



## Validación de los subtests de comprensión auditiva y denominación del test de Boston para el diagnóstico de la afasia

- ¶ Ana María Casas Monsegny, Fonoaudióloga, Docente de Práctica, Facultad de Salud, Programa de Fonoaudiología, Corporación Universitaria Iberoamericana, Práctica Privada Unidad Fonoaudiológica Colombiana.
- ¶ Clemencia Cuervo Echeverri, Terapeuta del Lenguaje, M.A., M.A. (Ed.), Profesora Asociada, Departamento de Terapias.
- ¶ Nancy Patricia Rojas Arias, Fonoaudióloga. Facultad de Medicina, Universidad Nacional de Colombia.

Se investigó la validez de los subtests de comprensión auditiva y denominación del test de Boston para el diagnóstico de la afasia de Goodglass y Kaplan (1-2). La determinación del grado de validez se hizo mediante la comparación estadística entre los resultados obtenidos con una muestra de 40 sujetos colombianos y los datos normativos originales del test, los cuales se utilizaron como medida criterio. Los criterios de selección de los sujetos -ausencia de neuropatología, edad, sexo y escolaridad- fueron los mismos utilizados para la validación española (2). Los resultados revelaron que los subtests ofrecen un buen grado de validez para su uso con poblaciones colombianas ya que no se encontraron diferencias estadísticamente significativas entre los resultados obtenidos en la aplicación de los subtests a los 40 sujetos y los datos normativos originales. Sin embargo, los hallazgos de la investigación señalan la necesidad de continuar con el proceso sistemático de validación de los subtests restantes para obtener información de la validez total del instrumento.

### INTRODUCCION

La evaluación fonoaudiológica del lenguaje y de los comportamientos de comunicación de las personas afásicas incluye la utilización de instrumentos estandarizados y no estandarizados, medidas funcionales de la comunicación y medidas especiales que tienen como propósito aportar criterios adicionales para orientar el diagnóstico diferencial. Ejemplos de estas medidas son las baterías del habla motora que se usan para evaluar la disartria o la apraxia del habla, los tests de comprensión auditiva y los tests de comprensión de lectura, entre otros (3).

Según Guzmán (4), los instrumentos estandarizados de mayor utilización en Colombia son: el *Boston Diagnosis Aphasia Examination - BDAE* de Goodglass y Kaplan (1986), el

*Western Aphasia Battery* de Kertesz (1982), el *Multilingual Aphasia Examination* de Spreen y Benton (1969), el *Minnesota Test for Differential Diagnosis of Aphasia - MTDDA* de Schuell (1965) y el *Porch Index of Communication Abilities - PICA* de Porch (1979).

Estos instrumentos fueron construidos y validados por fonoaudiólogos en otras latitudes con poblaciones diferentes a las colombianas. De éstos, el más utilizado en el país es la versión en castellano del BDAE realizada por García, Sánchez y Del Viso (2) en 1986 en la Universidad Complutense de Madrid, España (Test de Boston para el Diagnóstico de la Afasia). La preferencia en el uso de este test se explica no únicamente por el hecho de que existe una versión en castellano, sino debido a que se encuentra comercialmente disponible en Colombia. No obstante, ni éste último, ni los demás instrumentos mencionados han sido sometidos a procesos de validación rigurosos y sistemáticos con poblaciones colombianas. Únicamente se localizaron dos estudios de validación en la Universidad Nacional de Colombia: uno del *Test de Vocabulario complementario* del BDAE (5) y el otro de MTDDA (6). Sin embargo, ninguno de estos estudios condujo a la revisión de los instrumentos a pesar de que los dos detectaron problemas de validez.

También en el campo de la psicología colombiana se identifica el interés investigativo y clínico en la evaluación y el tratamiento de las alteraciones del lenguaje de las personas afásicas. Específicamente, en el subdominio de la neuropsicología se han registrado desarrollos particularmente valiosos durante la última década. Una de sus principales áreas de investigación ha sido la relacionada con estudios normativos en poblaciones

de personas afásicas colombianas (7). En la actualidad uno de los principales programas de investigación se orienta hacia la normalización de pruebas de diagnóstico neuropsicológico con poblaciones colombianas (8).

Internacionalmente, la fonoaudiología tiene una importante tradición de investigación psicométrica. Esto es explicable debido a que la evaluación del lenguaje humano se basa fundamentalmente en la observación de comportamientos externos para derivar inferencias respecto a atributos de naturaleza cognitiva y neurológica, no directamente observables. La investigación psicométrica en fonoaudiología también ha incluido la evaluación de instrumentos diseñados para medir dimensiones del habla (voz, articulación, fluidez, resonancia), aspectos psicoperceptuales de la audición y actitudes frente a un sinnúmero de eventos relacionados con la comunicación humana y sus desórdenes y la profesión de la fonoaudiología.

Por ejemplo, McCauley y Swisher (9) estimaron las características psicométricas de 30 tests estandarizados para la evaluación del habla y el lenguaje. Demorest y Walden (10) describieron principios de selección, interpretación y evaluación psicométrica de instrumentos fonoaudiológicos. Becerra y León (5) evaluaron las características psicométricas de un test de lenguaje infantil. Duncan y Perozzi (11) realizaron un estudio sobre el uso del lenguaje. Burns (12) realizó una investigación con relación a los aspectos específicos del lenguaje, más particularmente en el caso de la sintaxis. Finalmente, en cuanto a la evaluación psicométrica de instrumentos para evaluar el lenguaje adulto se encuentran investigaciones como la de Nicholas, McLennan y Brookshire (13), Nicholas y Brookshire (14), Lomas, Pickard, Bester, Elbard, Finlayson y Zoghaib (15), Biniek, Huber, Glindemann, Willmes y Klumm (16), García, Sánchez y Del Viso (2) y Cuervo (6), entre muchas otras.

La precisión de los instrumentos que se utilizan en el diagnóstico fonoaudiológico es vital, ya que su capacidad para medir de manera válida y confiable los atributos que pretende medir se relaciona directamente con la calidad de los servicios, la ética del ejercicio profesional y la posibilidad de documentar la eficacia de los tratamientos. El cumplimiento de estas condiciones es consistente con la filosofía y con las exigencias de la nueva ley de seguridad social en salud (17). Esta legislación habla específicamente de calidad, acreditación, auditoría, documentación de resultados, relaciones costo - eficiencia y del derecho de los usuarios a recibir información sobre la atención que se les suministra.

El propósito de esta investigación fue estimar la validez con referencia a criterio de los subtests de comprensión auditiva y

denominación de la versión en castellano del test de Boston para el diagnóstico de la afasia de Goodglass y Kaplan (2), con el fin de valorar la real utilidad del uso de este instrumento con personas colombianas. Se partió del siguiente razonamiento: si el instrumento falla significativamente en la evaluación de su validez, los resultados que se han obtenido con sujetos colombianos estarían siendo dudosos. Específicamente, se compararon dos conjuntos de datos: los datos normativos originales de cada uno de los subtests seleccionados y los datos obtenidos de la aplicación de estos mismos subtests a una muestra colombiana equivalente a la utilizada en el estudio original.

En su estudio psicométrico del test, Goodglass y Kaplan (1) determinaron que los dos subtests seleccionados para la presente investigación fueron los que mayor confiabilidad demostraron. Por esta razón, se escogieron para este estudio por considerarse los más potentes del test para evaluar el lenguaje adulto.

## MATERIALES Y METODOS

La muestra estuvo constituida por 40 adultos colombianos sin neuropatologías del habla y del lenguaje. Los criterios de selección fueron: la edad (17-72 años), el sexo (masculino-femenino), y la escolaridad (primaria, secundaria y superior) (tabla 1). También se determinó que la lengua materna de la muestra participante debía ser el castellano aprendido en Colombia. No se tuvo en cuenta como variable de selección el nivel socioeconómico, porque los sistemas taxonómicos para esta variable son ambiguos y difíciles de utilizar de manera consistente. Se asumió que esta variable podía identificarse con la variable escolaridad ya que esta última determina de una manera más fuerte la ejecución en el tipo de tareas de lenguaje que se utilizan en los tests (18, 19). La escogencia de los criterios se basó en los utilizados en el estudio de García, Sánchez y Del Viso (2) para la traducción al castellano del test de Boston para el diagnóstico de la afasia con 40 sujetos afásicos adultos hispanohablantes.

**Tabla 1.** Características de los sujetos de la muestra seleccionada (N=40).

Edad	< 25		26 - 40		> DE 40	
	Sexo		M	F	M	F
Primaria				5		2
Secundaria	2	6			2	1
Superior	1			5	4	1
						3

El test de Boston para el diagnóstico de la afasia es una prueba para la evaluación y el diagnóstico de la afasia, publicada en inglés por Goodglass y Kaplan en 1983, y traducido al

castellano por García, Sánchez y Del Viso en 1986. Este consta de 10 subtests, cada uno constituido por diferentes ítems que se complementan con 16 láminas para su ejecución. También incluye pruebas verbales complementarias, pruebas no verbales complementarias y el test de vocabulario de Boston. Se aplicaron los siguientes subtests para validarlos: el subtest de comprensión auditiva, conformado por cuatro categorías, cada una con diferente número de ítems. Estas son: discriminación de palabras (36 ítems), identificación de partes del cuerpo (20 ítems), órdenes (15 ítems) y material ideativo complejo (12 ítems) (1). El subtest de denominación consta de cuatro categorías, cada una con diferente número de ítems, estas son: respuesta de denominación (10 ítems), denominación por confrontación visual (32 ítems), nombrar animales y denominación de partes del cuerpo (10 ítems) (1).

Se aplicó una prueba piloto que permitió detectar una interferencia en la ejecución de los individuos relacionada con el uso de dos términos que no se utilizan en el castellano hablado en Colombia. Los dos términos se encontraron en el subtest de comprensión auditiva en las categorías de órdenes y material ideativo complejo. Estos términos fueron "mechero" y "cala" los cuales se cambiaron respectivamente por "encendedor" y "mojar". Esta modificación se introdujo con el fin de que los dos términos formaran parte del repertorio del castellano hablado en Colombia. A excepción de lo anterior, no se hizo ningún otro tipo de variación en cuanto a los ítems, las instrucciones de aplicación o el sistema de calificación de los subtests.

**Datos normativos originales.** Para determinar los puntajes que indican que no existe compromiso neurológico en un sujeto, Borod, Googlass y Kaplan especificaron en 1980 rangos para cada categoría de cada subtest por medio de la administración de la totalidad del test de Boston para el diagnóstico de la afasia (1). No se derivaron datos normativos para siete de las categorías del instrumento completo ya que se esperaba que no hubiera fracaso de los sujetos. Entre éstas se encontraba la categoría de denominación por confrontación visual, por lo tanto, se asumió el puntaje máximo de 96 como puntaje normativo, para poder así incluir todas las categorías contempladas en los dos subtests (Tabla 2).

**Procedimiento.** Para estimar la validez del BDAE se compararon sus datos normativos publicados en la versión traducida al castellano por García, Sánchez y Del Viso (2) y los datos obtenidos de la aplicación del instrumento en una muestra de 40 sujetos colombianos equivalente a la muestra utilizada por estos autores (2) para la traducción al castellano. Las investigadoras que aplicaron el instrumento fueron dos estudiantes de último año de fonoaudiología con una experiencia de 536 horas de práctica clínica con adultos y niños con

desórdenes de comunicación. Las dos investigadoras aplicaron una prueba piloto a una población de 20 sujetos. De esta manera, adquirieron experiencia en la consignación de los datos relacionados con edad, sexo y escolaridad. Así mismo, se entrenaron en la aplicación de los subtests de comprensión auditiva y denominación.

**Tabla 2. Datos normativos originales**

Subtest/Categoría	Rango/Calificación
<b>Comprensión auditiva</b>	
Discriminación de palabras	67-72
Identificación de partes del cuerpo	18-20
Órdenes	13-15
Material ideativo complejo	7-12
<b>Denominación</b>	
Respuesta de denominación	27-30
Denominación por confrontación visual	96
Denominación de partes del cuerpo	24-30
Nombrar animales	9-15

Para la ubicación y selección de los sujetos que participaron en la investigación, se recurrió a círculos familiares y sociales de las investigadoras. En todos los casos, se solicitó la participación voluntaria de las personas. Una vez seleccionados los sujetos, se aplicaron los dos subtests a los 40 individuos. El tiempo empleado en la entrevista y la aplicación de los subtests fue de aproximadamente 20 minutos por sujeto. Los lugares seleccionados para la aplicación del instrumento cumplieron con requisitos básicos de iluminación, control del ruido y mobiliario.

**Confiabilidad.** Se obtuvo una confiabilidad inter-examinador del 95% utilizando la fórmula de McReinolds y Kearns (20).

## RESULTADOS

El análisis cuantitativo de los resultados obtenidos se realizó a través de tres estrategias estadísticas: un diseño de cuadros de salida en el que se cruzaron los puntajes de las ocho categorías de los subtests de comprensión auditiva y denominación con las variables descriptivas escolaridad, edad y sexo con el propósito de determinar si los puntajes obtenidos se encontraban dentro de los rangos normativos de Goodglass y Kaplan (1). Se calcularon la media y la desviación estándar; la aplicación de estadísticas de prueba con el fin de establecer si se aceptaba o se rechazaba la hipótesis de igualdad, es decir, se observó si existían diferencias estadísticamente significativas dentro de cada grupo (escolaridad, edad y sexo), y si la ejecución de alguno de los grupos se encontraba en un nivel más bajo que la de otro grupo (21) y la aplicación de una prueba T para diferencia de medias para determinar la

significancia entre la diferencia de los resultados obtenidos por el grupo de 40 sujetos colombianos sin neuropatologías y los datos normativos de los subtests presentados por Borod, Goodglass y Kaplan en 1980, citados en el texto original de Goodglass y Kaplan (1). Los datos se procesaron con el paquete estadístico SPSS en un computador personal Compaq Prolinea 4/33s.

**Cuadros de salida.** El comportamiento normal en las tareas del lenguaje fue expresado en rangos en el trabajo original de validación del instrumento (1). En los cuadros de salida se observaron los puntajes máximos y los puntajes mínimos. Se calcularon la desviación estándar y la media de la ejecución de los sujetos en cada una de las ocho categorías que se incluyen en los dos subtests, y se observó si los puntajes mínimos del grupo colombiano se encontraban dentro de los rangos de normalidad.

En cuanto a la variable educativa se encontró que en el nivel primaria ( $N=11$ ), hubo tres categorías (identificación de partes del cuerpo, material ideativo complejo y denominación por confrontación visual), en las cuales los sujetos se ubicaron en un nivel más bajo del rango normal. Estos hallazgos demuestran que la variable escolaridad influye de manera importante en realización de este tipo de tareas de lenguaje (6,18,19). En la variable edad, en el grupo de menos de 25 años ( $N=14$ ), se observó el mismo comportamiento que con la variable escolaridad en la categorías de identificación de partes del cuerpo y material ideativo complejo. Los puntajes mínimos se ubicaron por debajo del rango normal. En el grupo de 26 a 40 años ( $N=14$ ), los puntajes obtenidos en la categoría de denominación por confrontación visual estuvieron por debajo del puntaje máximo. En lo relacionado con la variable sexo, el grupo femenino ( $N=27$ ) se comportó de manera semejante a las dos variables anteriores en las tres categorías mencionadas. En el grupo masculino ( $N=13$ ) se registraron puntajes que por debajo del puntaje máximo normativo en la categoría denominación por confrontación visual.

Los resultados obtenidos con las variables escolaridad, edad y sexo revelaron que éstas son importantes. La realización de las tareas por parte de los sujetos colombianos mostró ejecuciones por debajo de los datos normativos. Sin embargo, estas diferencias no alcanzaron niveles que afectaran negativamente la validez de los subtests debido a que el número de sujetos que produjeron los menores puntajes correspondió a un porcentaje del 2.5% en la categoría de identificación de partes del cuerpo, del 5.0% en la categoría de material ideativo complejo y del 10 % en la categoría de denominación por confrontación visual. El peso que representan estas proporciones de individuos no fue suficiente para hacer

que las diferencias fueran finalmente significativas (22). Se presentan las tablas 3 y 4 con el fin de ilustrar los resultados de los cruces de escolaridad con los subtests .

**Tabla 3.** Puntajes de las categorías en el subtest de comprensión auditiva por nivel educativo

Comprensión auditiva	Primaria	Secundaria	Superior	Total
<b>Discriminación de palabras</b>				
Número	11	15	14	40
Máximo	72.00	72.00	72.00	72.00
Mínimo	68.00	72.00	70.00	68.00
Media	71.00	72.00	71.86	71.68
Desviación	1.26	0.00	0.53	0.83
<b>Identificación de partes del cuerpo</b>				
Número	11	15	14	40
Máximo	20.00	20.00	20.00	20.00
Mínimo	16.50	19.00	19.00	16.50
Media	18.45	19.57	19.89	19.38
Desviación	1.06	0.37	0.29	0.85
<b>Ordenes</b>				
Número	11	15	14	40
Máximo	15.00	15.00	15.00	15.00
Mínimo	14.00	14.00	14.00	14.00
Media	14.91	14.93	14.93	14.93
Desviación	0.30	0.26	0.27	0.27
<b>Material ideativo complejo</b>				
Número	11	15	14	40
Máximo	11.00	12.00	12.00	12.00
Mínimo	5.00	8.00	9.00	5.00
Media	8.73	10.53	10.86	10.15
Desviación	1.68	0.991	1.23	1.55

**Estadísticas de prueba.** Se determinaron las hipótesis de igualdad ( $H_0$ ) o diferencia ( $H_1$ ) al máximo puntaje (21) con alfa igual a 0.01, 0.05 y 0.10 para cada una de las ocho categorías de los subtests de comprensión auditiva y denominación, respectivamente. Si las hipótesis arrojaban resultados que tendieran a cero se encontraban diferencias estadísticamente significativas. Por el contrario, si las hipótesis arrojaban resultados que tendieran a uno, no se encontraban diferencias estadísticamente significativas. Se observó, por una parte, el comportamiento de los resultados obtenidos totales en cada una de las hipótesis y, por otra parte, qué diferencias estadísticas existían entre grupos.

El comportamiento total en cada una de las categorías fue semejante, ya que para la mayoría de las hipótesis se aceptó

**Tabla 4.** Puntajes de las categorías en el subtest de denominación por nivel educativo.

Denominación	Primaria	Secundaria	Superior	Total
<b>Respuesta de denominación</b>				
Número	11	15	14	40
Máximo	30.00	30.00	30.00	30.00
Mínimo	28.00	30.00	27.00	27.00
Media	29.55	30.00	29.79	29.80
Desviación	0.82	0.00	0.80	0.65
<b>Denominación por confrontación visual</b>				
Número	11	15	14	40
Máximo	96.00	96.00	96.00	96.00
Mínimo	92.00	96.00	96.00	92.00
Media	95.00	96.00	96.00	95.73
Desviación	1.48	0.00	0.00	0.88
<b>Denominación de partes del cuerpo</b>				
Número	11	15	14	40
Máximo	30.00	30.00	30.00	30.00
Mínimo	24.00	27.00	30.00	24.00
Media	27.45	29.47	30.00	29.10
Desviación	1.75	0.92	0.001	1.48
<b>Nombrar animales</b>				
Número	11	15	14	40
Máximo	15.00	15.00	15.00	15.00
Mínimo	12.00	15.00	15.00	12.00
Media	14.73	15.00	15.00	14.93
Desviación	0.90	0.00	0.00	0.47

H1. En los grupos de escolaridad se observaron diferencias estadísticamente significativas con resultados que tendieron a cero, debido a que un grupo de sujetos que representaba el 10% de la muestra afectó el rendimiento de los resultados totales.

Esto quiere decir que estos resultados no afectan la validez de los subtests para medir procesos normales del lenguaje porque la proporción no es significativa para la muestra total (22). A diferencia de lo que ocurrió con los grupos de escolaridad, en los resultados de los grupos de edad y sexo no se encontraron resultados que tendieran a cero. En consecuencia, no hubo diferencias estadísticamente significativas que determinaran la significancia de las diferencias entre los resultados obtenidos con los 40 sujetos y los datos normativos originales reportados para los subtests.

**Prueba T para diferencia de medias.** En la aplicación de la prueba T para diferencia de medias entre los grupos establecidos para cada variable, se observó si el resultado de los cruces

entre los grupos tenía a cero o uno. En los cruces educación primaria versus educación secundaria, y educación primaria versus educación superior, en las categorías de identificación de partes del cuerpo y material ideativo complejo, el resultado de la aplicación de la prueba T tendió a cero para los dos casos. Esto significa que se dieron diferencias estadísticamente significativas entre la educación secundaria y la educación superior a diferencia de la educación primaria en estas dos categorías. Es decir, el grupo de educación primaria afectó los resultados totales de los cruces con los grupos escolaridad secundaria y escolaridad superior.

A pesar de que la categoría de denominación por confrontación visual presentó calificaciones por debajo del puntaje máximo, no se encontraron resultados que tendieran a cero con la aplicación de la prueba T. O sea, que no se registraron diferencias estadísticamente significativas que afectaran la validez de la categoría dentro del test.

En conclusión, el análisis de los resultados reveló que los subtests de comprensión auditiva y denominación ofrecen un buen grado de validez para su uso con la población colombiana. Esta afirmación se sustenta en el hecho de que no se encontraron diferencias significativas entre los resultados obtenidos en la aplicación a los 40 sujetos colombianos sin neuropatologías y los datos normativos originales del test. Sin embargo, es conveniente tener presente el hecho de que hubo sujetos normales que se comportaron por debajo de la norma. Estos comportamientos no afectaron estadísticamente la validez de los dos subtests debido a que el número de individuos que presentaron estos puntajes fueron muy pocos. En la categoría de identificación de partes del cuerpo la proporción fue de un 2.5%, en la categoría de material ideativo complejo la proporción fue de un 5.0% y en la categoría de denominación por confrontación visual la proporción fue de un 10%. Esta circunstancia se puede interpretar como una limitación metodológica relacionada con el tamaño de la muestra utilizada en este estudio para validar el instrumento. Sería lógico hipotetizar que si se hubiera trabajado con una muestra mayor, los resultados hubieran podido revelar diferencias estadísticamente significativas en contra de la validez del instrumento.

## DISCUSION

El propósito de este estudio fue estimar el grado de validez con referencia a criterio de dos subtests de un instrumento para evaluar el lenguaje adulto: el test de Boston para el diagnóstico de la afasia de Googlass y Kaplan (1) traducido al castellano por García Sánchez y Del Viso (2). Los resultados de esta investigación mostraron que esta prueba tiene capacidad para medir los procesos normales del lenguaje en una población

colombiana. Hubo una proporción pequeña de sujetos dentro del grupo de escolaridad primaria que estuvo por debajo del rango normativo reportado en el trabajo original de validación (1).

Además, estos resultados ofrecen confianza para la utilización del test completo con poblaciones colombianas. Sin embargo, es importante no perder de vista el aporte relativo de los instrumentos estandarizados en el proceso total de evaluación-diagnóstico y en la documentación de la eficacia de la intervención fonoaudiológica con poblaciones que presentan desórdenes del lenguaje de origen neurológico (3). En todas estas decisiones también interviene la información derivada de las evaluaciones cualitativas y de las medidas funcionales de comunicación. Por otro lado, no se puede olvidar que el juicio clínico del profesional juega un papel vital en todo el proceso de toma de decisiones (23).

Con relación a los antecedentes que motivaron la realización de esta investigación, en Colombia se han reportado dos estudios que estimaron la validez de dos instrumentos estandarizados utilizados por fonoaudiólogos para medir el lenguaje adulto y el lenguaje infantil: los de Cuervo (6) y Becerra y León (5).

El primer estudio estimó la validez del test de Minnesota para el diagnóstico diferencial de la afasia (19). Se trabajó bajo el supuesto de que las personas sin antecedentes neurogénicos debían producir un mínimo de errores en la ejecución de las tareas en comparación con los datos normativos (6). Este estudio no arrojó un buen grado de validez del instrumento debido a que una de las variables utilizadas, la escolaridad, contaminó significativamente los resultados. La muestra incluyó diez sujetos más de los que participaron en el presente estudio lo cual fue suficiente para que la variable escolaridad arrojara diferencias estadísticamente significativas.

En esta investigación, la variable escolaridad sugirió cierta influencia en la ejecución de los sujetos pero no alcanzó niveles de significancia. Esta discrepancia en los resultados de los dos estudios puede atribuirse al tamaño de las muestras utilizadas. En el primer estudio la muestra fue de 50 sujetos y en la presente investigación fue de 40. Es posible hipotetizar que si este estudio hubiera utilizado un número mayor de sujetos, el nivel de escolaridad habría podido vulnerar la validez del instrumento. La coincidencia radica en que la variable escolaridad se perfila como un factor de riesgo que puede afectar la ejecución de los sujetos en tareas de lenguaje, independientemente de la existencia de daño neurológico (18-19).

En lo referente a las características de la muestra, se debe tener

en cuenta que no únicamente la muestra utilizada fue pequeña sino que el procedimiento de selección no garantiza la generalización de los resultados. Brookshire (18) afirma que para que los resultados de las investigaciones se generalicen a una determinada población, la muestra de los sujetos que participe en ésta debe ser representativa de la población. Es decir, la muestra debe reflejar la población con respecto a las variables que ella misma afecte.

El segundo estudio (5) estimó la validez de test de vocabulario de Boston, prueba complementaria del test de boston para el diagnóstico de la afasia. Se encontraron diferencias estadísticamente significativas entre los datos normativos y los datos suministrados por la muestra colombiana lo cual afectó la validez del instrumento. Los resultados, por el hecho de ser una muestra grande, revelaron que el instrumento puede ser definitivamente vulnerable cuando se aplica a poblaciones colombianas. Existen diferencias con el presente estudio porque la población fue de niños, en oposición a una muestra de adultos, una condición que sin duda influyó en los resultados obtenidos.

El presente estudio vio la necesidad de hacer algunos cambios en el contenido de los ítems de algunas categorías. El estudio de Becerra y León (5) reveló que hubo algunas fallas en la validez de contenido que interfirieron con una ejecución eficiente en las tareas porque algunos sujetos no estaban familiarizados con estos términos. Estos resultados corroboran la necesidad de que para evitar al máximo las fallas de validez, los fonoaudiólogos tienen la responsabilidad de adelantar investigaciones sistemáticas que se dirijan a analizar los ítems de los tests a fin de validar su contenido.

Por otra parte, la validez de un instrumento estandarizado depende del rigor, tanto de su construcción como de los procesos de evaluación para obtener evidencia empírica sobre su validez y su confiabilidad. Esta es la única manera de asegurar un uso responsable y ético de este tipo de medidas (9). Los errores de construcción que se observaron en este estudio y en estudios anteriores ponen en evidencia las deficiencias de los instrumentos cuando estos aplican a poblaciones diferentes a aquella en la cual se estandarizó el instrumento. De tal forma, se hace necesario que los instrumentos sean evaluados y gradualmente adaptados por medio de estudios psicométricos empíricos que aseguren su validez cuando se aplican a poblaciones diferentes a la original.

Finalmente, los resultados de esta investigación señalan la necesidad de continuar con el proceso sistemático de validación de uno de los instrumentos más utilizados en Colombia para evaluar personas que presentan desórdenes neurogénicos del lenguaje adulto. Estos estudios deben ocuparse de todos los

aspectos psicométricos que señala la literatura (3,9,10,15,24).

La consolidación de toda esta información psicométrica contribuirá a mejorar la calidad en la prestación de servicios fonoaudiológicos ajustados a normas éticas.

En lo que compete a la versión en castellano del BDAE, futuras investigaciones deberán completar su validación para uso con poblaciones colombianas.

Por ahora, se puede utilizar con un buen grado de confianza, siempre y cuando el profesional valore adecuadamente las condiciones de cada persona, en particular su nivel educativo.

## SUMMARY

The purpose of this study was to estimate the degree of

psychometric validity of the Spanish version of the auditory comprehension and naming subtests of the Boston Diagnosis Aphasia Examination developed by Goodglass and Kaplan (1-2)). The degree of validity was obtained by statistically comparing two sets of data: The results obtained by 40 Colombian subjects and the original normative data used as criterion measure. The selection criteria - lack of neuropathologies, age, sex , and education - were the same used for the Spanish validation process (2).

Results revealed that the two subtests offer a reasonable degree of validity for its use with Colombian populations, since the observed differences between the two groups did not reach statistical significance. However, the evidence obtained suggests the need of continuing the validation process with the remaining subtests as to obtain psychometric information for the complete test.

## REFERENCIAS

1. Goodglass H, Kaplan E. La evaluación de la afasia y trastornos relacionados. 2a. Ed. Madrid: Panamericana. 1986: 7-190.
2. García J, Sánchez ML, Delviso S. Traducción al español del Test de Boston para el Diagnóstico de la Afasia. 2a. Ed. Madrid: Panamericana 1986: 7-190.
3. Simmons N. Beyond standardized measures: special tests, language in context, and discourse analysis in aphasia. SSL 1986; 7: 141-158.
4. Guzman E. Neuropsicología. 1a. Ed. Bogotá: Empresa Editorial Universidad Nacional de Colombia 1983: 7-351.
5. Becerra AI, León MC. Test de vocabulario de Boston: Estimación de validez. Tesis de grado sin publicar. Bogotá: Universidad Nacional de Colombia 1988.
6. Cuervo C. Una medida del lenguaje adulto: Estimación de validez. Trabajo de promoción docente sin publicar. Bogotá: Universidad Nacional de Colombia 1986.
7. Ardila R. Psicología en Colombia. Santafé de Bogotá: Tercer Mundo Editores 1993: 1-654.
8. Ardila A, Roselli M, Puente A. Neuropsychological assessment: A clinical guide. New York: Plenum. En prensa.
9. Mccauley RJ, Swisher L. Psychometric review of language and articulation test for preschool children. JSHD 1984; 49; 34-42.
10. Demorest ME, Walden BE. Psychometric principles in the selection, interpretation, and evaluation of communication self-assessment inventories. JSHD 1984: 49; 226-240.
11. Duncan JC, Perozzi JA. Concurrent validity of a pragmatic protocol. LSHSS 1987: 1; 80-85.
12. Burns E. Psychometric deficiencies of the short form version of West Syntax Screening Test. JSHD 1982; 4; 331-333.
13. Nicholas LE, McLennan DL, Brookshire RH. Validity of multiple-sentence reading comprehension tests for aphasic adults. JSHD 1986; 51; 82-87.
14. Nicholas LE, Brookshire RH. Error analysis and passage dependency of tests items from Standardized Test of Multiple Sentence Reading Comprehension for aphasic and non brain damaged adults. JSHD 1987; 52; 358-356.
15. Lomas J, Pickard L, Bester S, Elbard H, Finlayson A, Zoghaib C. The Communicative Effectiveness Index: Development and psychometric evaluation of a functional communication measure for adult aphasia. JSHD 1989; 54: 113-124.
16. Biniek R, Huber W, Glindemann R, Willmes K, Klumm H. The Aachen Aphasia Bedside Test . NK RWTH Aachen 1992.
17. Congreso de la República de Colombia. Sistema de Seguridad Social Integral. Ley 100 1993.
18. Brookshire RH. Subject description and generality of results in experiments with aphasic adults. JSHD 1983; 48: 342-346.
19. Schuell H. Booklet for the Minnesota Test for Differential Diagnosis of Aphasia. Minneapolis: University of Minnesota Press. 1965.
20. McReynolds LV, Kearns KP. Single-subject experimental design incommunicative disorders. 1983. Citados por Duncan y Perozzi concurrent validity of a pragmatic protocol. LSHSS 1987; 18: 80-85.
21. Kendal MG, Buckland WR. Diccionario de estadística. Madrid: Pirámide S.A. 1976.
22. Spiegel MR. Statistics. 1a. Ed. New York: Schaum Publishing Co. 1970: 1-359.
23. Records NL, Weiss AL. Clinical judgment: An overview. NSSLHAJ 1992; 20: 105-111.
24. Lyon JG. Standardized Test Batteries: Advances in aphasia testing. SSL 1986; 7: 159-180.