

## **Propiedades psicométricas de la prueba de inteligencia Otis en estudiantes de Educación mención Orientación\***

*Marhilde Sánchez de Gallardo\*\* y Ligia Pirela de Faría\*\*\**

### **Resumen**

El objetivo de esta investigación fue determinar las propiedades psicométricas de la prueba de inteligencia Otis en estudiantes de Educación mención Orientación de la Universidad del Zulia. Fue un trabajo de campo, donde se hicieron aplicaciones colectivas a 388 estudiantes de todos los semestres, con edad promedio de 21 años. La confiabilidad se calculó a través del método de división por mitades y la consistencia interna Alpha, resultando coeficientes de 0.81 y 0.84 respectivamente. El coeficiente de estabilidad de la prueba se determinó para 88 casos con resultado de 0.44. La validez concurrente se determinó correlacionando los puntajes con otras pruebas de inteligencia Raven, Purdue y Dominó, así como con el rendimiento académico, siendo los resultados de 0.56 con Raven, 0.47 con Purdue, 0.57 con Dominó, y 0.36 con rendimiento, interpretándose éstas como significativas. Se obtuvieron normas para la interpretación de los puntajes en estos estudiantes universitarios.

\* Este artículo corresponde al del proyecto de investigación: Estudio comparativo de los resultados de pruebas de inteligencia, adscrito al Consejo de Desarrollo Científico y Humanístico (CONDES) a través del Centro de Investigaciones Pedagógicas (CEDIP).

\*\* Psicóloga. Magíster en Orientación. Doctora en Ciencias Gerenciales. Profesora Titular del Departamento de Psicología, Facultad de Humanidades y Educación de la Universidad del Zulia. PPI, Nivel II. 0261-7596243. marsanchezg@cantv.net, marsanchezg@gmail.com

\*\*\* Psicóloga. Especialista en Metodología de la Investigación. Magíster en Gerencia de Empresas. Doctora en Ciencias Humanas. Profesora Titular del Departamento de Investigación de la Facultad de Humanidades y Educación de la Universidad del Zulia. PPI, Nivel I. ligiapirela@hotmail.com, ligia.pirela@gmail.com.

Dirección de correspondencia: Av. Guajira, sector Ziruma. Facultad de Humanidades y Educación, Escuela de Educación, Departamento de Psicología. Teléfono: 0261-7596243 Maracaibo, Estado Zulia, Venezuela.

**Recibido:** 12/01/09 • **Aceptado:** 30/07/09

Como conclusión, la prueba Otis es confiable, posee validez concurrente con otras pruebas de inteligencia y el rendimiento académico y se establecieron normas adaptadas a esta población.

**Palabras clave:** Inteligencia, Otis, confiabilidad, validez, normas.

## *Psychometric properties of the intelligence test Otis in students of Education mention Orientation*

### **Abstract**

The purpose of this investigation was to determine the psychometric properties of the intelligence test Otis in students of Education mention Orientation in the University of Zulia. It was a fieldwork, where collective applications were done to 388 students of all semesters, with average age of 21 years. The reliability was calculated with method of division by halves and Alpha internal consistency, turning out to be coefficients of 0.81 and 0.84 respectively. The coefficient of test stability decided for 88 cases with result of 0.44. The competing validity decided correlating the rates with other intelligence tests Raven, Purdue and Dominoes, as well as with the academic performance, being the results of 0.56 with Raven, 0.47 with Purdue, 0.57 with Dominoes, and 0.36 with performance, being interpreted these as significant. There were obtained procedures for the interpretation of punctuations in these university students. As conclusion, the test Otis is reliable, possesses competing validity with other intelligence tests and the academic performance and there were established procedures adapted to this population

**Key words:** Intelligence, Otis, reliability, validity, procedure.

### **Introducción**

La inteligencia ha sido un constructo clave en la psicología y en la evaluación psicológica, lo que hace necesario examinar de manera exhaustiva las pruebas de inteligencia utilizadas, tomando en cuenta la inmensa responsabilidad de los profesionales que administran dichos instrumentos (Cohen y Swerdlik, 2006).

Entre éstos la prueba Otis-B, una herramienta de aplicación grupal para cuantificar la inteligencia, que constituye, según Hogan (2004), uno de los sucesos de mayor influencia en la historia de la psicometría. Es una descen-

diente directa de la primera prueba aplicada a grupos, de uso generalizado, cuando un grupo de psicólogos de la Asociación Americana de Psicología contribuyó a procesar reclutas militares para participar en la Primera Guerra Mundial, en 1917.

Uno de dichos psicólogos, Arthur Otis, se interesó en convertir la prueba Stanford Binet, de uso en el contexto educativo, en un instrumento de aplicación a grupos y calificada con objetividad, derivándose, con la versión verbal y no verbal, de la prueba Army Alpha y Beta, administrada a un extenso número de militares. Posteriormente, se dispuso de las pruebas para uso general, denominándolas Escala de Inteligencia Otis para grupos, la cual es referencia obligada, junto con la Stanford-Binet, cuando se desea establecer el uso de una sola calificación para representar la inteligencia.

La prueba Otis-B ha sido utilizada en el proceso de exploración aptitudinal de la mención Orientación, en la Escuela de Educación, Facultad de Humanidades y Educación de la Universidad del Zulia.

Esto contempla una serie de enunciados, con alternativas de respuestas, a seleccionar sólo una de ellas, que genera una puntuación única, indicador del nivel intelectual de la persona examinada. El contenido de los reactivos se basa en información aprendida, producto del aprendizaje y la experiencia, descrito, según Cattell, como inteligencia cristalizada.

Sánchez de Gallardo y Pirela de Faría (2007a) han estudiado los instrumentos utilizados en el proceso de ingreso antes mencionado, en el cual se aplican de manera colectiva pruebas psicológicas en las áreas intelectual, emocional-social, de personalidad y una entrevista individual, en concordancia con el perfil de ingreso expresado en el Diseño Curricular de la Escuela de Educación (1995) y el perfil de ingreso a la mención Orientación.

Las autoras efectuaron una investigación en la que se analizaron cuantitativamente las propiedades psicométricas de las pruebas aplicadas, tomando en cuenta la importancia de efectuar una evaluación psicológica con instrumentos confiables, válidos y con normas locales. Para ello procesaron los resultados de los aspirantes aptos, desde julio de 1997 hasta mayo de 2004. El coeficiente de confiabilidad obtenido fue de 0.83 considerado un moderado a alto nivel de consistencia, la validez predictiva con el rendimiento académico resultó una correlación de 0.268 con probabilidad de 0.007, significativa a nivel de 0.01, clasificada como media baja y significativa, por lo que puede considerarse el resultado del test un predictor del rendimiento académico. Se obtuvieron normas de interpretación estandarizadas para ese grupo.

Asimismo, se compararon los resultados de quienes ingresan luego de culminar sus estudios de ciclo diversificado y quienes se integran utilizando la modalidad de cambios de carrera, mención de Educación, reincorporación, doble carrera o aquellos que han culminado estudios de técnico superior universitario y desean continuar en Orientación, este último grupo evidenció calificaciones universitarias significativamente más bajas. No se encontraron diferencias significativas en nivel de inteligencia e indicadores emocionales (corrección adolescente), mientras que la corrección de adultos reveló una media significativamente mayor de indicadores emocionales en quienes ingresan por modalidad. En la prueba de personalidad, se evidenciaron diferencias significativas en las escalas Afecto Interpersonal (A), Dominancia (E), Conformismo (G), Audacia (H), Susplicacia (L), Imaginación (M) y Autodisciplina (Q3), mientras que en las restantes no se encontraron diferencias importantes: Inteligencia (B), Estabilidad Emocional (C), Impulsividad (F), Sensibilidad (I), Sagacidad (N), Inseguridad (O), Radicalismo (Q1), Autosuficiencia (Q2) y tensión (Q4) (Sánchez de Gallardo y Pirela de Faría, 2006a).

Posteriormente, se contrastaron los resultados de las pruebas de los aspirantes aprobados y no aprobados. En la prueba de inteligencia se encontraron diferencias significativas entre los grupos, siendo el aprobado el que obtuvo un puntaje promedio mayor, mientras que el grupo no aprobado en dicha prueba se ubica en el rango inferior al término medio.

En cuanto a la prueba socio emocional, tanto en la corrección de adolescentes como de adultos, igualmente se encontraron diferencias significativas entre los grupos, siendo mayor el número de indicadores socioemocionales en el grupo que no aprueba.

Con respecto a la prueba de personalidad, se encontraron diferencias significativas entre los grupos, en los factores: expresividad emocional, lealtad grupal, actitud situacional, e impulsividad, donde el promedio de los puntajes del grupo no aprobado fueron mayores y en los factores dominancia, confianza y posición social, el grupo aprobado obtuvo un puntaje promedio mayor. No se encontraron diferencias significativas entre los grupos en: inteligencia, estabilidad emocional, temperamento, realismo, convencionalismo, franqueza, ansiedad, dependencia grupal y frustración, considerándose que estos factores de la prueba no discriminan entre los estudiantes aprobados y no aprobados (Sánchez de Gallardo y Pirela de Faría, 2006b).

Las investigaciones previas han arrojado resultados que apoyan el uso de la prueba de inteligencia Otis utilizada en el proceso de exploración aptitudinal, por ser confiable, con validez predictiva con rendimiento académico,

haberse diseñado un baremo local, presentar diferencias significativas entre los aspirantes aptos y no aptos. Sin embargo, es necesario contar con datos actualizados de las propiedades psicométricas de confiabilidad, validez y normas de la prueba Otis, confirmando la utilidad de dicha prueba, tomando en cuenta la importancia concedida al resultado de la medición de inteligencia, en el proceso de ingreso a la mención antes descrita, lo cual se plantea como objetivo en la presente investigación.

## **Desarrollo Inteligencia**

Desde que la psicología existe como disciplina, los psicólogos han tenido diferencias acerca de la definición de la inteligencia, así como del mejor método para medirla. Cohen y Swerdlik (2006) la refieren como una capacidad multifacética que se manifiesta de diferentes maneras a lo largo del ciclo vital. Señalan que incluye las habilidades de adquirir y aplicar el conocimiento, razonar de manera lógica, planear de modo efectivo, hacer inferencias a partir de la percepción, realizar juicios sólidos y resolver problemas, comprender y visualizar conceptos, poner atención, ser intuitivo, encontrar con facilidad las palabras y pensamientos correctos, enfrentarse, ajustarse y aprovechar situaciones nuevas.

Santrock (2006) describe la inteligencia como la habilidad de solución de problemas y la capacidad de adaptarse y aprender de las experiencias cotidianas de la vida. Morris y Maisto (2005) la enuncian como un término general relacionado con la habilidad o habilidades implicadas en el aprendizaje y la conducta adaptativa. Papalia y Olds (2004) utilizan el término para referirse a una constante interacción activa entre las capacidades heredadas y las experiencias ambientales, cuyo resultado capacita al individuo para adquirir, recordar y utilizar conocimientos, entender conceptos concretos y eventualmente abstractos, comprender las relaciones entre los objetos, los hechos y las ideas, así como aplicarlo con el propósito específico de resolver los problemas de la vida cotidiana.

Por lo tanto, Santrock (2006) considera que siendo un concepto tan abstracto y amplio, no es sorprendente encontrar tantas definiciones sobre ella. Asimismo este autor, Santrock (2004), refiere que el problema de la inteligencia es la dificultad para observarla y medirla en forma directa, ya que sólo puede evaluarse de manera indirecta, al estudiar y comparar los actos inteligentes realizados por las personas.

Además de diversas definiciones, existen diferentes teorías sobre inteligencia, las cuales influyen significativamente en el contenido de los tests de inteligencia. Existen dos teorías clásicas sobre inteligencia, que han predominado en los estudios sobre la naturaleza de la inteligencia.

Hogan (2004) describe al inglés Charles Spearman, (1904, 1927), como el creador de lo que generalmente se considera la primera teoría formal sobre la capacidad mental humana. Spearman basó su teoría en el examen de correlaciones entre una extensa variedad de pruebas de funciones sensoriales simples. Consideraba que dichas correlaciones eran lo suficientemente elevadas para concluir que el desempeño en estas pruebas dependía mayoritariamente de una capacidad mental general. A esta capacidad general denominó “g”. Por supuesto, las correlaciones entre dichas correlaciones no eran perfectas, ya que cada prueba poseía una varianza única o específica, independiente de “g”, así que cada conjunto de pruebas poseía una serie de factores “s” y un factor “g”. Dado que la teoría cuenta con estos dos tipos de factores, Spearman la denominó teoría bifactorial; no obstante, el factor esencial de la teoría es la “g” y suele conocerse como la teoría de la inteligencia unifactorial o “de un solo factor”. Morris y Maisto (2005) refieren que Spearman señalaba que los individuos buenos en un área, frecuentemente destacan en otras áreas. Una persona inteligente comprende con rapidez, toma decisiones acertadas, sostiene conversaciones interesantes y, en general, se comporta inteligentemente en una amplia variedad de situaciones. Spearman señalaba que cada individuo puede ser más rápido en unas áreas que en otras, producto de la manera como la misma inteligencia general subyacente se revela en diferentes actividades.

Hogan (2004) describe cómo, difiriendo con Spearman, el psicólogo Thurstone planteó una teoría que constituyó la principal competencia de la teoría “g”, quien estimaba que las correlaciones eran lo suficientemente bajas como para considerar que median varios factores bastante independientes. El autor argumentaba que la inteligencia está integrada por nueve habilidades mentales distintas: habilidad espacial, rapidez perceptual, habilidad numérica, significado verbal, memoria, fluidez verbal, razonamiento, inducción y deducción. Aunque Thurstone no fue el único en proponer una teoría multifactorial de la inteligencia, su obra es considerada la que mayor influencia ejerció.

Cohen y Swerdlik (2006) señalan que posteriormente el psicólogo Cattell en 1971 en contraste con Thurstone planteó la existencia de dos grupos de habilidades mentales únicamente: inteligencia cristalizada e inteligencia fluida. La inteligencia cristalizada (Gc) la constituyen habilidades y conocimientos adquiridos, dependientes de la exposición a una cultura particular, así

como a la educación formal e informal. La inteligencia fluida (Gf), integrada por habilidades no verbales, relativamente independientes de la cultura, así como de cualquier tipo de instrucción específica. Dicha teoría ha recibido gran atención por parte de los investigadores y usuarios de pruebas.

Más recientemente, Robert Sternberg en 1986 propuso una teoría triárquica de la inteligencia humana, en la cual señala que la inteligencia se manifiesta en tres formas principales: analítica, creativa y práctica. La primera es descrita como la capacidad para analizar, juzgar, evaluar, comparar y contrastar; la creativa se refiere a las habilidades para crear, diseñar, inventar, originar e imaginar, y la práctica abarca las habilidades para utilizar, aplicar, implementar y poner en práctica. De manera que algunos estudiantes son capaces en las tres áreas, otros sólo en una o dos. Los estudiantes con patrones triárquicos diferentes se comportan de manera distinta en las instituciones educativas, por lo cual resulta conveniente que las tareas asignadas por los docentes requieran la combinación de estas habilidades y que dichos educadores equilibren la instrucción con relación con los tres tipos de inteligencia, más explícitamente deben brindar oportunidades a los estudiantes para aprender a través del pensamiento analítico, creativo y práctico (Santrock, 2006).

Adicionalmente, surge una influyente alternativa a la teoría de la inteligencia de Sternberg, expuesta por Howard Gardner y colaboradores en 1983, denominada de las inteligencias múltiples. Plantea la inteligencia conformada por muchas capacidades separadas o inteligencias múltiples, cada una de las cuales es relativamente independiente de las otras, porque diferentes partes del cerebro procesan distintas clases de información y expone ocupaciones en las que cada tipo de inteligencia se consideran ventajas.

Gardner (citado por Papalia, Olds y Feldman, 2005) describe ocho inteligencias: Lógico-Matemática, Lingüística, Espacial, Musical, Cinestésico-corporal, Interpersonal, Intrapersonal, así como Habilidades naturalistas. Posteriormente, Gardner (citado por Kail y Cavanaugh, 2006) añade una novena inteligencia, la existencialista, en la cual se analizan asuntos “supremos”, como el propósito de la vida y la naturaleza de la muerte.

Papalia, Olds y Feldman (2005) señalan que para Gardner un grado de inteligencia elevado en un ámbito no necesariamente va acompañado de un grado de inteligencia elevado en cualquiera de los otros. Aunado a esto, Papalia y Olds (2004) enuncian que para el autor antes señalado cada una de las ocho inteligencias puede ser destruida por un daño cerebral y además, como implican habilidades cognoscitivas únicas, se pueden mostrar de un modo exagerado en personas dotadas como en individuos con retardo mental o autismo.

Al comparar estos diferentes planteamientos sobre la inteligencia, Spearman creía que las personas difieren en cuanto al grado en que poseen la “energía mental” denominada inteligencia general. Teóricos posteriores, como Thurstone y Cattell, especificaron la estructura de las capacidades mentales, mientras que dos teóricos contemporáneos de gran influencia, Sternberg y Gardner, presentan teorías similares con énfasis en las habilidades prácticas manifestadas en el mundo real y difieren en asuntos básicos: Gardner se basó fundamentalmente en una aproximación por medio de un estudio de casos y exploró el desarrollo de una inteligencia específica de una persona en particular, mientras que Sternberg mostró su ingenio en la creación de tests mentales para medir distintos aspectos de la inteligencia. Otra comparación importante se refiere a que las teorías de Spearman, Thurstone y Gardner tienden a describir la manera como los individuos difieren en el contenido de la inteligencia, las diferentes capacidades subyacentes en la conducta inteligente, en cambio, Sternberg, en sus trabajos más recientes sobre psicología cognoscitiva, ha enfatizado los procesos de pensamiento, que pueden ser comunes a todas las personas, en cuanto a cómo reúnen y utilizan información los humanos para resolver problemas y comportarse con inteligencia.

También Gardner (citado por Santrock, 2004) señala que su teoría, como la de Sternberg, incluye una o dos categorías de inteligencia social. En los planteamientos de Gardner, las categorías son inteligencia interpersonal e inteligencia intrapersonal. En la teoría de Sternberg, la categoría es inteligencia práctica. En la actualidad se insiste en otra teoría que resalta la importancia de los aspectos interpersonal, intrapersonal y práctico, denominada inteligencia emocional, ampliamente difundida por Goleman. Reviste especial importancia el estudio de las teorías de inteligencia, dado que determinan el contenido de los test de inteligencia y de habilidades mentales, los cuales se utilizan con frecuencia en la evaluación de las personas.

## **Pruebas de inteligencia**

En la actualidad todavía se discute qué es la inteligencia y se evalúan los diferentes enfoques para describirla. Sin obviar esta diatriba, Kail y Cavanaugh (2006) plantean como innegable que las personas se distinguen por su capacidad intelectual y existe un elevado número de pruebas diseñadas para medir dichas diferencias. Desde 1904, el Ministro de Instrucción Pública de Francia solicitó a dos famosos psicólogos de la época, Alfred Binet y Theodophile Simon, construir un instrumento para identificar a los niños que no aprenderían en la escuela convencional sin una enseñanza especial, para lo cual eligieron

algunas tareas simples que los niños franceses de diferentes edades deberían ser capaces de realizar.

La edad mental de un niño se refería a la dificultad de los problemas que podía resolver correctamente. Un niño capaz de resolver problemas que un niño normal de siete años podría resolver tendría una edad mental de 7. Binet y Simon usaron la edad mental para distinguir a los niños “brillantes” de los “lentos”. Así, un niño “brillante” tendría la edad mental de otro mayor y un niño “lento” tendría la edad mental de uno de menos edad. Ambos psicólogos confirmaron que los niños “brillantes” identificados mediante su prueba tenían más éxito en la escuela, por lo cual se considera que fue la primera prueba objetiva de inteligencia.

Según Morris y Maisto (2005), durante la década que siguió a la incursión de la Escala Binet-Simon se publicaron numerosas adaptaciones de Binet, siendo la más conocida la Escala de Inteligencia de Stanford-Binet, preparada en la Universidad de Stanford por L. M. Terman en 1916, quien además introdujo el término Coeficiente Intelectual o CI, para establecer un valor numérico de la inteligencia, asignando la puntuación de 100 a una persona de inteligencia promedio. Desde 1916, la Stanford-Binet se ha revisado en diferentes ocasiones, considerándose más apropiada para niños, adolescentes y adultos muy jóvenes, aunque puede utilizarse con gente mayor.

En la actualidad, junto a la prueba Stanford-Binet, se utilizan los instrumentos creados por David Wechsler, las pruebas WISC y WAIS, considerada esta última de uso frecuente en adultos. Wechsler desarrolló la prueba fundamentalmente porque deseaba un instrumento que fuera más adecuado para adultos que la Stanford-Binet y obtuvo como resultado calificaciones separadas para la parte verbal y de desempeño, así como un CI global. Las pruebas antes mencionadas son pruebas individuales, donde el examinador lleva a la persona a una habitación aislada y pasa entre 30 y 90 minutos aplicando la prueba, posteriormente el examinador puede tardar luego otra hora para calificar la prueba de acuerdo con las detalladas instrucciones del manual, lo cual es una operación costosa que exige tiempo, además el criterio y la conducta del examinador pueden tener una influencia considerable en la calificación.

Aunado a esto, refiere Hogan (2004), se desarrolló la primera prueba de inteligencia aplicada a grupos, en el contexto de la selección de reclutas militares, de Estados Unidos de Norteamérica, para participar en la Primera Guerra Mundial. Se solicitó la asistencia de una comisión de la American Psychological Association, liderada por Lewis Terman, quien acababa de producir la Stanford-Binet de aplicación individual. A Arthur Otis, en ese momento

estudiante de doctorado de Terman, le interesaba fundamentalmente convertir la Stanford-Binet en una prueba de administración colectiva y emprendió dicha tarea para su disertación doctoral. Como resultado, se produjo la versión verbal y no verbal de la prueba Army Alpha and Beta, respectivamente, utilizada con cerca de dos millones de militares. Luego, se usó para la población general, titulada la Escala de Inteligencia Otis para grupos, las cuales en el presente se han ido actualizando y empleando cuando se desea usar una sola calificación para cuantificar la inteligencia.

Luego de comentar sobre el surgimiento de las pruebas grupales, Cohen y Swerdlik (2006) las describe como herramientas útiles para la exploración cuando se debe evaluar a una gran cantidad de personas de forma simultánea o en un período breve. La utilizada en este estudio, OTIS, corresponde a este tipo. De ser requerido, los datos obtenidos se podrán explorar con mayor profundidad mediante métodos de evaluación más individualizados.

Señalan como ventajas de las pruebas grupales el uso eficiente del tiempo y de los recursos, reactivos de la prueba ubicados en un formato fácil de calificar por una persona o una máquina, tiempo de trabajo independiente, a su propio ritmo; el aplicador de la prueba no necesita una capacitación rigurosa, ya que la tarea podría requerir tan sólo leer las instrucciones, medir el tiempo y supervisar a los evaluados; conjuntamente puede tener menos efecto en la puntuación del evaluado que en una situación de uno a uno; son menos costosas que las individuales, se ha comprobado su valor para propósitos de exploración, por lo general están normalizadas con base en grandes cantidades de personas y funcionan bien con personas que pueden leer, seguir instrucciones, tomar un lápiz y que no requieren mucha asistencia.

Según Aiken (2003), muchas pruebas de inteligencia colectiva se han diseñado específicamente con el propósito de medir la aptitud para el trabajo académico, por lo que se les designa pruebas de aptitud académica. Algunas pruebas de inteligencia en grupo tienen un enfoque más amplio que éste, pero aun así, su contenido es similar al de las medidas de habilidad académica: tienen un gran contenido de reactivos verbales, numéricos y otros de tipo escolar. A lo largo del tiempo, se han utilizado distintas pruebas con propósitos de admisión a las universidades e instituciones profesionales.

En este sentido, Morris y Maisto (2005) señalan que las pruebas de inteligencia tienen una fuerte correlación con las calificaciones escolares, lo cual se vincula con lo planteado por Kail y Cavanaugh (2006) al señalar que en términos generales las pruebas logran predecir el éxito escolar en forma razonable. Estos últimos también refieren que las pruebas de inteligencia no

sólo predicen bastante bien el desempeño escolar, sino también el del trabajo, sobre todo en puestos de elevada complejidad.

No obstante, los usos de tests de inteligencia han dado lugar a una gran cantidad de controversias respecto a la naturaleza y significado de la inteligencia y las consecuencias personales y sociales que se confía dependen de estas pruebas. Por ejemplo, una de las críticas más reiteradas, según expone Santrock (2004), se refiere a que reflejan los valores de la cultura dominante, tales como la importancia del tiempo, idioma predominante entre diferentes grupos lingüísticos, culturas con experiencias en dibujos y fotografías, diversas actitudes, valores, motivación. También refiere que la mayoría de los reactivos de las pruebas de inteligencia consisten en habilidades e información aprendida y en consecuencia, personas provenientes de ambientes con privaciones tienen pocas posibilidades de haber podido adquirir los conocimientos que se requieren. Asimismo, Santrock (2003) expone dificultades en la aplicación, más que todo en las administradas grupalmente, ya que el examinador no puede establecer una relación personal, ni determinar el nivel de ansiedad de los participantes, pudiendo ocurrir que algunos estudiantes no entiendan las instrucciones o se distraigan con los compañeros de aplicación.

Sin embargo, a pesar de lo antes descrito, Aiken (2003) expone como méritos de las pruebas de inteligencia su importante contribución en la identificación de individuos situados en todos los niveles de las habilidades mentales y en el diseño de programas para tratarlos, entrenarlos y educarlos, tales como personas con retardo mental o superdotados; así como también para diferenciar las habilidades en un solo individuo.

Además, las pruebas suministran una o más puntuaciones o proporcionan evidencia de que una persona pertenece a una categoría y no a otra. La suposición implícita es que las pruebas miden las diferencias individuales en cuanto a rasgos o características que existen. En la mayor parte de los casos se supone que todas esas personas poseen ese rasgo o característica medida, aunque en diferentes cantidades, por cuanto el propósito de la prueba consiste en estimar la cantidad del rasgo o cualidad que posee el individuo. En este aspecto es importante efectuar dos advertencias: la primera, toda puntuación de prueba siempre reflejará cierto grado de error de medición, por lo que quien desarrolla la prueba debe procurar que dicho error sea muy pequeño, aunque nunca se le puede eliminar por completo, como tampoco se puede conocer su impacto exacto sobre un caso individual. La segunda, los resultados no presentan una "cosa" que tenga realidad física, por lo común reflejan una abstracción que se ha demostrado tiene utilidad para pro-

nosticar conductas externas a la prueba, como por ejemplo el rendimiento académico (Gregory, 2001).

## **Propiedades psicométricas**

Hernández, Fernández y Baptista (2006) definen la medición como el proceso de vincular conceptos abstractos con indicadores empíricos. Dicha definición incluye dos consideraciones, la primera, desde el punto de vista empírico, se refiere a que el centro de atención es la respuesta observable y la segunda, desde la perspectiva teórica, se vincula con que el interés se sitúa en el concepto subyacente no observable, representado por la respuesta. En el mencionado proceso, el instrumento de medición o de recolección de los datos desempeña un papel fundamental y son descritos por los autores como el recurso que utiliza el investigador para registrar información o datos sobre las variables de interés. Asimismo, Rodríguez y Pineda (2003), citados por Pelekais, Finol, Newman y Parada (2005), lo describen como el recurso del que se vale el investigador para acercarse a la realidad y extraer información sobre su objeto de estudio. Dichos autores también señalan su utilidad en el registro de datos objetivos o subjetivos, primarios o secundarios, que se encuentran en el contexto estudiado.

Pelekais *et al.* (2005), por su parte, enuncian la importancia de cumplir con el proceso técnico de confiabilidad y validez en toda prueba o instrumento de carácter científico.

Ningún instrumento de medición puede tener valor a menos que mida algo en forma consistente o confiable. Por ello, otro de los primeros aspectos que deben determinarse es si las pruebas son lo suficientemente confiables o no para medir aquello para lo cual se creó.

Chávez (2007) define la confiabilidad como el “grado de congruencia con que se realiza la medición de una variable”. Si en ausencia de cualquier cambio permanente en una persona (producto del crecimiento, aprendizaje, enfermedad o accidente) las calificaciones de las pruebas varían en gran medida de tiempo en tiempo o en diferentes situaciones, es probable que la prueba no sea confiable y no pueda utilizarse para explicar o realizar predicciones sobre el comportamiento de las personas (Aiken, 2003).

Münch y Ángeles (2003: 55) definen la confiabilidad como “la estabilidad, consistencia y exactitud de los resultados”, considerándola siempre como un intento por procurar que los resultados obtenidos por el instrumento sean similares si se vuelven a aplicar sobre las mismas muestras en igualdad de condiciones.

Santrock (2004) señala que una prueba estable y consistente no debe fluctuar de manera significativa debido a factores aleatorios como: cuánto durmió la noche anterior la persona que respondió la prueba, quién es el examinador o cuál es la temperatura del lugar donde se aplica la prueba.

Según Papalia y Olds (2004), se han diseñado varios métodos estadísticos para estimar el grado de confiabilidad de las mediciones: fiabilidad test-retest, formas paralelas y de división por mitades. En la fiabilidad test-retest, se aplica la misma prueba al mismo grupo o persona más de una vez. El método de formas alternativas o paralelas del test, con elevadas similitudes en formato, contenido y nivel de dificultad. El otro método utilizado en esta investigación es el método de las dos mitades, mitades partidas o *split-halves*, el cual requiere una sola aplicación, especificado por Hernández *et al.* (2006) como la comparación entre las dos mitades del conjunto total de ítems o componentes. Gregory (2001) expone su utilidad para estimar la correlación de los pares de puntuaciones obtenidas de mitades equivalentes de una prueba administrada en una única oportunidad a una muestra representativa de examinados, específicamente consiste en calcular una “r” de Person entre las puntuaciones de las dos mitades equivalentes de la prueba, e implica un paso adicional: ajustar la confiabilidad utilizando la fórmula Spearman–Brown.

Asimismo, el coeficiente alfa consiste en la media de todos los posibles coeficientes por mitades, corregidos según la fórmula Spearman–Brown, la cual es un índice de consistencia interna de los reactivos, es decir, la tendencia a correlacionarse en sentido positivo entre sí. Por lo general, el test se divide en dos partes, correspondientes a las preguntas pares e impares, dicho método genera una medida de consistencia interna del test.

En esta investigación se utiliza este último método descrito para comprobar la confiabilidad de la aplicación de la prueba Otis, al ser administrada a estudiantes de Educación, mención Orientación.

La confiabilidad es un elemento necesario, pero no suficiente, de una buena prueba; de manera que además de ser confiables, las pruebas deben ser precisas de una manera razonable; es decir, deben ser válidas (Cohen y Swerdlik, 2006).

La validez, Hernández *et al.* (2006) la definen como “el grado en el que un instrumento realmente mide la variable que pretende medir”. De dicho concepto pueden presentarse distintos tipos de evidencia, la vinculada con el contenido, la relacionada con el criterio y la referida al constructo.

La validez relacionada con el criterio, según Cohen y Swerdlink (2006), es un juicio de cuán adecuadamente puede ser utilizada la puntuación de una prueba para inferir la posición más probable de un individuo con respecto a cierta medida de interés, siendo el criterio esa medida de interés. También Hernández *et al.* (2006) señalan que se mide dicha validez comparándola con algún criterio externo; dicho criterio es un estándar con el que se juzga la validez del instrumento. Mientras los resultados del instrumento de medición se relacionen más al criterio, la validez de criterio será mayor.

Se exponen dos tipos de evidencia de la validez relacionada con el criterio: la validez concurrente y la predictiva. La validez concurrente, para Cohen y Swerdlink (2006), consiste en un índice del grado en que se relaciona la puntuación de una prueba con alguna medida de criterio obtenida al mismo tiempo que la puntuación. Las declaraciones de validez concurrente indican el grado en que las puntuaciones de una prueba pueden servir para estimar la posición actual de un individuo frente a un criterio. Una vez que se ha establecido la validez de inferencia de las calificaciones, la prueba puede proporcionar una forma más rápida y menos costosa para ofrecer una decisión de clasificación o un diagnóstico, lo cual la hace atractiva para futuros usuarios, en cuanto a disminuir dinero y tiempo profesional.

En cuanto a la validez predictiva, según Hernández *et al.* (2006), consiste en un índice del grado en que las puntuaciones de una prueba pronostican alguna medida de criterio en el futuro. En la presente investigación, se determinó la validez concurrente de la prueba Otis.

Para Feldman (2006) la validez y la confiabilidad de una prueba son requisitos para una evaluación precisa de inteligencia, así como para cualquier otra medición que realicen los psicólogos. Reseña que suponiendo que una prueba es válida y confiable, se necesita interpretar el significado de la puntuación: el establecimiento de normas. Las normas son criterios de desempeño en las pruebas que permiten comparar la puntuación de una persona en una prueba, con las puntuaciones de otras personas a quienes se les administró la misma prueba. Aquellas pruebas para las que se han desarrollado normas se denominan pruebas estandarizadas.

El autor plantea que para el desarrollo de las normas, los diseñadores de pruebas calculan la puntuación promedio de un determinado grupo para el cual se elaboró el instrumento y determinan en qué grado difiere la puntuación de cada persona con respecto a las puntuaciones de los demás que se sometieron anteriormente a la prueba, de manera que las personas evaluadas están en posibilidades de considerar el significado de sus puntuaciones naturales en

relación con la de otras que también hicieron la prueba, lo cual brinda una idea cualitativa de su desempeño.

Es importante destacar que las muestras de las personas evaluadas empleadas en el establecimiento de normas son cruciales para el proceso de normalización, tomando en cuenta que la gente cuyos valores sirven para determinar las normas debe ser representativa de los individuos a quienes se dirige la prueba.

Al respecto, existen diferentes grupos normativos, enunciados por Hogan (2004) como: normas nacionales, normas internacionales, grupos normativos convenientes, normas de usuarios, normas de subgrupos, normas locales y normas institucionales. Las normas nacionales son aquellas basadas en un grupo que sea representativo del segmento de la población del país para la cual se ideó la prueba: adultos, niños en octavo grado, personas que quieran ingresar a la universidad. Por lo general, el grupo o población meta se define junto con la finalidad de la prueba. Las internacionales, provenientes de varios países, quienes optaron por participar en estos estudios, por lo general de naciones económicamente avanzadas; los grupos normativos convenientes, se refieren a grupos que provienen de un solo lugar geográfico, por lo general homogéneos en cuanto a antecedentes culturales y quizá limitados en términos de rangos de edad, nivel educativo y otras variables importantes.

También, las normas de los usuarios, basadas en cualquier grupo que en realidad se somete a la prueba, por lo general en un periodo específico. No es un intento a priori por asegurar que el grupo sea representativo de alguna representación bien definida. Constituyen, en realidad, una modalidad de norma conveniente. Además, las normas de subgrupos extraídos del grupo normativo en su totalidad, por ejemplo, sexo, raza, grupo socioeconómico, grupo ocupacional o región geográfica, útiles cuando hay diferencias sustanciales entre los subgrupos en la variable que mide la prueba. Las normas locales, provenientes de personas que están en situación local, por lo general presentadas en percentiles y las normas institucionales, sustentadas en promedios de individuos dentro de organizaciones. En resumen, dependiendo del objetivo de la evaluación psicológica, es posible preferir utilizar un tipo de norma.

## **Metodología**

La presente es una investigación de tipo descriptiva, correlacional, de campo, diseño no experimental, transeccional.

Según Hernández, Fernández y Baptista (2006), el estudio se cataloga de tipo descriptivo, explicado como aquellos donde se recolectan datos sobre diversos conceptos, aspectos, dimensiones o componentes del fenómeno a estudiar, con el propósito de especificar propiedades, características y rasgos importantes y describir tendencias de un grupo, en este caso los resultados de la aplicación de la prueba de inteligencia en estudiantes de Educación de la Universidad del Zulia. También se considera correlacional, ya que en el mismo se miden dos o más aspectos o variables, en los cuales se pretende conocer si están o no relacionados en los mismos sujetos y luego se analizó la correlación, más explícitamente, la vinculación entre los resultados de la administración grupal de las pruebas Otis, Raven, Pordue y Dominó en los alumnos antes mencionados, así como el rendimiento académico recopilado de la base de datos del Centro de Computación SADIA de la Universidad del Zulia y suministrados por la Secretaría Docente de la Facultad de Humanidades y Educación.

Asimismo se califica como de campo, según lo declarado por Arias (2006), ya que la recolección de datos se realiza directamente en los sujetos investigados o de la realidad donde ocurren los hechos, particularmente en el interior de los salones de la facultad de Humanidades de la Universidad antes mencionada, donde los estudiantes asisten a sus clases presenciales; sin manipular o controlar las variables, de allí su carácter de investigación no experimental.

La población estuvo conformada por 326 estudiantes de uno u otro sexo, de la mención Orientación de la Escuela de Educación que asisten a las materias que administra el Departamento de Psicología, cursantes de la totalidad de los semestres de la carrera, con edad promedio de 21 años, con diversos años de escolaridad en la Universidad, desde primero a décimo semestre, con diferentes circunstancias de ingreso a la Universidad del Zulia: vía Consejo Nacional de Universidades o C.N.U. y modalidades de cambios de carrera, equivalencias, carreras simultáneas.

En cuanto a la técnica e instrumento de recolección de datos, se administró de manera grupal la prueba de inteligencia Otis, se leyó con detenimiento las instrucciones, en cuanto a cómo responder los ítems y el tiempo de culminación de la prueba. Se constató que los participantes comprendieran las instrucciones, mediante la ejecución de los ejemplos correspondientes.

La prueba Otis (forma B) es una prueba estandarizada que se aplica de manera grupal integrada por un cuadernillo con los enunciados y alternativas de respuestas, entre las cuales el individuo debe escoger sólo una; así como

una hoja de respuestas donde se deberán escribir. Consta de tres ejemplos y 75 ítems. Se evalúa con una plantilla de corrección y el puntaje es 1 ó 0, resultando como máxima puntuación, 75. El tiempo máximo permitido es de 45 minutos.

La prueba Raven, también está compuesta por un cuadernillo donde aparecen las figuras a las cuales les falta una parte y el sujeto debe seleccionar entre las alternativas planteadas. Además contempla una hoja aparte para anotar las respuestas y se evalúa con una plantilla de corrección. El puntaje es 1 ó 0, resultando como máxima puntuación, 60. No presenta límite de tiempo.

Las pruebas Purdue y Dominó ( D-48) son instrumentos estandarizados que consisten en un cuadernillo donde aparecen las figuras correspondientes; para el primero se identifica entre cinco figuras, aquella que es completamente diferente a las otras, para el segundo se anota la figura que debe continuar, en similitud a las secuencias de los modelos que aparecen. Ambos se corrigen con una plantilla de corrección, siendo el puntaje 1 ó 0, resultando como mayor puntuación, 48. El tiempo máximo permitido es de 45 minutos. Cada una de las cuatro pruebas mencionadas se aplica en una ocasión diferente y resultan en una única calificación de inteligencia.

Las pruebas Raven, Purdue y Dominó de acuerdo con el contenido de sus reactivos podrían vincularse con la inteligencia fluida, expuesta por Cattell, mientras que el Otis B se relaciona con la inteligencia cristalizada.

## **Resultados**

Para el análisis de los datos aportados por los sujetos de este estudio se aplicó un análisis inferencial utilizando el programa SPSS 10.0 para Windows, determinando la confiabilidad por el método de división por mitades y consistencia interna de la prueba por el método de Crombach, la validez de criterio a través de su correlación con el criterio de otras pruebas de inteligencia y el rendimiento académico. Este fue medido a través de los índices o promedios de las materias inscritas por los estudiantes y suministrados por la Secretaría Docente de la Facultad de Humanidades y Educación. Luego se obtuvieron las normas de interpretación para la prueba.

En la tabla 1 se puede apreciar que el coeficiente de división por mitades fue de 0.81 y el de consistencia interna de 0.84, los cuales se interpretan como alto, indicando que la prueba Otis tiene alta confiabilidad, es decir, que existe correlación entre las mitades y los ítems de cada una de las subpruebas.

En cuanto a la estabilidad de la prueba al correlacionar los puntajes de la aplicación de la misma con los puntajes obtenidos al momento de ingresar a la escuela en 88 estudiantes, dio como resultado 0.44, que es media y significativa.

**Tabla 1**  
**Resultados de la confiabilidad de la prueba de inteligencia Otis**

Prueba	División por mitades	Consist. Interna
Otis	0.81	0.84

Estabilidad de la prueba

Correlación	Sig.	No. Casos
0.44	0.00	88

Para determinar la validez de la prueba comparándola con un criterio se correlacionaron los puntajes con los obtenidos en otras pruebas de inteligencia general, también de aplicación colectiva como Raven, Purdue y Dominó, así como con el rendimiento académico. En la tabla 2, se puede apreciar que la correlación entre la prueba de Raven fue de 0.56 media y significativa, con Purdue fue de 0.47, y con Dominó fue de 0.57, también media y significativa. La correlación del rendimiento académico fue de 0.36, media baja y significativa. Por lo tanto se plantea que la prueba es válida para medir inteligencia.

**Tabla 2**  
**Resultados de las correlaciones entre la prueba de inteligencia Otis con otras pruebas y el rendimiento académico**

Prueba	Raven	Purdue	Dominó	Rendim
Otis	R=0.56** Sig. 0.00	R=0.47** Sig. 0.00	R=0.57** Sig. 0.00	R=0.36** Sig. 0.00

A continuación se presentan las normas de interpretación obtenidas para la población de estudiantes de la mención Orientación de la Escuela de Educación de la Universidad del Zulia, el cual se construyó determinando los percentiles 90, 70, 30 y 10 de los datos procesados identificando a qué puntaje corresponde, para luego ser considerados en el baremo de interpretación que se presenta en la tabla 3.

Tabla 3

### Baremos para la interpretación del puntaje total de la prueba de inteligencia Otis para los estudiantes de Educación mención Orientación

Rango	Nivel	Otis
I	Superior (>90)	>40
II	Superior al término medio (70- 90)	35-40
III	Término medio (30-70)	25-34
IV	Inferior al término medio (30 – 10)	17-24
V	Inferior (<10)	<17

## Conclusiones

La información producida al examinar las propiedades psicométricas de la prueba para cuantificar la inteligencia, lo determinan como un instrumento confiable, válido y con normas locales, los cuales son requisitos importantes que debe cumplir toda prueba.

A través de esta investigación se pudo comprobar que la prueba de inteligencia Otis posee datos consistentes, mide el constructo que se pretende medir, al correlacionar con otras pruebas que miden el mismo constructo: Raven 0.56, Purdue 0.47 y Dominó 0.57, ( media y significativa). La correlación del rendimiento académico fue 0.36 baja y significativa. Se elaboró un baremo local, por lo cual se recomienda continuar su aplicación en el contexto educativo para evaluar a estudiantes preferiblemente de la carrera educación.

También se recomienda incluir la elevada cantidad de datos recopilados de investigaciones previas (aptos y no aptos para ingresar a la mención) desde los años 1997 y obtener información psicométrica que confirme su uso en el proceso de exploración aptitudinal de los estudiantes a la mención Orientación.

Se sugiere continuar administrando dicha prueba en diferentes grupos de la población, así como contactar con otras universidades e investigadores interesados y con datos de aplicaciones grupales del Otis, a fin de establecer normas regionales y nacionales.

## **Referencias Bibliográficas**

- AIKEN, Lewis (2003). **Test psicológicos y evaluación**. México. Editorial Pearson. Undécima Edición, pp 154-161.
- ARIAS, Fideas (2006). **El proyecto de investigación**. Venezuela. Editorial Episteme. Cuarta edición, pp.31.
- CHÁVEZ, Nilda (2007). **Introducción a la Investigación Educativa**. Venezuela. Editorial Ars Gráfica. Cuarta edición, pp. 27.
- COHEN, Ronald y SWERDLINK, Mark (2006), **Pruebas y Evaluación Psicológicas**. España, Editorial McGraw-Hill Interamericana. Sexta Edición, pp. 233-262.
- Escuela de Educación de La Universidad del Zulia (1995), **Diseño Curricular**. Maracaibo, Venezuela.
- FELDMAN, Robert (2006). **Psicología con aplicaciones en países de habla hispana**. México, Editorial McGraw-Hill Interamericana. Sexta edición, pp. 282-286.
- GREGORY, Robert (2001). **Evaluación Psicológica: Historia, Principios y Aplicaciones**. México. Editorial El Manual Moderno. Tercera Edición, pp. 36-41-101-102.
- HERNÁNDEZ, Roberto, FERNÁNDEZ, Carlos y BAPTISTA, Pilar (2006). **Metodología de la Investigación**. México. Editorial McGraw-Hill Interamericana. Cuarta edición, pp.102- 277.
- HOGAN, Thomas (2004). **Pruebas Psicológicas**. México, Editorial El Manual Moderno. Primera edición, pp. 75-77-195.
- KAIL, Robert y CAVANAUGH, John (2006). **Desarrollo Humano. Una perspectiva del ciclo vital**. México, Thompson Editores. Tercera Edición, pp. 220-225.
- MORRIS, Charles y MAISTO, Albert (2005). **Introducción a la Psicología**. México, Editorial Pearson Educación. Duodécima Edición, pp. 255-269.
- MUNCH Galindo, Lourdes y ÁNGELES, Ernesto (2003). **Métodos y técnicas de investigación**. México, Editorial Trillas, pp. 55.
- PAPALIA, Diane y OLDS, Sally (2004), **Psicología**. México, Editorial McGraw-Hill Interamericana, pp. 248-249.

- PAPALIA, Diane, OLDS, Sally y FELDMAN (2005). **Psicología del Desarrollo**. México, Editorial McGraw-Hill Interamericana. Novena edición, pp. 371-373.
- PELEKAIS, Cira; FINOL, Mineira; NEUMAN, Noel; PARADA, José. (2005). **El ABC de la Investigación. Una aproximación teórico-práctica**. Venezuela. Ediciones Astro Data. Primera edición, pp. 64- 72.
- SÁNCHEZ DE GALLARDO, Marhilde y PIRELA DE FARÍA, Ligia (2006a). “¿Qué hacer con los resultados de las pruebas psicológicas de estudiantes de educación, mención Orientación?” **Revista PSICODEBATE**, Mayo-Agosto 2007. Volumen 7. Buenos Aires, Argentina, Universidad de Palermo. Pp.169-191.
- \_\_\_\_\_ (2006b), “Comparación de resultados de pruebas de inteligencia, emocional-social y de personalidad utilizadas para el ingreso a la mención orientación, de la Escuela de Educación de la Universidad del Zulia.” **Revista Humanitas**. Septiembre 2006. Volumen 3, N° 3. San José, Costa Rica. Universidad Católica de Costa Rica. Pp 34-44.
- \_\_\_\_\_ (2007a). “La Inteligencia como criterio de selección en el proceso de ingreso a la mención Orientación, Escuela de Educación de la Universidad del Zulia”. **Revista MULTICIENCIAS**, Mayo-Agosto 2007. Volumen 7, N° 2. Núcleo Punto Fijo-Universidad del Zulia. Pp.144-155.
- SANTROCK, John (2003). **Psicología del desarrollo en la infancia**. México, Editorial McGraw-Hill Interamericana. Séptima Edición, pp. 318-325.
- \_\_\_\_\_ (2004). **Introducción a la Psicología**. México, Editorial McGraw-Hill Interamericana. Segunda Edición, pp. 302-311.
- \_\_\_\_\_ (2006). **Psicología de la Educación**. México, Editorial McGraw-Hill Interamericana. Segunda edición, pp. 110-111.