

**UNIVERSIDAD DE SEVILLA**  
**DEPARTAMENTO DE PSICOLOGIA EXPERIMENTAL**



**EVALUACION DEL DIF EN EL TEST DE RECUERDO VERBAL  
SELECTIVO ENTRE POBLACIÓN ESPAÑOLA, MEXICANA Y DE  
HABLA HISPANA EN E.U.A.**

Tesis presentada por:

***Fabiola Peña Cárdenas***

Para la obtención del Grado de  
Doctor en Aprendizaje y Cognición

Trabajo dirigido por:

***Dr. Manuel Morales Ortíz***

Profesor Titular de la Universidad de Sevilla

Sevilla, España., abril de 2015

## INDICE

1. Introducción .....	3
2. Estudio empírico .....	7
2.1. Objetivo general .....	8
2.2. Método.....	8
2.2.1. Participantes.....	8
2.2.1.1. Muestra de referencia española .....	8
2.2.1.2. Muestra focal mexicana.....	10
2.2.1.3. Muestra focal estadounidense .....	10
2.2.2. Materiales.....	12
2.2.2.1. Descripción de la prueba .....	12
2.2.3. Procedimiento .....	14
2.2.3.1. Análisis de datos.....	16
3. Resultados .....	16
4. Conclusión y discusión.....	24
Bibliografía .....	26

## 1. INTRODUCCION

La práctica neuropsicológica requiere conocimiento, flexibilidad, curiosidad y creatividad en el trabajo cotidiano pues, ésta es útil para identificar trastornos neurológicos en pacientes no psiquiátricos; documentar el estado cognoscitivo del paciente en las distintas condiciones neurológicas, y para identificar y localizar zonas de compromiso funcional, que en ocasiones escapan a los registros para clínicos convencionales. La evaluación neuropsicológica es útil también al estudiar enfermedades que no modifican en sus estadios iniciales la anatomía del cerebro, como las demencias o las encefalopatías tóxicas (Lezak, Howieson, Loring, Julia, & Fischer, 2004).

El estudio de la memoria ha sido de gran interés para los investigadores, en parte porque en muchos padecimientos, tanto neurológicos como psiquiátricos, uno de los primeros síntomas que se manifiestan en los pacientes son las alteraciones en esta función (Fuster, 2005). En un principio se avanzó poco en esta área, debido a que se estudiaba la memoria como una función única, pero ahora se sabe que esta habilidad está formada por diferentes subsistemas o procesos relacionados entre sí, que se organizan en distintas regiones del Sistema Nervioso.

Sin embargo, a pesar de la importancia que reviste, en algunos países se cuenta con pocas herramientas para realizar correctamente este trabajo. La

evaluación neuropsicológica de la memoria en los países de habla hispana es un claro ejemplo de ello (Llorente, 2008; Puente & Ardila, 2000). Es poca la investigación científica encaminada al desarrollo de instrumentos de evaluación confiables para esta población y la deficiencia es aún mayor en los países en vías de desarrollo como México (Ostrosky-Solis et al., 2007).

Por otra parte, el crecimiento del porcentaje de población de habla hispana en Estados Unidos de Norteamérica, ha incrementado la necesidad de contar con normas y tests apropiados para estudiar a ese grupo poblacional, el U.S. Census Bureau (1990; U. S. Census Bureau, 2005), ha estimado que para el 2020, habrá más de 52 millones de latinos en E.U.A. lo que se reflejará en un incremento notable en los servicios de salud requeridos por esta población (Llorente, 2008).

El presente documento, sin ser exhaustivo, pretende sumarse a los esfuerzos realizados en el campo de la evaluación neuropsicológica de la memoria verbal en población de habla hispana, mediante la adaptación del Test de Recuerdo Verbal Selectivo desarrollado por Buschke & Fuld, (1974), el cual ha sido estandarizado para su uso con población española por Campo & Morales, (2004). Se han estudiado sus propiedades psicométricas y obtenido datos normativos (Morales et al., 2010), pretendiendo a partir de esta investigación, ampliar su uso para población mexicana y de habla hispana en E.U.A.

La tesis ha sido dividida en seis apartados para facilitar su lectura, en el capítulo uno, se presenta la revisión bibliográfica en torno al tema de la evaluación neuropsicológica de la memoria, mostrando desde las principales clasificaciones de la memoria, los estratos neurológicos y fisiológicos implicados, pasando por las principales técnicas y pruebas desarrolladas para la evaluación neuropsicológica de la memoria verbal, aquellas que han sido desarrolladas y/o adaptadas en el idioma español, concluyendo con una revisión de la situación actual de la evaluación neuropsicológica en México y los latinos de E.U. A.

En el capítulo dos se presenta una revisión de la evaluación neuropsicológica transcultural, mostrando los principales errores en los cuales se puede caer al realizar esta labor (Ardila, 1995; Matsumoto & Van de Vijver, 2010), realizando una revisión desde la situación de los hispanos en E.U.A, una de las muestras focales del presente estudio (Ardila, 2013; Artiola I Fortuny, Heaton, & Hermosillo, 1998; Puente & Ardila, 2000). También se aborda el tema de la necesidad de adaptar pruebas para su uso en diferentes idiomas y culturas; las recomendaciones y directrices de la International Test Commission (ITC) en ese campo (Hambleton & Jong, 2003; Hambleton, Merenda, & Spielberger, 2006; Hambleton & Patsula, 1998; Hambleton & Zenisky, 2010; International Test Commission, 2010); posteriormente se realiza una revisión de las principales técnicas para comprobar la equivalencia de las mediciones para asegurar una evaluación justa a los clientes/pacientes de distinta

procedencia cultural (Arffman, 2013; Byrne & Campbell, 1999; Byrne et al., 2009; Milfont & Fischer, 2010; Penfield, 2010; Sireci, 2010).

El capítulo tres se centra en el tema del Funcionamiento Diferencial de Ítems (DIF), antecedentes históricos, así como las principales ventajas y desventajas de las diferentes técnicas estadísticas para calcularlo (Andrich & Hagquist, 2012; Bandeira Andriola, 2003; Hambleton & Patsula, 1999; Zenisky, Hambleton, & Robin, 2003), centrándose principalmente en los procedimientos utilizados en nuestro estudio: las técnicas de Mantel-Haenszel (Camilli & Penfield, 1997; Mantel, 1963) y de Regresión Logística (Ferrerres Traver, González Romá, & Gómez Benito, 2000; French & Miller, 1996; Hidalgo Montesinos, Gómez Benito, & Padilla García, 2005; Swaminathan & Rogers, 1990).

En el capítulo cuatro, se presenta la descripción del estudio empírico llevado a cabo para la realización de nuestra investigación: objetivos, método, participantes, instrumento utilizado y procedimiento.

En el capítulo cinco se presentan los resultados obtenidos en el estudio del DIF en el vSRT, se detallan las diferentes etapas seguidas en el análisis de resultados, mediante los procedimientos elegidos para tal efecto: Regresión Logística y Mantel-Haenszel.

Por último; en el capítulo seis se exponen las conclusiones generales, discusión, limitantes del estudio y recomendaciones para investigaciones futuras.

## **2. ESTUDIO EMPÍRICO**

En un estudio exploratorio previo, realizado como parte de los trabajos para la obtención del DEA de la autora, se evaluó el funcionamiento diferencial del ítem entre población española y mexicana, utilizando para tal efecto en la muestra focal (mexicana) una muestra de 90 participantes. En aquel estudio se encontró un Funcionamiento Diferencial del Ítem “bajo” entre estas dos poblaciones. En el presente estudio se espera corroborar estos resultados contando con unas muestras más grandes; y además ampliarlos a la población de habla hispana en E.U.A. La pregunta de investigación a la que se tratará de dar respuesta es: ¿Existe Funcionamiento Diferencial del Ítem en el TRVS entre poblaciones española, mexicana y de habla hispana en E.U.A.?

Para tratar de responder la pregunta de investigación, se establecieron los objetivos general y específicos, así como la metodología del estudio que a continuación se detallan.

## **2.1. Objetivo general**

El objetivo general del presente estudio es evaluar el Funcionamiento Diferencial del Test de Recuerdo Verbal Selectivo entre población española, mexicana y de habla hispana en E.U.A.

## **2.2. Método**

Este estudio sigue un diseño no experimental, de tipo transversal y comparativo (intercultural) con un enfoque cuantitativo.

### **2.2.1. Participantes**

#### **2.2.1.1. Muestra de referencia española**

Los datos de los participantes de la muestra de referencia fueron 211 españoles seleccionados de entre los participantes del estudio para obtener datos normativos para dos versiones en español de la prueba vSRT con 6 ensayos. (Morales et al., 2010). El primer requisito fue ser de nacionalidad española. Los criterios de inclusión en el proceso de estandarización fueron:

- a) No tener antecedentes de enfermedad neurológica.
- b) No presentar antecedentes de hospitalización por cuestiones psicopatológicas (esquizofrenia, trastorno depresivo mayor).
- c) No antecedentes de desarrollo psicomotriz anormal.
- d) No tener uso de psicofármacos que pudieran interferir con cuestiones como concentración, atención o producir somnolencia.
- e) No tener historial de consumo de drogas o alcoholismo.
- f) Tener el idioma español como primera lengua.

Los individuos con condiciones médicas crónicas como hipertensión, diabetes, pérdida auditiva leve, no fueron excluidos del estudio. Todos los participantes fueron considerados como capacitados cognitivamente para el funcionamiento independiente. La mayoría de los participantes fueron reclutados en el sur y sureste de España, intentando obtener datos tanto de áreas rurales como urbanas. De los 804 sujetos evaluados en aquel país se procedió a seleccionar mediante un muestreo no probabilístico por conveniencia a los 211 sujetos, que fueron equiparados con las muestras de las poblaciones focales: mexicanas y estadounidenses en cuanto a las variables, edad, escolaridad y sexo.

### **2.2.1.2. Muestra focal mexicana**

La muestra del primer grupo focal (mexicano) está conformada por un total de 201 sujetos adultos mexicanos voluntarios, seleccionados en la ciudad de H. Matamoros, Tamaulipas. El tipo de muestreo utilizado fue intencional, seleccionando a aquellos sujetos que aceptaron participar en el estudio y que cumplieron con los mismos criterios de inclusión que la muestra española pero con nacionalidad mexicana.

### **2.2.1.3. Muestra focal estadounidense**

La muestra del segundo grupo focal: población de habla hispana en E.U.A., estuvo conformada por un total de 205 sujetos adultos voluntarios, seleccionados en la ciudad de Brownsville, Texas. El tipo de muestreo fue intencional, seleccionando a aquellos sujetos que aceptaron participar en el estudio y que cumplieron con los mismos *criterios de inclusión* que las muestras anteriores más los siguientes:

- a) Ser de origen hispano.
- b) Estar viviendo de manera permanente en E.U.A.
- c) Tener el idioma español como lengua.

En las tablas 1 y 2 se presentan los datos demográficos de las muestras (edad, escolarización y sexo).

Tabla 1

*Distribución de las muestras Edad y Escolarización*

Variable	N	Media	DE	Error típico	Intervalo de confianza para la media al 95%		Mínimo	Máximo
					Límite inferior	Límite superior		
<b>Edad</b>								
España	211	42.09	14.411	.992	40.13	44.04	15	77
México	201	43.77	18.327	1.293	41.22	46.32	15	79
EUA	205	41.96	14.337	1.001	39.99	43.94	16	77
Total	617	42.59	15.767	.635	41.35	43.84	15	79
<b>Escolarización</b>								
España	211	11.25	3.635	.250	10.75	11.74	2	21
México	201	11.46	5.602	.395	10.68	12.24	0	23
EUA	202	11.00	4.095	.288	10.43	11.57	0	21
Total	614	11.24	4.505	.182	10.88	11.59	0	23

Tabla 2

*Distribución de muestras x Sexo*

Población	Sexo				Total	
	Masculino		Femenino		Frecuencia	Porcentaje
	Frecuencia	Porcentaje	Frecuencia	Porcentaje		
España	62	27%	149	71%	211	100%
México	54	27%	147	73%	201	100%
EUA	57	28%	148	72%	205	100%
Total	173	27%	444	72%	617	100%

### **2.2.2. Materiales**

Para la realización de la presente investigación se utilizó el Test de Recuerdo Verbal Selectivo (TRVS) en su versión en español, desarrollado por Campo, Morales, & Juan-Malpartida, (2000). Existen dos formas del test, pero para efectos de la presente investigación sólo se utilizó la forma uno, también se aplicó un cuestionario de datos personales para la identificación de los participantes en cuanto a las variables sociodemográficas.

#### **2.2.2.1. Descripción de la prueba**

La prueba consta de dos fases: la fase del *recuerdo inmediato* y la del *recuerdo demorado*, en la primera de ellas, se realizan a su vez dos evaluaciones, la primera evaluación que es la de *aprendizaje*, consiste en que el evaluador lee al sujeto la lista de palabras completa (*dado, cinta, norte, jarro, pollo, frente, llave, cruz, fuego, pena, modelo, oído*); se realiza a una velocidad de una palabra cada 2 segundos; para que posteriormente intente recordar la mayor cantidad posible de ellas, teniendo para tal efecto seis ensayos como máximo y siendo recordado sistemáticamente con aquellas palabras que no ha dicho en el ensayo precedente.

Posteriormente se realiza la evaluación de *reconocimiento en elección múltiple*, en la cual se le presentan al sujeto doce tarjetas blancas separadas

con cuatro palabras escritas en color negro, una en cada esquina. Cada tarjeta contiene una palabra de la lista y tres distractores: un distractor fonético, otro semántico y uno no relacionado, de los cuales el sujeto tiene que seleccionar la palabra que considera pertenece a la lista.

En la prueba de aprendizaje inicial se evalúan seis medidas distintas, Recuerdo Total (Recall), Recuerdo a corto plazo (STR), Almacenamiento a largo plazo (LTS), Recuerdo a largo Plazo (LTR), Recuerdo consistente a largo plazo (CLTR), Recuerdo aleatorio a largo plazo (RLTR). Además de reconocimiento con opción múltiple tanto inmediato como demorado y adiciones o intrusiones.

La segunda fase que es *recuerdo demorado* se evalúa 30 minutos después de haber concluido las dos primeras actividades del recuerdo inmediato. Consiste en solicitarle al sujeto que mencione la mayor cantidad de palabras que recuerde de la lista, pero teniendo en esta ocasión un solo ensayo, posteriormente se pasa al reconocimiento de elección múltiple siguiendo el mismo proceso que en la fase de recuerdo inmediato.

### 2.2.3. Procedimiento

Con aquellos sujetos que aceptaron participar y que contaban con los criterios de inclusión, se concertó la cita para proceder a la aplicación del instrumento.

La aplicación del instrumento se llevó a cabo en las áreas que fueron facilitadas por las instituciones. En España: despachos del Departamento de Psicología Experimental de la Facultad de Psicología y el Aula de la Experiencia de la Universidad de Sevilla, así como en diversos centros de mayores. En México, en el Centro de Atención Psicológica y la Cámara de Gesell de la Unidad Académica Multidisciplinaria Matamoros-UAT, en el departamento de psicología del Centro para la Juventud y la Familia y en el consultorio de enfermería del Gimnasio Multidisciplinario de la UAT. En el caso de la muestra Estadounidense, la mayoría de las evaluaciones se realizaron en la *Cameron Park Clinic of the Brownsville Community Health Center* (Centro Comunitario de Salud de Brownsville, Clínica Cameron Park), se utilizó el consultorio destinado a *Social Work* (Trabajo social). Para el caso de los estudiantes universitarios de la *University of Texas at Brownsville* (Universidad de Texas en Brownsville), se utilizaron espacios de la biblioteca y aulas de la institución. En todos los casos se buscó que la aplicación se llevara a cabo en cubículos privados que contaran con sillas y escritorio, que hubiera iluminación y ventilación adecuada, con la menor cantidad posible de ruido proveniente del exterior, que pudiera interferir con la atención de los sujetos.

La aplicación se realizó de manera individual con cada sujeto, con una duración promedio de 50 a 60 minutos; la ubicación del evaluador y del sujeto fue sentados cara a cara, uno a cada lado del escritorio. La primera actividad dentro del proceso de evaluación consistió en obtener el consentimiento informado por parte del sujeto, para ello, se elaboró un formato por escrito en el cual se describió a grandes rasgos la investigación, los objetivos del estudio, los procedimientos, implicaciones para los participantes, etc. Se le entregó a cada sujeto para que lo leyera detenidamente y una vez que concluyó, se procedió a aclarar cualquier duda que le pudiera haber surgido durante la lectura.

Una vez leído y firmado el consentimiento informado por parte del evaluador y el evaluado, se procedió a la evaluación.

Para efectos de la investigación, en el intervalo de 30 min. entre la fase de recuerdo inmediato y demorado, se aplicó una batería de pruebas que miden diversos constructos, con el fin de que el sujeto pasara los 30 minutos ocupado en otras actividades no relacionadas con la lista de palabras. Cabe hacer mención que para interferir lo menos posible con los resultados del TRVS, ninguna de las pruebas utilizadas midió memoria semántica verbal.

### 2.2.3.1. Análisis de datos

Los análisis estadísticos generales se realizaron con SPSS versión 20. Para el análisis del DIF mediante Mantel-Haenszel se utilizó el paquete estadístico DIFAS (Randall D. Penfield, 2005), mientras que el análisis del DIF mediante Regresión Logística también se realizó con SPSS versión 20.

## 3. RESULTADOS

Debido a la extensión de los análisis realizados para evaluación del DIF, los resultados se han dividido en seis etapas para su mejor comprensión. En la **primera** de ellas se presentan las comparaciones mediante ANOVA univariante realizado a las puntuaciones globales de la prueba (Ver Tabla 3).

Tabla 3

*Análisis de la varianza puntuaciones globales del vSRT.*

Variables	ANOVA		Estadístico de Levene		Brown-Forsythe	
	Valor de F	p	Valor de F	p	Valor de F*	p
RECALL	-	-	7.717	<b>.000*</b>	23.783	<b>.000*</b>
LTR	-	-	5.894	<b>.003*</b>	24.260	<b>.000*</b>
STR	-	-	123.198	<b>.000*</b>	58.934	<b>.000*</b>
LTS	64.576	<b>.000*</b>	1.602	.202	-	-
CLTR	13.079	<b>.000*</b>	.620	.538	-	-

RLTR	-	-	3.070	<b>.047*</b>	15.498	<b>.000*</b>
DR	63.477	<b>.000*</b>	2.798	.062	-	-
INT	-	-	23.047	<b>.000*</b>	25.413	<b>.000*</b>
MCRT	1.893	.152	1.008	.366	-	-
DMCRT	-	-	3.771	<b>.024*</b>	1.785	<b>.169*</b>

\* p &lt; 0.05

En todos los casos en que se encontraron diferencias significativas en las puntuaciones medias, éstas fueron a favor de la población española, con excepción de la variable STR, en donde la mayor puntuación fue para los estadounidenses, seguidos por los mexicanos; sin embargo, estas puntuaciones son menos deseables o puntuaciones que nos indican una mayor memoria a corto plazo en lugar de a largo plazo. En la variable adiciones (INT) los mexicanos tuvieron la puntuación más alta seguidos por los estadounidenses; esta variable también es negativa, ya que las adiciones, son indicador de confusión con otras palabras que no están en la lista, por lo que tampoco es deseable tener puntuaciones altas (Ver Tabla 4). Por lo anterior, se puede afirmar, que la muestra española tuvo un funcionamiento más alto en todas las sub-escalas del vSRT seguido por la mexicana y estadounidense.

Tabla 4

*Puntuaciones globales del vSRT por muestras.*

Variables	Media	DE	Error típico	Intervalo de confianza para la media al 95%		Mínimo	Máximo
				Límite inferior	Límite superior		
<b>RECALL</b>							
España	47.55	8.752	.602	46.37	48.74	23	69
México	44.47	9.810	.692	43.11	45.84	20	66
EUA	40.82	11.153	.779	39.29	42.36	18	71
Total	44.31	10.306	.415	43.50	45.13	18	71
<b>LTR</b>							
ESPAÑA	36.35	13.052	.899	34.58	38.12	2	69
MÉXICO	32.81	14.309	1.009	30.82	34.80	0	66
E.U.A.	26.48	16.281	1.137	24.24	28.72	0	66
Total	31.92	15.138	.609	30.72	33.12	0	69
<b>STR</b>							
España	11.30	5.659	.390	10.53	12.07	0	27
México	11.74	5.910	.417	10.91	12.56	0	28
EUA	19.94	13.447	.939	18.09	21.79	0	65
Total	14.31	9.898	.398	13.53	15.10	0	65
<b>LTS</b>							
España	39.28	13.112	.903	37.50	41.06	3	69
México	35.56	14.537	1.025	33.54	37.58	0	67
EUA	23.74	15.808	1.104	21.56	25.92	0	68
Total	32.91	15.942	.642	31.65	34.17	0	69
<b>CLTR</b>							
España	25.81	14.808	1.019	23.80	27.82	0	69
México	26.20	14.056	.991	24.24	28.15	0	64
EUA	19.46	15.979	1.116	17.26	21.66	0	64
Total	23.83	15.263	.614	22.62	25.03	0	69
<b>RLTR</b>							
España	10.53	7.147	.492	9.56	11.50	0	42
México	8.46	6.042	.426	7.62	9.30	0	36
EUA	7.05	5.994	.419	6.23	7.88	0	31
Total	8.70	6.575	.265	8.18	9.22	0	42
<b>DR</b>							
España	9.47	2.684	.185	9.10	9.83	0	12
México	7.53	2.563	.181	7.18	7.89	0	12
EUA	6.74	2.369	.165	6.42	7.07	0	12
Total	7.93	2.789	.112	7.71	8.15	0	12
<b>INT</b>							
España	.95	1.241	.085	.78	1.12	0	5
México	2.20	2.055	.145	1.92	2.49	0	10
EUA	1.47	1.947	.136	1.20	1.74	0	10

Total	1.53	1.848	.074	1.39	1.68	0	10
MCRT							
España	12.35	6.910	.476	11.41	13.29	8	12
México	11.61	1.095	.077	11.45	11.76	4	12
EUA	11.77	.563	.039	11.69	11.84	9	12
Total	11.91	4.108	.165	11.59	12.24	4	12
DMCRT							
España	11.75	.767	.053	11.64	11.85	6	12
México	11.60	.981	.069	11.46	11.73	4	12
EUA	11.63	.822	.057	11.52	11.74	6	12
Total	11.66	.861	.035	11.59	11.73	4	12

$N=617$ ,  $n$  España=211,  $n$  México=201,  $n$  E.U.A.=205

En la **segunda etapa** se realiza análisis del DIF para ítems dicotómicos a cada uno de los ítems del vSRT en los distintos ensayos, se presentan resultados del análisis Mantel-Haenszel, utilizando el programa estadístico DIFAS, para cada una de las comparaciones dos a dos de las muestras (focal y referencia).

Para analizar los datos mediante esta técnica se construyó una base de datos en la que se registró si cada sujeto recordó o no cada ítem en cada ensayo. Por tanto, se analizaron 72 variables (12 ítems por seis ensayos). Los grupos fueron equiparados en el nivel de habilidad, utilizando el puntaje total de recuerdo (RECALL) de la prueba como indicador de habilidad ( $\theta$ ). Las comparaciones en este caso siempre se hacen sólo entre dos grupos (G) a la vez (uno de referencia y otro focal). Se corrieron los análisis mediante el comando de *modelos no paramétricos en ítems dicotómicos (DIF analysis: Nonparametric tests for dichotomous ítems)*.

Para determinar el grado de DIF se utilizaron los siguientes estadísticos: *Chi cuadrado de Mantel-Haenszel (MH CHI)*, *Mantel-Haenszel Common log-odds ratio (MH LOR)*, *Logaritmo de la razón de ventajas estandarizado del estadístico Mantel-Haenszel (LOR Z)*, *Chi cuadrado de Breslow-day (BD)*, *Regla de decisión combinada (CDR)*.

Siguiendo la Regla de decisión combinada (CDR); en la comparación España-México se detectaron 11 ítems “flag” (sesgados), en la comparación España-EUA se detectaron 12; mientras que en la comparación México-EUA, solamente 4.

En una **tercera etapa** se presentan resultados de la aplicación de la técnica de Regresión Logística para ítems dicotómicos a cada uno de los ítems en los 6 ensayos del vSRT, también para cada una de las comparaciones dos a dos entre las muestras del estudio.

Para este análisis se realizaron tres regresiones logísticas para cada uno de los 12 ítems en los 6 ensayos, utilizando el procedimiento de comparación de modelos anidados descrito por (Hidalgo Montesinos et al., 2005). Se probaron tres modelos distintos en cada uno de ellos, utilizando el acierto al ítem o no acierto al ítem como variable dependiente (puntuando 0 en no aciertos y 1 el acierto). En el primer modelo se incluye la variable RECALL como variable independiente y considerado como nivel de habilidad ( $\theta$ ); en el

segundo se agrega otra variable independiente que es la población o grupo (focal o referencia) a la que se le asigna valor de 0 al focal y 1 al de referencia. En el tercer modelo se agrega la variable interacción entre la variable RECALL y el grupo. Una vez concluidas las tres regresiones, se comparan los aportes de cada variable a la razón de verosimilitud en cada modelo.

En la comparación España-E.U.A. un total de 19 ítems fueron detectados por la Regresión Logística con posible DIF. Al aplicar la técnica de RL en los ítems del vSRT entre población española y mexicana se detectaron con posible DIF un total de 23 ítems. En la comparación entre México (población de referencia) y EUA (población focal), mediante este procedimiento se identificaron un total de 6 ítems con posible DIF.

En la **cuarta etapa** se realizaron las purificaciones bietápicas para cada una de las comparaciones y se volvieron a realizar los análisis DIF, mediante Mantel-Haenszel. Se encuentran resultados muy similares entre los procedimientos de purificación tanto con resultados de MH y RL, sin embargo al encontrar principalmente DIF uniforme, el procedimiento de MH parece ser más adecuado. En la comparación España- México fueron detectados 12 y 13 ítems “flag”, utilizando resultados de MH y RL respectivamente. En la comparación España-EUA fueron detectados 12 ítems con ambos procedimientos; mientras que al comparar México-EUA, solamente 4 ítems con ambos procedimientos.

Debido a la mayor similitud en los análisis previos encontrados entre las muestras focales (mexicana y estadounidense); en la **quinta etapa** se decide fusionarlas y realizar análisis DIF en comparación con la muestra española, se utiliza nuevamente procedimiento Mantel-Haenszel. Se encontró un total de 17 ítems marcados “flag” (CDR).

Una ventaja agregada del paquete estadístico DIFAS, es que además del análisis del funcionamiento diferencial de cada uno de los ítems, permite evaluar el Funcionamiento Diferencial del Test, que es un indicador global del nivel de DIF, presente en una prueba.

En la **sexta** y última etapa se evalúa el Funcionamiento Diferencial del Test (DTF; *Differential Test Functioning*) mediante DIFAS para las diferentes comparaciones.

La varianza del efecto del DIF a través del conjunto de ítems de un test o escala ha sido propuesta como un medida de DTF para conjuntos de ítems dicotómicos (Randall D. Penfield, 2005). La varianza del efecto del DIF entre ítems dicotómicos se simboliza como  $\tau^2$  (tau-squared) (Longford, Holland, & Thayer, 1993), o la versión ligera  $\tau_{\omega}^2$  (weighted tau-squared), propuesta por Camilli & Penfield (1997). Ambos estadísticos son calculados en el paquete estadístico DIFAS y su interpretación puede servir como indicador para

determinar el funcionamiento diferencial del test en su conjunto (Rao & Sinharay, 2007).

Para efectos del presente estudio y considerando las características de la prueba se utilizó el procedimiento de análisis del Funcionamiento Diferencial del Test (DTF) no paramétrico para ítems dicotómicos, en cada par de comparaciones. Para la interpretación de los estadísticos se ha seguido la interpretación del estimador  $\tau^2$ (tau-squared), siguiendo el procedimiento descrito por Brown (2010).

En el análisis del DTF entre las muestras de España y México; podemos observar que este presenta un nivel medio  $\tau^2 = 0.10 > 0.07$ , al realizar la purificación el valor de DTF baja ligeramente, pero se mantiene en un nivel medio  $\tau^2 = 0.093 > 0.07$ .

En cuanto al primer análisis del DIF mediante MH en la comparación población española y estadounidense el Funcionamiento Diferencial del Test fue medio  $\tau^2 = 0.11 > 0.07$ . Posterior al proceso de purificación el DTF en esta comparación aumentó ligeramente; no obstante, se mantiene en un nivel medio  $\tau^2 = 0.116 > 0.07$ .

Al analizar el Funcionamiento Diferencial del Test en su conjunto en la comparación realizada entre la muestra mexicana y estadounidense, se observa que este es bajo o inexistente  $\tau^2 = 0.25 < 0.07$ . Al realizar el proceso de purificación el DTF baja ligeramente, pero se sigue manteniendo en nivel bajo o inexistente  $\tau^2 = 0.022 < 0.07$ .

#### **4. CONCLUSIÓN Y DISCUSIÓN**

En México, como en la mayoría de los países en vías de desarrollo, el uso de pruebas psicométricas desarrolladas para otros países y/o culturas, es una práctica común, sin embargo como se pudo observar en capítulos anteriores, esto conlleva una disminución considerable e incluso la anulación de su confiabilidad y validez. El campo de la evaluación neuropsicológica de la memoria no es la excepción en nuestro país; motivo por el cual, con la presente investigación se pretende hacer extensivo el uso del Test de Recuerdo Verbal Selectivo en su versión en español (Campo et al., 2000), al evaluar el Funcionamiento Diferencial del Test de Recuerdo Verbal Selectivo entre población española (población de referencia), mexicana e hispana en Estados Unidos de Norteamérica (poblaciones focales). Esto debido a que la prueba cuenta con datos normativos (Campo & Morales, 2004; Morales et al., 2010), de confiabilidad (Campo et al., 2000) y validez (Campo, Morales, & Martínez-Castillo, 2003) para población española, y de no existir DIF o DFT, la prueba podría ser utilizada en nuestras poblaciones focales con los mismos criterios de

confiabilidad, validez y normatividad que los obtenidos en la población de referencia.

Con base en el análisis estadístico realizado, podemos determinar que en su conjunto, el tRVS presentó funcionamiento diferencial mayor al comparar las muestras españolas contra las mexicanas y estadounidenses.

En cambio, al realizar las comparaciones de DIF entre las muestras mexicana (referencia) y estadounidense (focal), los ítems y el grado de DIF fue significativamente menor al presentado en las comparaciones contra la muestra española. A pesar de que se detectaron DIF moderados y no homogéneos en algunos de los ítems en los diversos ensayos, estos no llegaron a ser favorecedores para ninguna de las dos poblaciones estudiadas.

Además, al evaluar el Funcionamiento Diferencial del Test en su conjunto entre estas dos poblaciones, el grado de DTF, fue bajo o nulo, incluso sin necesidad de realizar purificación en la puntuación global. Esto podría ser indicio de una mayor similitud en aspectos socio-culturales que pudieran influir en el desempeño en el vSRT y en otras pruebas neuropsicológicas, entre estas dos poblaciones, que con respecto a la población española.

Por lo anterior, podemos concluir que **no** se recomienda utilizar esta prueba en población mexicana e hispana en EUA, con los datos normativos, confiabilidad y validez establecidos para población española; sino que es necesario realizar un nuevo procedimiento de estandarización para estas poblaciones.

Sin embargo a la luz de estos resultados, podríamos considerar la posibilidad de realizar un proceso de estandarización con los resultados de las muestras focales fusionadas; es decir, las muestras mexicanas y estadounidenses, ya que como se observó en los análisis previos, estos grupos parecen tener un funcionamiento muy similar en el tRVS; por lo que la línea de investigación futura y más próxima sería ésta.

## BIBLIOGRAFIA

- Andrich, D., & Hagquist, C. (2012). Real and Artificial Differential Item Functioning. *Journal of Educational and Behavioral Statistics*, 37(3), 387–416. <http://doi.org/10.3102/1076998611411913>
- Ardila, A. (1995). Directions of research in cross-cultural neuropsychology. *Journal of Clinical and Experimental Neuropsychology*, 17(1), 143–150. <http://doi.org/10.1080/13803399508406589>

- Ardila, A. (2013). The impact of culture in neuropsychological test performance. In B. P. Uzzell, M. Ponton, & A. Ardila, *International Handbook of Cross-Cultural Neuropsychology*. Psychology Press.
- Arffman, I. (2013). Problems and Issues in Translating International Educational Achievement Tests. *Educational Measurement: Issues and Practice*, 32(2), 2–14. <http://doi.org/10.1111/emip.12007>
- Artiola I Fortuny, L., Heaton, R. K., & Hermsillo, D. (1998). Neuropsychological comparisons of Spanish-speaking participants from the U.S.-Mexico border region versus Spain. *Journal of the International Neuropsychological Society*, 4(4), 363–379.
- Bandeira Andriola, W. (2003). Descripción de los principales métodos para detectar el funcionamiento diferencial del ítem (DIF) en el área de la evaluación educativa. *Bordón: Revista de Orientación Pedagógica*, 55(2), 177–189.
- Brown, A. (2010). Introducing concepts of measurement invariance. Investigating Differential Item Functioning (DIF) using various approaches (Mantel-Haenszel and Confirmatory Factor Analysis (CFA) with covariates). Retrieved from [http://www.psychometrics.cam.ac.uk/uploads/documents/ESRC\\_RDI\\_Se p\\_10/summer-school-2-day4.pdf](http://www.psychometrics.cam.ac.uk/uploads/documents/ESRC_RDI_Se p_10/summer-school-2-day4.pdf)
- Buschke, H., & Fuld, P. A. (1974). Evaluating storage, retention, and retrieval in disordered memory and learning. *Neurology*, 24(11), 1019–1019. <http://doi.org/10.1212/WNL.24.11.1019>
- Byrne, B. M., & Campbell, T. L. (1999). Cross-Cultural Comparisons and the Presumption of Equivalent Measurement and Theoretical Structure A

- Look Beneath the Surface. *Journal of Cross-Cultural Psychology*, 30(5), 555–574. <http://doi.org/10.1177/0022022199030005001>
- Byrne, B. M., Oakland, T., Leong, F. T. L., van de Vijver, F. J. R., Hambleton, R. K., Cheung, F. M., & Bartram, D. (2009). A critical analysis of cross-cultural research and testing practices: Implications for improved education and training in psychology. *Training and Education in Professional Psychology*, 3(2), 94–105. <http://doi.org/10.1037/a0014516>
- Camilli, G., & Penfield, D. A. (1997). Variance Estimation for Differential Test Functioning Based on Mantel-Haenszel Statistics. *Journal of Educational Measurement*, 34(2), 123–139.
- Campo, P., & Morales, M. (2004). Normative data and reliability for a Spanish version of the verbal Selective Reminding Test. *Archives of Clinical Neuropsychology*, 19(3), 421–435. [http://doi.org/10.1016/S0887-6177\(03\)00075-1](http://doi.org/10.1016/S0887-6177(03)00075-1)
- Campo, P., Morales, M., & Juan-Malpartida, M. (2000). Development of two Spanish versions of the verbal selective reminding test. *Journal of Clinical and Experimental Neuropsychology*, 22(2), 279–285.
- Campo, P., Morales, M., & Martínez-Castillo, E. (2003). Discrimination of Normal from Demented Elderly on a Spanish Version of the Verbal Selective Reminding Test. *Journal of Clinical and Experimental Neuropsychology*, 25(7), 991–999. <http://doi.org/10.1076/jcen.25.7.991.16492>
- Ferreres Traver, D., González Romá, V., & Gómez Benito, J. (2000). Comparación del estadístico Mantel-Haenszel y la regresión logística en el funcionamiento diferencial de los ítems en dos pruebas de aptitud

- intelectual en un contexto bilingüe. *Psicothema*, 12(Suplemento), 214–219.
- French, A. W., & Miller, T. R. (1996). Logistic Regression and Its Use in Detecting Differential Item Functioning in Polytomous Items. *Journal of Educational Measurement*, 33(3), 315–332. <http://doi.org/10.1111/j.1745-3984.1996.tb00495.x>
- Fuster, J. M. (2005). *Cortex and Mind: Unifying Cognition: Unifying Cognition*. Oxford University Press, USA.
- Hambleton, R. K., & Jong, J. H. A. L. de. (2003). Advances in translating and adapting educational and psychological tests. *Language Testing*, 20(2), 127–134. <http://doi.org/10.1191/0265532203lt247xx>
- Hambleton, R. K., Merenda, P. F., & Spielberger, C. D. (2006). *Adapting Educational and Psychological Tests for Cross-Cultural Assessment*. Routledge.
- Hambleton, R. K., & Patsula, L. (1998). Adapting tests for use in multiple languages and cultures. *Social Indicators Research*, 45(1-3), 153–171.
- Hambleton, R. K., & Patsula, L. (1999). Increasing the validity of adapted tests: Myths to be avoided and guidelines for improving test adaptation practices. *Association of Test Publishers*, 1(1), 1–13.
- Hambleton, R. K., & Zenisky, A. L. (2010). Translating and adapting tests for cross-cultural assessments. In D. Matsumoto & F. J. R. Van De Vijver (Eds.), *Cross-Cultural Research Methods in Psychology* (pp. 46–70). Cambridge University Press.

Hidalgo Montesinos, M. D., Gómez Benito, J., & Padilla García, J. L. (2005).

Regresión logística: alternativas de análisis en la detección del funcionamiento diferencial del ítem. *Psicothema*, 17(3), 509–515.

International Test Commission. (2010). Guidelines for translating and adapting tests, 1, 2012.

Lezak, M. D., Howieson, D. B., Loring, D. W., Julia, H., & Fischer, J. S. (2004).

*Neuropsychological assessment (4th ed.)* (Vol. xiv). New York, NY, US: Oxford University Press.

Llorente, A. M. (2008). Chapter 2. In A. M. Llorente (Ed.), *Principles of*

*Neuropsychological Assessment with Hispanics: Theoretical Foundations and Clinical Practice*. Springer.

Longford, N. T., Holland, P. W., & Thayer, D. T. (1993). Stability of the MH D-

DIF statistics across populations. Retrieved from

<http://psycnet.apa.org/?fa=main.doiLanding&uid=1993-97193-009>

Mantel, N. (1963). Chi-square tests with one degree of freedom; extensions of

the Mantel-Haenszel procedure. *Journal of the American Statistical Association*, 58(303), 690–700.

Matsumoto, D., & Van de Vijver, F. J. R. (2010). *Cross-Cultural Research*

*Methods in Psychology*. Cambridge University Press.

Milfont, T. L., & Fischer, R. (2010). Testing measurement invariance across

groups: Applications in crosscultural research. *International Journal of Psychological Research*, 3(1), 111–121.

Morales, M., Campo, P., Fernandez, A., Moreno, D., Yanez, J., & Sanudo, I.

(2010). Normative Data for a Six-Trial Administration of a Spanish

- Version of the Verbal Selective Reminding Test. *Archives of Clinical Neuropsychology*, 25(8), 745–761. <http://doi.org/10.1093/arclin/acq076>
- Ostrosky-Solis, F., Gomez-Perez, M. E., Matute, E., Rosselli, M., Ardila, A., & Pineda, D. (2007). NEUROPSI ATTENTION AND MEMORY: A neuropsychological test battery in Spanish with norms by age and educational level. *Applied Neuropsychology*, 14(3), 156–170.
- Penfield, R. D. (2005). DIFAS: Differential Item Functioning Analysis System. *Applied Psychological Measurement*, 29(2), 150–151. <http://doi.org/10.1177/0146621603260686>
- Penfield, R. D. (2010). Modeling DIF Effects Using Distractor-Level Invariance Effects: Implications for Understanding the Causes of DIF. *Applied Psychological Measurement*, 34(3), 151–165. <http://doi.org/10.1177/0146621609359284>
- Puente, A. E., & Ardila, A. (2000). Neuropsychological assessment of hispanics. In *Handbook of Cross-Cultural Neuropsychology* (pp. 87–104). Springer Science & Business Media.
- Rao, C. R., & Sinharay, S. (2007). *Psychometrics*. Elsevier.
- Sireci, S. G. (2010). Evaluating test and survey items for bias across languages and cultures. In D. Matsumoto & F. J. R. Van De Vijver (Eds.), *Cross-Cultural Research Methods in Psychology* (pp. 216–240). Cambridge University Press.
- Swaminathan, H., & Rogers, H. J. (1990). Detecting Differential Item Functioning Using Logistic Regression Procedures. *Journal of Educational Measurement*, 27(4), 361–370. <http://doi.org/10.1111/j.1745-3984.1990.tb00754.x>

U. S. Census Bureau, D. I. S. (2005). 2005 Interim State Population Projections.

Retrieved June 26, 2013, from

<http://www.census.gov/population/projections/data/state/projectionsagesex.html>

Zenisky, A. L., Hambleton, R. K., & Robin, F. (2003). Detection Of Differential Item Functioning In Large-Scale State Assessments: A Study Evaluating A Two-Stage Approach. *Educational and Psychological Measurement*, 63(1), 51–64. <http://doi.org/10.1177/0013164402239316>