 de las 30 preguntas resultaron con un nivel de dificultad superior al Promedio, eso significa que 27 preguntas fueron respondidas en forma correcta por menos del 40% de los aspirantes. Es importante destacar que eso incluye 16 preguntas que fueron respondidas de forma correcta por menos del 20% de los bachilleres que presentaron la prueba. Este resultado indica que la prueba de matemática resultó muy difícil para Demanda.

Para los admitidos, la mayoría de los ítems se encuentran entre Promedio y Difícil, incluso se presentan 4 preguntas clasificadas como Fáciles y solo 3 se incluyen en la categoría Muy Difícil. Estos resultados sugieren la dificultad de la prueba se puede ubicar como Promedio. Esta tendencia que se observa en el grupo de admitidos de fijar los niveles de dificultad hacia el promedio $p=0.5$, facilita la definición de grupos diferenciados y eleva los niveles de discriminación de las preguntas que pasan en el nivel muy alto de 7 en la demanda a 19 en los admitidos.

PRUEBA DE RAZONAMIENTO NUMÉRICO Y LÓGICO DE LA EVALUACIÓN DIAGNÓSTICO EN EL ÁREA DE CIENCIAS DE LA SALUD

Para ilustrar el tipo de pregunta clasificada en cada nivel, a continuación se encuentran algunos ejemplos.

Ítem 15. El siguiente cuadro numérico está formado por sumas resueltas en filas y columnas. En ella hay un número equivocado. Indique el número de que se trata.

36	18	12		66
45	11	6		62
21	16	14		50
<hr/>				
101	45	32		178

A. 21 B. 18 C. 36 D. 16

Este ítem fue resuelto por el 59,3% de la Demanda, el porcentaje más alto de respuestas correctas en este grupo, por lo que se le considera de dificultad *Promedio*, mientras que el caso de los asignados fue resuelto por el 76,8%, por lo que se catalogaría como *Fácil*, para ese grupo.

La información que se les ofrece a los estudiantes le debe permitir suponer que los totales de filas y columnas que se encuentran en la tabla son correctos. No obstante, como ítems se indica que en el cuerpo de la tabla hay un número equivocado, alguno de esos totales debería ser diferente si se sumen los números que se encuentra en la tabla. Luego de verificar las sumas de las filas, se encuentra en la última da 51 aun cuando se indica 50. Es de suponer entonces que allí se encuentra el número equivocado. Al hacer el mismo proceso con las columnas, se puede identificar una diferencia en el total de la primera columna, se indica 101 pero la suma es 102. Por esta razón el estudiante debería indicar que el número equivocado es 21, alternativa A. El distractor que capta mayor la atención de los aspirantes es la alternativa D (14%), lo cual indica que identifican la fila donde se encuentra el error, pero se equivocan de columna.

El ítem de mayor dificultad para la Demanda fue el siguiente:

Ítem 9. El área de un triángulo rectángulo es de 12 centímetros cuadrados. Si los catetos están en relación 2:3, entonces la medida en centímetros de la hipotenusa será:

A. $\sqrt{13}$ B. $\sqrt{26}$ C. $3\sqrt{13}$ D. $\sqrt{52}$

Esa pregunta sólo fue resuelta en forma correcta por el 9,6% de la demanda, por lo que se incluye en la categoría de *Muy difícil*. El 19,9% de los Asignados ofreció la respuesta correcta para esa pregunta, aproximadamente un 10% más de respuestas correctas, que la Demanda, pero igual se cataloga como un ítem *Muy Difícil*, para ellos.

Esta es una pregunta donde el estudiante debe relacionar varios contenidos matemáticos. El estudiante debe partir de la idea de área de un triángulo rectángulo y la relación de ese cálculo con los catetos. Luego debe considerar que esos catetos guardan una cierta relación en su longitud, para posteriormente aplicar el teorema de Pitágoras y hallar la medida de la hipotenusa. La cantidad de pasos involucrados en la solución de la pregunta aumentan la posibilidad de incurrir en un error y terminar señalando una respuesta incorrecta. Eso parece ser lo que pensaron 2612 aspirantes que optaron por dejar la respuesta en blanco (28%), lo cual sin duda indica que la pregunta es considerada de alta dificultad. El número de respuestas en blanco es superior al de cualquier alternativa de respuestas, lo que ratifica que los aspirantes consideran que es una pregunta con un elevado nivel de dificultad, por lo cual es preferible no ofrecer una respuesta.

La siguiente pregunta fue resuelta de forma correcta por el 38,6% de la Demanda y el 56,1% de los Asignados. Es una pregunta clasificada como *Difícil* para la Demanda y *Promedio* para los asignados.

Ítem 11. ¿Cuál es el 10% de $\frac{1}{3}x$ si $\frac{2}{3}x$ es 10% de 60?

- A. 0,1 B. 0,2 C. 0,3 D. 0,4

En este ítem se sabe que $\frac{2}{3}x$ es 10% de 60 y se solicita que se indique el valor del 10% de $\frac{1}{3}x$. Entonces es necesario operar con la primera expresión y darse cuenta que se debe solucionar la ecuación $\frac{2}{3}x = 6$, para hallar el valor de x . Luego trabajar con ese valor de x , para encontrar primero $\frac{1}{3}x$ y luego el 10% de ese nuevo valor. Se trata de una pregunta con varios pasos para hallar su solución, donde un grupo considerable de los aspirantes opta por dejarla en blanco. Un total de 2191 (24%) aspirantes decidieron no ofrecer ninguna respuesta, no 38,6% de la Demanda ofrece la respuesta correcta. La diferencia con la pregunta anterior, es que una vez que se comprende lo que se solicita, el

estudiante puede trabajar algebraicamente con expresiones y algoritmos conocidos para encontrar la respuesta.

Los tres ejemplos anteriores son ilustrativos de como los mismos ítems tienen distintos niveles de respuestas correctas en los grupos considerados, lo cual se refleja en las puntuaciones logradas en las pruebas y en sus posibilidades de ingreso en las carreras de salud.

Para el grupo Demanda, las 10 preguntas con mayores niveles de dificultad corresponden a Geometría (3), Ecuaciones (2), Fracciones (2), Trigonometría (1), Polinomios (1) y Razonamiento lógico (1). En el caso de los Asignados, las preguntas de mayor dificultad corresponden a tópicos de: Geometría (3), Ecuaciones (3), Fracciones (1), Trigonometría (1), División de polinomios (1) y Razonamiento lógico (1). Se observa bastante coincidencia en los dos grupos en cuanto a los temas con preguntas con mayores niveles de dificultad, lo cual podría sugerir fallas en la formación en educación media de los aspirantes en esos temas o en la forma en que se presenta la pregunta es poco familiar.

Los temas de las preguntas de menor dificultad en el caso de la demanda son: Razonamiento lógico (5) Ecuaciones (3) Aritmética (1) y Fracciones (1). Para los Asignados los temas son: Ecuaciones (5) Razonamiento lógico (3) Aritmética (1) y Fracciones (1). Nuevamente se presenta bastante coincidencia entre los dos grupos. Llama la atención que el tema de Ecuaciones se encuentre tanto en los ítems con mayor y menor dificultad, esto podría explicarse por los niveles cognitivos exigidos por la pregunta. Cuando se trata de ecuaciones planteadas en forma directa o de problemas que se puedan llevar a ecuaciones con cierta facilidad, las preguntas son de bajo nivel de dificultad. Pero cuando se trata de problemas que requiere una buena comprensión del texto para poder plantear la ecuación, el nivel de dificultad es alto. En el caso de las fracciones ocurre algo semejante. Cuando se trata de preguntas que requieren para su solución una única operación con fracciones, la pregunta suele ser de baja dificultad, pero cuando involucra dos o más operaciones con fracciones la dificultad es alta. Las preguntas de Razonamiento lógico en general no involucran conocimientos matemáticos más allá de las operaciones básicas, por lo que tienden a ser de baja dificultad, aunque cuando se aumenta el nivel de exigencia

cognitiva, involucrando realizar conexiones entre informaciones, la pregunta sube en el nivel de dificultad.

En la tabla de niveles de discriminación se observa que solo una pregunta resulto con bajo poder discriminativo entre los admitidos ⁽¹²⁾.

Tabla 6: Nivel de discriminación

Nivel de discriminación	Demanda	Asignados
Bajo	10	1
Medio	5	2
Alto	8	8
Muy Alto	7	19

El poder de discriminación de los ítemes es apreciable por la conducta de los aspirantes en la prueba, al observar que una tercera parte de las preguntas, diez en total, tiene un bajo poder discriminativo en la demanda conduce al analista a revisar en detalles el comportamiento de la población ante estas preguntas en especial, una vez que constatamos la veracidad de la clave de respuestas, esto fue lo que se observó en las salidas del paquete Lertap: en todas ellas la opción correcta no fue la más seleccionada como respuesta, en 8 de las 10 fue superada en la selección por al menos dos de las alternativas incorrectas y en tres de ellas fue superada en la selección por las tres alternativas de respuesta incorrectas, esto sugiere que los participantes no logran identificar las respuestas correctas descartando las que parezcan incorrectas, evidenciándose un bajo dominio de los contenidos que sirven de base a la elaboración de estas preguntas.

El aumento del poder discriminativo de los ítems en el grupo admitidos permite la definición más homogénea de los grupos, su conducta ante la prueba está mejor definida y esto aumenta el índice de consistencia interna, asociado a esto el hecho de que para la selección la prueba resultara de dificultad promedio con tendencia a 0.5 hace que aumenten los promedios de tendencia central y facilita el incremento de los índices de discriminación en las diferentes preguntas.

El ítem con menor poder de discriminación en la demanda fue el siguiente:

Ítem 13. Si cada uno de los lados de un rectángulo se incrementa en un 100%, entonces el área se incrementa en:

- A. 100% B. 200% C. 300% D. 400%

El r_{pb} de esta pregunta fue 0,09; es una pregunta que confunde tanto a los sujetos que lograron obtener un alto número de respuesta correcta como aquellos que tienen un bajo número de respuestas acertadas dentro del grupo de la Demanda, por lo cual se le califica como de baja discriminación. En el caso de los Asignados, ese mismo ítem tiene un r_{pb} de 0,15, también es la pregunta con menor poder de discriminación para este grupo. Aunque la correlación, entre las puntuaciones de la prueba y responder correctamente o no este ítem, es mayor aún se le considera con un bajo poder de discriminación.

El mejor distractor de esta pregunta es la alternativa D, señalada por el 37% de la Demanda. Los estudiantes que señalan esta alternativa parece que se observaron que la pregunta es sobre el incremento en el área de rectángulo, luego de aumentar la medida de cada lado en 100%. Si bien el área del nuevo rectángulo ($4b.h$) es 400% mayor que el original ($b.h$), el incremento del área, respecto a ese rectángulo inicial, es de 300% (respuesta correcta C). Es importante señalar que la alternativa correcta sólo fue señalada por el 16% de la Demanda, siendo la menos indicada en esa pregunta. El comportamiento de los Asignados es similar a la Demanda. La pregunta sugiere que hay problemas con la comprensión de lo que se le requiere, de allí que la alternativa más señalada sea la D; pero también podría sugerir problemas con los contenidos que involucra la pregunta, cálculo de área, cálculo de porcentajes, áreas expresadas en porcentajes.

El ítem 6 es de mayor poder de discriminación en el grupo Asignados.

Ítem 6. La suma de dos números es 108 y su cociente 5. ¿Cuál es el mayor de estos números?

- A. 90 B. 85 C. 95 D. 75

Esta pregunta tiene un r_{pb} de 0,56, por lo que se le incluye en los ítems con un poder de discriminación *Muy Alto*. Este resultado indica que en el grupo de los asignados, los que lograron un alto número de respuestas correctas tienen a responder de forma correcta este ítem. Mientras que los de menor número de respuestas correctas en la prueba, por lo general, ofrecen respuestas incorrectas en esta pregunta. El r_{pb} de esa misma pregunta en el

grupo Demanda es 0,39, razón por la cual también se le incluiría en el grupo de poder de discriminación *Muy Alto*.

La alternativa C (33%) es la más señalada por la Demanda. La forma normativa de hallar la respuesta a esta pregunta es mediante la solución de un sistema sencillo de dos ecuaciones con dos incógnitas. Al solucionar el sistema se encuentran dos números, 90 y 18; por lo cual la respuesta correcta es la alternativa A. El número 95 es el producto de un error en la solución del sistema de ecuaciones. Otra posible explicación es que los estudiantes hayan identificado a 95, como el número mayor de la lista que es divisible entre 5.

La pregunta con mayor poder de discriminación en el grupo Demanda es el siguiente:

Ítem 26. La suma de dos números es 21, y su diferencia es 9. Señale esos dos números.

A. 17 y 8 B. 12 y 9 C. 15 y 6 D. 15 y 5

Esa pregunta tiene un r_{pb} de 0,44, razón por la cual se le considera de un *Muy Alto* poder de discriminación. El r_{pb} en el grupo Asignados para esta pregunta fue de 0,43, lo que lo ubicaría también entre los ítems con poder de discriminación *Muy Alto*.

Esta pregunta también se puede solucionar mediante un sistema sencillo de dos ecuaciones con dos incógnitas, no obstante, también se puede solucionar por inspección simple. Si se observan las alternativas, los números de A y D suman 25 y 20 respectivamente, con lo cual se incumple con la condición de que “suma de dos números es 21”. Entonces esas son dos alternativas descartables con rapidez. De los números que quedan en las alternativas restantes, los únicos que cumple con la condición “su diferencia es 9”, son los que están en la alternativa C (respuesta correcta). La alternativa B es el mejor distractor y capta el 38% de las respuestas indicadas. Los números 12 y 9 se producen al incurrir en un error en la solución del sistema de ecuaciones. Otra posible explicación de por qué se señala esta alternativa, es que los estudiantes observen que los dos números suman 21 y por la premura de ofrecer respuestas rápidas no verifique el cumplimiento de la segunda condición del problema. Es importante señalar que el 21% de los estudiantes optó por dejar la respuesta en blanco, lo cual indica que un grupo importante de la Demanda

considera que la pregunta es muy difícil para ello o que no tiene los conocimientos necesarios para resolverla.

Los temas de las preguntas con mayor poder discriminativo en el grupo Demanda son: Ecuaciones (5), Razonamiento lógico (2), Fracciones (1) y Aritmética (1). Para el grupo Asignados, los temas de Ecuaciones (6), Razonamiento lógico (1), Fracciones (3) corresponde a las 10 preguntas con mayor poder discriminativo. Las preguntas con menor poder discriminativo en el grupo demanda corresponden a los temas: Ecuaciones (4), Geometría (3), Razonamiento Lógico (2), Trigonometría (1). Para el grupo de Asignados las 10 preguntas con menor poder de discriminación se distribuyen por tema de la siguiente manera: Ecuaciones (4), Razonamiento lógico (3), Geometría (2), Trigonometría (1).

Conclusiones

Las distintas medidas estadísticas calculadas indican que en la prueba de Razonamiento Numérico y Lógico, el grupo Asignados presenta un mejor rendimiento que el grupo Demanda, tal como se espera en un evento de este tipo. Asimismo la prueba se puede catalogar como difícil para el grupo total de aspirantes y de dificultad promedio para el grupo Asignados.

La prueba en general presenta un nivel de confiabilidad adecuado, aunque es mayor en el grupo Asignados, razón por la cual se puede afirmar que hay mayor consistencia interna en los resultados de este grupo con referencia al grupo total. El comportamiento de los Admitidos ante la prueba es mucho más homogéneo que el observado en la demanda. El nivel de discriminación de los ítems que conforman la prueba es adecuado, aunque, de acuerdo con los resultados de la Demanda hay varios ítems que deberían revisarse.

En la evaluación psicométrica o edumétrica, considerando, la validez, confiabilidad, error de medición, niveles de dificultad y niveles de discriminación, sugieren que se trata de una prueba de calidad suficiente y el uso de los resultados con fines de selección es adecuado. Sin embargo, el hecho de no disponer de un plan de construcción, o Tabla de Especificaciones en la que se indique el número de la pregunta, que contenido evalúa, como lo hace, que tipo de pregunta se formula y con qué nivel de dificultad, limita el uso de los

resultados con fines predictivos. Las nuevas versiones de la prueba pudieran ofrecer un nivel de dificultad diferente o variar los contenidos de que se trate, lo cual hace que las mediciones sean distintas, no comparables. Mientras las versiones de la prueba puedan variar libremente, podremos computar los resultados obtenidos cada año pero no podríamos trazar expectativas para el año siguiente con base a los resultados obtenidos en un año dado.

Las estadísticas que evalúan el rendimiento del grupo Demanda en la prueba, así como las que corresponden al grupo Asignados, indican que hay un bajo dominio de los contenidos matemáticos en algunos de los temas evaluados por lo que las facultades debieran ofrecer alguna actividad remedial que les permitiera, al menos a los seleccionados, subsanar las deficiencias para enfrentar con suficiencia las dificultades propias de las carreras a las cuales se postularon.

Para las Autoridades de las diferentes facultades que conforman el Área de Ciencias de la Salud, así como para los equipos profesionales encargados de coordinar la Evaluación Diagnóstica, disponer del Plan de Construcción ofrecería ventajas adicionales: les permite un mayor control sobre el desarrollo del trabajo de las Comisiones o equipos técnicos, y en caso de necesitar la contratación de agentes externos, dispondrían de un documento técnico en el que se describen las características del trabajo a contratar.

Referencias

1. González, J. Asignación de aspirantes a través del Sistema Nacional de Ingreso a la Educación Superior (SINIES), UCV 2008. Revista de Pedagogía. vol.31, n.88 pp. 39 – 60.
2. Méndez Castellano, H. Sociedad y Estratificación. Método Graffar-Méndez Castellano. Caracas: Fundacredesa. 1994.
3. Magnusson, D. Teoría de los Test. México: Trillas. 1972.
4. Nunnally, J. Teoría Psicométrica. México: McGraw-Hill. 1995.
5. Thorndike, R y Hagen, E. Test y técnicas de Medición en Psicología y Educación. México: Trillas. 1970.
6. Sachs Adams, G. Medición y Evaluación. Barcelona. Editorial Herder. 1975.

7. EDACI 2011. Materiales y archivos digitales.
8. Pérez, C. Técnicas Estadísticas con SPSS 12. Madrid: Pearson Prentice Hall. 2005.
9. Nelson, L.R. Item Analysis for Tests and Surveys. Using Lertap 5. Perth: Curtin University of Technology. 2001.
10. Crocker, L., y Algina, J. Introduction to Classical and Modern Test Theory. New York: CBS College Publishing. 1986.
11. Tavella, N. (). Análisis de los Ítemes en la construcción de Instrumentos Psicométricos. México: Trillas. 1978.
12. Ponce, M. (). Reporte Técnico de la Prueba Interna de la Facultad de Humanidades. Mimeografiado. UCV. 2005.