

BID T 1505

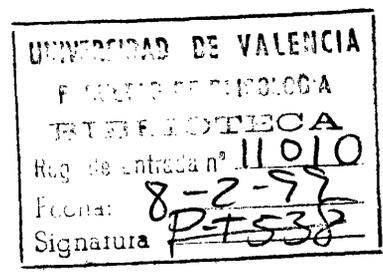
UNIVERSITAT DE VALENCIA

Facultat de Psicologia.



FUNCIONAMIENTO DIFERENCIAL DE LOS ITEMS DE UNA PRUEBA DE APTITUD INTELECTUAL EN FUNCION DE LA LENGUA FAMILIAR Y LA LENGUA DE ESCOLARIZACION

TESIS DOCTORAL



Presentada por: DORIS FERRERES TRAVER

DIRECTORES: Dr. VICENTE GONZALEZ ROMA  
Prof. Titular de Metodologia  
Universitat de València

Dra. JUANA GOMEZ BENITO  
Prof. Titular de Metodologia  
Universitat de Barcelona

València, Juliol de 1998.

Handwritten signature and stamp: P-538



60000767526  
Psicologia



UMI Number: U607369

All rights reserved

INFORMATION TO ALL USERS

The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if material had to be removed, a note will indicate the deletion.



UMI U607369

Published by ProQuest LLC 2014. Copyright in the Dissertation held by the Author.  
Microform Edition © ProQuest LLC.

All rights reserved. This work is protected against  
unauthorized copying under Title 17, United States Code.



ProQuest LLC  
789 East Eisenhower Parkway  
P.O. Box 1346  
Ann Arbor, MI 48106-1346

D. 769517  
L. 769526

Als meus pares,  
per estar sempre aquí



## **AGRADECIMIENTOS**

La tesis doctoral que se presenta a continuación ha sido posible gracias a la participación de un gran número de personas sin las cuales no hubiera sido probable la realización de la misma. Quisiera destacar que los momentos de preparación, desarrollo y posterior maduración de esta tesis han supuesto para su autora estar inmerso en un proceso de continuo aprendizaje y enriquecimiento tanto a nivel personal como intelectual, el cual ha sido posible a los consejos, conocimientos y habilidades aportadas por las personas más cercanas a mí. En este sentido, tienen especial mención los directores de esta tesis. Quiero manifestar mi más sincero agradecimiento a Vicente González Romá. Trabajar juntos ha supuesto una ayuda inestimable para mí, dándome en numerosas ocasiones valiosos consejos, abriéndome nuevos horizontes y fomentando mi interés por distintos temas. Asimismo, cabe señalar su afán de superación y perfeccionamiento el cual ha sabido transmitir a todo el equipo de investigación que dirige, y no menos importante, la comprensión y calor humano prestado en estos últimos momentos. En cuanto a Juana, agradezco todos sus consejos y directrices prestadas las cuales me han permitido analizar la problemática planteada desde perspectivas diferentes.

Existen muchas otras personas que también han hecho posible el desarrollo de la misma. Quisiera expresar mi agradecimiento a mis compañeros y compañeras de trabajo por su gran apoyo a nivel profesional como personal. Doy gracias a Ana por su amistad y comprensión dentro y fuera del ámbito laboral en especial durante estos últimos meses de intenso trabajo, a Inés por la ayuda prestada en la recogida de datos y las



intensas conversaciones mantenidas con ella durante nuestros viajes por toda la Comunidad Valenciana, a Begonia por su ayuda y ofrecimiento ante mis problemas con los ordenadores, a Susana por su amistad y acertados comentarios que me han permitido ver las cosas de otro modo, y en general a todos los miembros del área que han compartido conmigo el nacimiento de este trabajo. También, y no menos importantes son los agradecimientos dirigidos al personal de los diferentes gabinetes Psicopedagógicos participantes en este estudio, ya que sin su ayuda y colaboración no sería posible la presentación de este trabajo.

Finalmente, quiero recordar aquí la gran ayuda que me han proporcionado "los míos", en especial a mis padres y hermanas. Su cariño, apoyo y comprensión han sido muy importantes para mí. Y por último, a todos mis amigos que han sabido llegar hasta el final y permitir que se haga realidad este proyecto.

Vinaròs, Juliol de 1998.





De vegades, la memòria ens emporta a la nostra infantesa. En el seu record, aquesta va tindre un lloc privilegiat. El meu 'abuelito' i jo passavem molt de temps junts, assentats sobre tot al 'brancalet'. Des d'aquell lloc, el meu "abuelito" em va ensenyar moltes coses, però la poesia era una de les seues preferides. Per tant, avui día tan esperat vull recordar-li una de elles:

La palabra es don divino artífice del pensamiento,  
recojerá la esencia de la flor del camino que hubiera  
sin ella volado con el viento.  
Aquí está prisionera del idioma en la letra sonora  
Realidad verbal de la quimera o  
verbo quimérico de la realidad de la hora.  
Y a tí que sabes descifrar lo misterioso mujer o hombre  
se ofrece a tu espíritu espacio luminoso  
la luz imponderable que en el bloque del signo se estremece.

Per a tu.





## ÍNDICE.

<b>Problemática actual en la medición y evaluación de las aptitudes intelectuales en los escolares de la Comunidad Valenciana.....</b>	<b>1</b>
--	----------

### **Capítulo 1. Bilingüismo y Educación Bilingüe**

<b>1.1. CONCEPTO, TIPOS Y MEDIDA DEL BILINGÜISMO</b>	
1.1.1.Introducción.....	9
1.1.2.Concepto de Bilingüismo.....	10
1.1.3.Tipos y Situaciones Bilingües.....	14
1.1.4.Medida del Bilingüismo.....	17
1.1.4.1.El sesgo en la medida de la competencia bilingüe.....	21
<b>1.2.PERSPECTIVA HISTORICA, CONCEPTO Y PROGRAMAS DE EDUCACION BILINGÜE</b>	
1.2.1.Introducción.....	25
1.2.2.Factores socio-históricos en la Educación Bilingüe.....	26
1.2.3.Concepto de Educación Bilingüe.....	31
1.2.4.Clasificación de los programas de Educación Bilingüe.....	32
<b>1.3.LA INVESTIGACION EMPIRICA EN LA RELACION BILINGÜISMO E INTELIGENCIA.</b>	
1.3.1.Introducción.....	37
1.3.2.Primer Etapa: Comparaciones cuantitativas.....	39
1.3.3.Segunda Etapa: Diferencias en la estructura de las capacidades cognitivas.....	46
1.3.4.Tercera Etapa:Relaciones entre procesos cognitivos específicos y bilingüismo.....	51
<b>1.4.MARCO TEORICO DEL BILINGÜISMO EN EL ÁMBITO EDUCATIVO.</b>	
1.4.1.Principio o Hipótesis del umbral.....	57
1.4.2.Principio de Interdependencia o Transferencia Lingüística...65	
1.4.3.Principio del Bilingüismo Aditivo y Sustractivo.....	69
1.4.4.Principio de la Significatividad de Aprendizaje.....	72
1.4.5.Principios de Complejidad, Contextualización y Continuidad..74	
1.4.6.Principio del Valor Simbólico de la Lengua.....	76

1.5.ESTUDIOS EMPIRICOS: EXPERIENCIAS EN COMUNIDADES Y EN PAISES.

1.5.1.Introducción.....	79
1.5.2.Programas de inmersión en Canadá.....	80
1.5.3.Programas de Educación Bilingüe en el Estado Español.....	87
1.5.3.1.Evaluación de programas en Cataluña.....	89
1.5.3.2.Evaluación de programas en Galicia.....	96
1.5.3.3.Evaluación de programas en el País Vasco.....	97

**Capítulo 2.La situación bilingüe en la Comunidad Valenciana**

2.1.INTRODUCCION.....	103
2.2.PERSPECTIVA HISTORICA.....	105
2.3.LA SITUACION LINGÜISTICA EN LA SOCIEDAD VALENCIANA.....	109
2.4.LA EDUCACION BILINGÜE EN LA COMUNIDAD VALENCIANA.....	115
2.4.1.Evaluación de programas educativos valencianos.....	120
2.5. CONCLUSIONES.....	130

**Capítulo 3.Evaluación del funcionamiento diferencial de items**

3.1.INTRODUCCION.....	137
3.2.ANTECEDENTES HISTORICOS.....	138
3.3.FUNCIONAMIENTO DIFERENCIAL DEL ITEM (DIF) E IMPACTO.....	138
3.4.REVISION CRITICA DE MÉTODOS EN LA DETECCION DEL FDI.....	150
3.4.1.Procedimientos basados en la TRI.....	157
3.4.1.1. $\chi^2$ de Lord.....	161
3.4.1.2.Medidas de Área.....	164
3.4.1.3.Comparación de modelos.....	172
3.4.2.Procedimientos que no especifican ningún modelo de medida.....	174
3.4.2.1.Método de Mantel-Haenszel (MH).....	174
3.4.2.2.Estandarización (DPE).....	185
3.4.2.3.Modelos loglineales y modelos logit.....	189
3.4.2.4.Regresión Logística (RL).....	193
3.4.2.5.SIBTEST.....	196
3.4.3.Desarrollos actuales y futuros.....	202

3.5.RELEVANCIA DEL SESGO EN LA INVESTIGACION EMPIRICA.....	205
---	-----

#### **Capítulo 4. Objetivos y Metodología.**

4.1.PLANTEAMIENTO DE LA INVESTIGACION.....	229
4.2.VARIABLES.....	235
4.3.HIPOTESIS.....	243
4.4.INSTRUMENTOS DE MEDIDA:	
4.4.1.Prueba de Aptitud: "B.A.D.y.G.".....	253
4.4.1.1.Habilidad mental verbal (M.H.V).....	256
4.4.1.2.Habilidad mental no verbal (H.M.Nv).....	258
4.4.2.Encuesta sociolingüística.....	260
4.5.MUESTRA .....	261
4.5.1.Procedimiento de recogida de datos.....	261
4.5.2. Descripción de la muestra.....	265
4.6. ANALISIS.....	277
4.6.1. fiabilidad.....	278
4.6.2. Dimensionalidad.....	278
4.6.3. Análisis estadísticos de funcionamiento diferencial.....	281

#### **Capitulo 5.Resultados**

5.1.Análisis descriptivos.....	287
5.2.Unidimensionalidad.....	289
5.3.Evaluación del funcionamiento diferencial.....	293

#### **Capitulo 6.Conclusiones.....**

<b>Referencias bibliográficas.....</b>	<b>331</b>
--	------------

<b>Anexos.....</b>	<b>353</b>
--------------------	------------



**Problemática actual en la medición y evaluación de  
las aptitudes intelectuales en escolares de la Comunidad  
Valenciana: a modo de introducción**

\*\*\*\*\*

La situación lingüística de la enseñanza en el Estado Español ha sufrido un gran cambio en los últimos años, pasando de una política de unificación a otra donde se reconoce y promociona la pluralidad lingüística. Una de las consecuencias más relevantes ha sido la posibilidad de diversas Comunidades Autónomas de poder reconocer y utilizar la lengua propia de su territorio como lengua oficial e introducirla en el sistema educativo, generando un número importante de programas de Educación Bilingüe (Vila Moreno, 1986, 1991). Entre ellas, se encuentra la Comunidad Valenciana.

La Comunidad Valenciana se caracteriza por ser una comunidad bilingüe donde conviven individuos que hablan valenciano y castellano, y en la que ambas lenguas son oficiales. Tras la aprobación del "Estatut d'Autonomia" (1983) en este territorio, el proceso de normalización lingüística del valenciano ha tenido importantes consecuencias en la sociedad valenciana, sobre todo en el ámbito educativo. Quizás una de las más relevantes ha sido la posibilidad de poder cursar la enseñanza obligatoria bien en castellano o bien en valenciano, junto con sus posibles variantes curriculares.



Sin embargo, la implantación de la enseñanza bilingüe en el sistema educativo valenciano no ha estado exenta de ciertas controversias, las cuales han teniendo como principal elemento de discusión uno de los siguientes argumentos: (1) la utilidad o no de introducir el valenciano como materia de enseñanza y/o lengua de instrucción, y (2) relacionada con la anterior, las posibles carencias que la presencia de la lengua propia en el curriculum escolar puede originar en el dominio lingüístico del castellano de los niños/as valencianos/as (Neel, 1991). Precisamente, esta realidad ha determinado enormemente la política lingüística adoptada en la Comunidad Valenciana, la cual se ha relacionado más con los derechos individuales y/o colectivos de los propios valencianos de recibir o no una enseñanza bilingüe, que en los resultados procedentes de la evaluación de los diferentes modelos de Educación Bilingüe implantados en esta Comunidad (Bueno y Pascual, 1995; Díez, 1995; Gotor, Miralles y Perea, 1990, 1993; Pascual y Doménech, 1990; entre otros).

En la Comunidad Valenciana, la evaluación de los programas de enseñanza bilingüe se ha centrado básicamente en el estudio de cómo inciden los distintos modelos educativos implantados en el rendimiento académico global de los escolares valencianos. En éstos, la evaluación del rendimiento académico se ha operacionalizado fundamentalmente en dos formas diferentes: (a) mediante el estudio de las habilidades lectoescritoras adquiridas a lo largo del proceso de escolarización en las dos lenguas oficiales de dicho territorio, y (b) la aplicación de tests de inteligencia con la finalidad de medir y evaluar las aptitudes intelectuales de los escolares valencianos. El interés de esta investigación se centra en la

segunda opción, esto es, en la *medición y evaluación de las aptitudes intelectuales de los alumnos escolarizados en los distintos modelos educativos implantados en la Comunidad Valenciana.*

Hoy en día, la incorporación de los tests psicológicos en la evaluación de las aptitudes intelectuales es una práctica psicológica y educativa extendida en todo el mundo, y todo parece indicar que va a seguir siéndolo en el futuro. Sin embargo, una de las críticas más importantes a los tests psicológicos y educativos es su mala utilización, aspecto especialmente delicado en el ámbito educativo donde con frecuencia están en juego decisiones de suma importancia para la vida de las personas. En ocasiones, este uso inadecuado de los tests puede llevar implícito la posibilidad de que la aplicación de ciertos tests o items del test puedan favorecer injustamente a los miembros de un grupo sobre otro y, en consecuencia, identificar erróneamente diferencias entre ellos. La enorme trascendencia de este fenómeno se pone de manifiesto cuando se constata la excesiva frecuencia con que se aplica un mismo instrumento a grupos mayoritarios y a otros más o menos minoritarios que conviven en una misma zona geográfica o política; o si es uno de los dos grupos, precisamente el grupo dominante el que construye los tests para toda la sociedad. Así pues, dadas las serias implicaciones éticas, educativas, sociales y políticas que conlleva este hecho, esto es, la posibilidad de infravalorar sistemáticamente las capacidades psicológicas y/o educativas de ciertos grupos en función de su lengua, cultura o cualquier otra característica diferenciadora en los grupos a comparar, el presente trabajo de investigación surge ante la preocupación de conocer *qué tipo de repercusiones educativas y sociales tiene la aplicación colectiva de*

*pruebas aptitudinales elaboradas únicamente en la lengua mayoritaria en una situación de contacto de lenguas, donde coexisten dos lenguas oficiales y, además, los programas de educación bilingüe son una realidad educativa.*

Un caso práctico se localiza en la Comunidad Valenciana. Dentro de esta comunidad, la medición y evaluación de las aptitudes intelectuales de los/as niños/as valencianos/as es una práctica educativa generalizada en la mayoría de los centros escolares, sin embargo ésta presenta una importante peculiaridad: *la aplicación de tests redactados y baremados en castellano a alumnos en proceso de escolarización de programas de enseñanza bilingüe.* Esta situación ocasiona la aparición de una serie de interrogantes acerca de la medición y evaluación de las aptitudes intelectuales de los escolares valencianos:

*¿Los/as niños/as cuya lengua familiar es el Valenciano y que cursan un programa bilingüe en Valenciano, presentan ciertos "handicaps" cuando se enfrentan a tests elaborados en castellano, y/o cuando se les compara con niños/as cuya lengua familiar y escolar es el Castellano?*

*¿Tienen también ese "handicap" y dificultad añadida los/as niños/as cuya lengua familiar es el Valenciano pero su lengua de enseñanza es el Castellano?*

*¿Es posible que la existencia de este "handicap" esté determinado por los años de escolarización en una determinada lengua?*

Partiendo de esta realidad, el *objetivo principal* de este trabajo de investigación es analizar si los items de un test de aptitud, elaborado y baremado en castellano y de uso habitual en el ámbito educativo valenciano, favorecen sistemáticamente a los miembros de un grupo sobre otro, igualmente capaces, pero con características lingüísticas distintas, como por ejemplo poseer una lengua familiar y/o lengua de escolarización diferente.

Las conclusiones aportadas por este estudio, en caso de obtener resultados positivos en cuanto a la presencia de items con sesgo en el test analizado, contribuirán sin duda a introducir ciertas mejoras en la evaluación de las aptitudes intelectuales de los escolares valencianos y, en consecuencia, a mitigar las importantes repercusiones educativas y sociales que pueden derivarse del uso de tests psicológicos con problemas de funcionamiento diferencial en sus items tras ser aplicados a grupos lingüísticos diferentes para la realización de diagnósticos, asesoramiento y orientación vocacional en los escolares valencianos.

Por último y para finalizar, me gustaría expresar mi deseo de que la realización de este estudio de investigación contribuya modestamente a una reflexión sobre "el sistema educativo valenciano", teniendo en cuenta que la realidad educativa valenciana siempre debe abordarse desde una situación de contacto de lenguas. De esta forma, el conocimiento del sistema educativo valenciano, junto con una evaluación adecuada del mismo, nos permitirá progresar hacia una igualdad de oportunidades educativas para todos los escolares de la Comunidad Valenciana.



## ***CAPITULO 1***

### **Bilingüismo y Educación Bilingüe**



## **1.1. CONCEPTO, TIPOS Y MEDIDA DEL BILINGÜISMO**

### **1.1.1. Introducción**

En este primer capítulo se intentará delimitar conceptualmente el término bilingüismo. Dentro de este área de estudio, es imprescindible dejar claro los límites de algunas de las definiciones más utilizadas para este concepto, puesto que en el campo del bilingüismo y temas afines, se produce a veces una cierta confusión terminológica que dificulta con frecuencia una comprensión exacta de aquéllo que se está estudiando. A continuación, se presentará una clasificación de distintos tipos y situaciones bilingües, haciendo especial referencia a aquellos grupos bilingües que aparecen con más frecuencia en la bibliografía sobre bilingüismo y que, a partir de los estudios empíricos realizados han demostrado tener más importancia a la hora de diferenciar distintos comportamientos en rendimiento escolar o desarrollo cognitivo al comparar monolingües y bilingües.



Por último, se abordará el bilingüismo como objeto de medida, centrándonos en el problema de la definición y sesgo en la medida de la competencia lingüística del bilingüe. Cualquier profesional que tenga que tratar con personas bilingües tiene que conocer la importancia de una buena medida del dominio en los dos códigos lingüísticos de una persona bilingüe; incluso cuando la labor de éste no consista en realizar dicha medida. Así pues, es muy importante que todo profesional posea unos conocimientos básicos acerca de la medida que le permitan discernir si los tests relativos a la competencia lingüística de los bilingües son aplicados en las circunstancias adecuadas y con la finalidad para la que fueron creados.

#### **1.1.2. Concepto de Bilingüismo.**

En nuestros días, podemos encontrar entre 200 estados soberanos y, al menos, 4000 lenguas diferentes en todo el mundo (Siguán y Mackey, 1986), lo que nos da una idea de la diversidad de relaciones y situaciones que pueden producirse en las situaciones de lenguas en contacto, y la gran variedad de circunstancias de las personas en relación a sus propias lenguas. Habitualmente, todo conjunto de relaciones y situaciones en las cuales se encuentran presentes dos o más lenguas queda agrupado bajo un único término: BILINGÜISMO.

Sin embargo, es evidente que un único concepto es demasiado ambiguo para definir adecuadamente tanta infinidad de relaciones y contextos bilingües. Esta situación queda totalmente demostrada en los numerosos intentos por parte de diferentes investigadores, de ofrecer una definición

correcta y apropiada de este término. Prueba de ello es la amplia gama de definiciones encontradas en la literatura especializada sobre este tema. Además, a medida que avanza la investigación en el campo del bilingüismo, van surgiendo nuevas maneras de entender tanto la realidad como la persona bilingüe lo cual todavía dificulta más una delimitación adecuada de este término.

A continuación, de manera ilustrativa vamos a presentar algunas de las definiciones más destacadas en la historia del bilingüismo, sin dejar de negar que, hoy en día, es prácticamente imposible ofrecer una definición válida y ajustada para todas las situaciones donde convive la presencia de más de una lengua:

Bloomfield (1935): "El bilingüismo consiste en la capacidad de hablar dos lenguas de la misma manera que aquellas personas que las tienen como lengua materna".

Weinreich (1953): "Se utiliza el término *bilingüismo* al uso habitual de utilizar dos lenguas alternativamente, y a las personas que lo practican *bilingües*. Si no se especifica lo contrario, todo lo que decimos sobre el bilingüismo también es válido para el multilingüismo, que consiste en el hábito de utilizar tres o más lenguas alternativamente".

MacNamara (1967): "Tener competencia lingüística en un segundo idioma en una de las siguientes habilidades lingüísticas: comprender, hablar, leer y escribir".

Sánchez-López (1980): "es la maestría igual, activa y perfecta de dos o más lenguas".

Hamers y Blanch (1982): "Estado de una comunidad en la cual hay dos lenguas en contacto y, por tanto, dos códigos que pueden ser utilizados en una misma interacción. También, incluye la bilingüidad a nivel psicológico individual".

MacKey (1987): "Es evidente que si queremos estudiar el bilingüismo no hay más remedio que considerarlo como un fenómeno absolutamente relativo. Este fenómeno incluye no sólo el uso de dos lenguas sino, también, el uso de un número indeterminado de lenguas. Por tanto, consideraremos bilingüismo al uso alternativo de dos o más lenguas en el mismo sujeto".

Siguán y MacKey (1986): "Bilingüe es aquella persona que, a parte de su propia lengua, posee una competencia lingüística parecida en otra lengua, y además es capaz de utilizar las dos en cualquier circunstancia con una eficacia equivalente".

Genesee-Tucker (1990): "Capacidad para expresarse en dos idiomas en una de las siguientes maneras: como si ambas fueran lenguas nativas del propio hablante (bilingüismo auténtico), como si ninguna de ellas lo fuera (semilingüismo) o utilizando un idioma en determinados contextos culturales y el otro en otros contextos (disglosia)".

Como se aprecia, las definiciones presentadas aquí engloban grados y niveles de bilingüismo muy diferentes, debido a que el término bilingüe abarca desde definiciones poco exigentes como la presentada por MacNamara (1967), que marca un nivel inferior de dominio en ambas lenguas en alguna de las habilidades lingüísticas señaladas, hasta definiciones más estrictas como las ofrecidas por Bloomfield (1935) o Sánchez-López (1980), que exigen una competencia lingüística similar en ambas lenguas. Para un repaso de las principales definiciones de este término puede consultarse Sánchez-López (1980).

Generalizando a partir de ellas, podemos concluir que el término bilingüismo no debe ser considerado un **concepto absoluto** sino **relativo**. Es decir, la idea de un dominio equilibrado de dos o más lenguas en una persona bilingüe es realmente un tópico, puesto que el equilibrio en dos códigos lingüísticos en una misma persona es muy difícil de conseguir. Sin embargo, esta idea puede resultar útil como marco de referencia en el estudio del bilingüismo, ya que la propia experiencia muestra la existencia de muchos tipos de personas y situaciones bilingües muy diferentes entre sí, que se sitúan a lo largo de un continuo relativo a su competencia lingüística, desde niveles muy elementales hasta niveles superiores en el dominio lingüístico de cada una de las lenguas. Así pues, partiendo de este razonamiento y siguiendo a Etxeberria (1992), en la actualidad resulta mucho más apropiado defender que el bilingüismo implica algo más que un simple estado definido a priori de competencia lingüística en dos o más lenguas; es decir, que el bilingüismo se caracteriza por ser un proceso dinámico en situaciones tanto individuales como colectivas, que admite

muchas gradaciones y que dada la influencia recibida por diversos factores está constantemente cambiando.

Por lo tanto, atendiendo a todo lo expuesto aquí la idea defendida en este trabajo de investigación acerca del bilingüismo se basa en que éste representa a un fenómeno *activo, dinámico*, dado que admite multitud de variantes y posibilidades en las personas abarcando desde niveles simples hasta complejos en la competencia lingüística alcanzada, que se encuentra en especial *presente y*, a su vez, *en proceso de evolución* en la vida de aquellas personas y/o colectividades que conviven en situaciones donde coexisten dos o más lenguas diferentes.

### **1.1.3. Tipos y Situaciones Bilingües**

Basándonos en la conceptualización sobre el término bilingüismo propuesta aquí, la forma más idónea de estudiar este fenómeno no consistirá en formular preguntas acerca de si un individuo es o no bilingüe, sino más bien en averiguar en qué aspectos y/o contextos una determinada persona es bilingüe, lo cual dará lugar a diferentes tipos y situaciones bilingües. Por ejemplo, un mismo individuo puede ser bilingüe compuesto y su bilingüismo ser dominante y no equilibrado. Esta afirmación se hará más clara a medida que avancemos en la lectura de este apartado.

Sin embargo, dada la diversidad de relaciones y circunstancias que se producen en un determinada situación de contacto de lenguas, un recuento adecuado podría hacerse interminable. No obstante, a fin de que ello fuera posible, a continuación ofrecemos una clasificación de tipos y situaciones

bilingües diferentes, tomando en consideración aquellos criterios que mayor importancia y frecuencia de uso han tenido en la investigación científica sobre el bilingüismo.

Respecto a **las relaciones entre lenguaje y pensamiento**, la clasificación más conocida y que más trabajos de investigación ha producido es la distinción propuesta por Ervin y Osgood (1954) entre *bilingües compuestos* y *bilingües coordinados*. El *bilingüe compuesto* es aquél que aprende su segunda lengua (L<sub>2</sub>) relacionando cada palabra nueva con su equivalente de traducción en su primera lengua (L<sub>1</sub>), y da a estos equivalentes de traducción un significado idéntico en las dos lenguas. Esto parece deberse al hecho de que los dos idiomas han sido aprendidos en el mismo contexto. El *bilingüe coordinado* es aquél que adquiere su primera y segunda lengua en circunstancias distintas y, como consecuencia, no tiene un significado idéntico para los equivalentes de traducción en los dos idiomas. Esta distinción implica claramente que no todos los bilingües organizan el material verbal de la misma manera. De hecho, algunos estudios empíricos (Genesse y Hamayan, 1978; Diaz, 1986) afirman que los bilingües coordinados tienen una organización del contenido verbal más independiente que los bilingües compuestos. Los datos más recientes señalan la necesidad de considerar esta distinción, dado que los bilingües pueden variar entre dos polos de un mismo continuo, y no como dos categorías puras y excluyentes entre sí (Hamers y Blanc, 1983).

Según **la competencia alcanzada en cada uno de los dos idiomas**, se puede hablar de *bilingüismo equilibrado* cuando existe una competencia lingüística bastante similar en las dos lenguas, y de

*bilingüismo dominante* cuando un idioma se conoce mejor que otro. En este último caso pueden haber diversos grados de dominio o competencia de una lengua respecto a la otra u otras.

En relación al tercer criterio, **la edad de adquisición de las dos lenguas**, hablamos de *bilingüismo simultáneo* o *de infancia*, y *bilingüismo sucesivo* o *de adulto*. Se denomina *bilingüismo simultáneo* cuando el individuo aprende sus dos lenguas a la vez, por tanto es una adquisición que siempre se produce durante la primera infancia. Por el contrario, cuando se habla de *bilingüismo sucesivo*, el individuo aprende primero una lengua (que normalmente es su L<sub>1</sub>) y después la otra (que habitualmente será su L<sub>2</sub>).

El cuarto criterio alude al **prestigio sociocultural de las dos lenguas**. Aquí tiene cabida la distinción establecida por Lambert (1974) entre el *bilingüismo aditivo* y el *bilingüismo sustractivo*. El primero se da cuando ambas lenguas son valoradas socialmente al mismo nivel y, por tanto, se produce un *bilingüismo "armonioso y productivo"*. El *bilingüismo sustractivo* aparece cuando la valoración social de la L<sub>2</sub> (la lengua mayoritaria) se produce a expensas de la L<sub>1</sub> (la lengua minoritaria).

Por último, según **la identidad cultural**, se puede distinguir entre el *bilingüismo bicultural* cuando hay una doble pertenencia e identidad cultural, el *bilingüismo monocultural* en L<sub>1</sub> o en L<sub>2</sub>, cuando la identidad cultural se produce sólo en uno de los dos idiomas, y el *bilingüismo acultural con anomía*, cuando el sujeto no se identifica con ninguna de las dos culturas.

Para el tema de estudio que nos ocupa aquí, la importancia de esta clasificación reside en el hecho de que ésta permite un análisis de la realidad bilingüe desde puntos de vista diferentes. De este modo, la consideración de estos criterios permitirá la creación de grupos bilingües distintos, lo cual puede repercutir en la aparición de comportamientos o resultados diferentes en los grupos considerados ante el mismo fenómeno: el bilingüismo. Además, como se verá más adelante esta distinción será un elemento clave cuando se aborden los temas relacionados con la educación bilingüe.

#### **1.1.4. Medida del Bilingüismo**

La falta de una definición apropiada del término bilingüismo ha repercutido negativamente en la campo de la medida de la competencia lingüística en los bilingües. Hoy en día, determinar el grado de bilingüismo de una persona bilingüe no es una labor fácil. En el campo del bilingüismo, todavía no existe ningún acuerdo entre los investigadores acerca de cómo se debe medir la competencia lingüística en una persona bilingüe, o cómo determinar hasta qué punto una lengua del bilingüe influye en la otra, o cómo determinar el papel que juegan cada una de las lenguas en la vida cotidiana del bilingüe. Tal vez, el profesional pueda delimitar hasta qué punto un individuo puede realizar una actividad en una lengua o en otra, pero tratar de describir el comportamiento bilingüe de esa persona en su totalidad es una tarea mucho más compleja y difícil de realizar.



Los propios expertos en el tema son conscientes de la dificultad que representa la medición y evaluación de la competencia bilingüe en las personas, lo cual resulta bastante evidente en las investigaciones empíricas realizadas sobre bilingüismo. Precisamente, algunos estudios propuestos en este campo han sido realizados bajo el supuesto de que los sujetos son bilingües, o porque viven en un área geográfica bilingüe, o porque asisten a un colegio con enseñanza bilingüe o por cualquier otra razón, pero sin que haya habido realmente una comprobación empírica de dicho fenómeno y de su grado de bilingüismo. Sin embargo, a pesar de la dificultad que conlleva la medida de la competencia bilingüe, algunas investigaciones sí han intentado realizar una medición de la misma. En ellas, la medida de la competencia lingüística se ha abordado bajo la siguiente premisa:

"Todos los seres humanos normales poseen una capacidad innata para la adquisición del lenguaje, desarrollan esta capacidad y llegan a ser competentes en ella, es decir, llegan a comunicarse de forma efectiva en la lengua que les rodea desde su nacimiento" (Sanchez-Lopez y Rodriguez, 1997, pp.61).

Siguiendo esta propuesta, la mayoría de las medidas elaboradas para evaluar el bilingüismo del individuo defienden la idea de que tomar medidas de cada una de las lenguas en el individuo y, posteriormente, compararlas para determinar su grado de dominancia es la técnica más adecuada para medir la competencia lingüística de la persona bilingüe. Así pues, la medida de la competencia bilingüe se elabora intentando responder a las siguientes dos preguntas: (a) ¿Cuál es el nivel de competencia que tiene el

bilingüe en cada una de sus lenguas?, y (b) ¿Cuál de sus lenguas es la dominante?. Básicamente, lo que se busca es determinar el status de las dos lenguas en el individuo, poniendo en relación una con la otra.

En la investigación empírica, uno de los índices más extendidos para evaluar la competencia lingüística de los sujetos bilingües en el **índice de bilingüidad** (Hamers y Blanch, 1983), que pretende medir la relación subjetiva entre dos o más idiomas en un individuo. Por ejemplo, cuando un sujeto asiste a programas de enseñanza bilingüe la relación entre sus dos lenguas evoluciona a lo largo del proceso de escolarización, por tanto, en este caso el índice de bilingüidad aporta la información necesaria para conocer qué idioma tiene mayor peso en la relación de dos idiomas durante su enseñanza obligatoria. Una aplicación práctica de este índice puede consultarse en el trabajo de Elosua, López y Artamendi (1994).

Sin embargo, en la mayoría de las investigaciones la metodología empleada para averiguar la lengua dominante de un individuo ha sido la administración de una encuesta sociolingüística donde se solicita información acerca de cuál es la primera lengua del sujeto (Candell y Hulin, 1987; Ellis, 1989; Bontempo, 1993), o cuál es la lengua materna o familiar del alumno (Miralles y cols, 1989); Serra y Vila, 1991; Bel, Serra y Vila, 1993), o cuál es la lengua dominante del entorno sociolingüístico del individuo (Pascual y Doménech, 1990; Bueno y Pascual, 1995; Diez, 1995). En este último caso, el entorno sociolingüístico del individuo se refiere a su ámbito familiar, escolar, amigos cercanos, familiares más próximos, etc.

No obstante, conforme ha ido avanzando la investigación en el campo del bilingüismo, la propia realidad ha ido demostrando que la competencia bilingüe no es únicamente la suma de dos estados monolingües, sino que es una condición específica, propia de una situación de lenguas en contacto, y como tal, su estudio debe abordarse desde esta perspectiva. Sin embargo, pese a los esfuerzos realizados por los investigadores, en la actualidad el estudio del bilingüismo se encuentra en una situación bastante incierta, ya que todavía no contamos con una descripción de la "competencia bilingüe" lo suficientemente definida como para poder medirla, y menos aún baremos o estándares para comparar las competencias entre distintos tipos de bilingües.

Así pues, la comparación directa entre dos o más lenguas en una persona bilingüe casi nunca ha sido posible, y aún en el caso de que dispusiéramos de medidas válidas de la competencia, éstas tendrían que ser comparables, es decir, que la medida de la competencia lingüística se abordara de forma idéntica en ambas lenguas. El principal problema es ¿cómo hacerlo?, sobre todo teniendo en cuenta la existencia de idiomas tan diferentes como son el castellano y el árabe, y la gran cantidad de diferentes tipos de personas y situaciones bilingües en el mundo. En términos generales, lo que se suele hacer para salvar esta dificultad es una doble comparación: *se compara cómo se sitúa el bilingüe en cada una de sus lenguas con respecto a las normas de la población monolingüe*. Por ejemplo, si a un bilingüe castellano/alemán se le aplica un test de competencia lingüística en castellano y otro en alemán, y en ambos obtiene una puntuación que le sitúa por encima del 70% de los sujetos, se puede afirmar que esa persona tiene una buena competencia lingüística en ambas

lenguas y, por tanto, es un bilingüe equilibrado. En este caso, la comparación entre la competencia en alemán y castellano se realiza utilizando la distribución de las competencias de los sujetos nativos en esas pruebas.

Sin embargo, a pesar de la falta de una definición y medida operativa acerca de la competencia lingüística en los bilingües, esta situación sí debe ser tomada en consideración al interpretar los resultados de las pruebas monolingües aplicadas a individuos bilingües en una situación de lenguas en contacto, ya que la interpretación de los resultados obtenidos puede llevar implícito un **problema de sesgo** en la medida de la competencia bilingüe. Dada la importancia del sesgo en la medida de este constructo, este tema será abordado con más profundidad en el siguiente apartado.

#### **1.1.4.1. El sesgo en la medida de la competencia bilingüe**

El problemática del sesgo en la medida de la competencia lingüística de los bilingües aparece principalmente en la clase de bilingües resultantes de una situación de lenguas en contacto, como por ejemplo, el castellano y el valenciano en la Comunidad Valenciana, y el caso de las minorías lingüísticas como pueden ser los inmigrantes hispanos en los Estados Unidos, entre otros.

Para muchas de estas personas, determinar cuál es su lengua nativa es muy difícil, ya que muchas de ellas son personas con dos o más lenguas perfectamente desarrolladas y aprendidas en una situación de contacto. Sin

embargo, a la hora de medir la competencia lingüística de estos bilingües no hay que olvidar que los bilingües no viven dos vidas, es decir, no pueden tener la misma práctica en ambas lenguas en todas las actividades que realizan. Por ejemplo, en algunos ámbitos como el hogar y las relaciones afectivas será dominante una de sus lenguas y, en otros, como el trabajo o las relaciones sociales lo será otra. Además, esta dominancia puede ir cambiando a lo largo de la vida de las personas dependiendo de sus propias vivencias en cada una de las lenguas, e incluso puede darse el caso de que la lengua conocida a nivel de competencia nativa llegue a ser olvidada, al dejar de ser utilizada.

El problema del sesgo en la medida del bilingüismo puede surgir cuando se intenta evaluar la competencia lingüística de estas personas con tests cuyo objetivo es la norma monolingüe de la lengua mayoritaria (L<sub>2</sub>), o con tests desarrollados y baremados con normas monolingües pero que son aplicados a personas que todavía están en el proceso de aprendizaje de la L<sub>2</sub>, es decir, que aún no poseen una competencia a nivel nativo de la L<sub>2</sub>. En estos casos, la competencia lingüística de los bilingües se compara con la norma de una lengua que no es la que ellos hablan en algunos casos, y no es la lengua que hablan casi todo el tiempo. Esta situación puede resultar en una interpretación errónea de resultados respecto a la competencia lingüística de los bilingües y, posiblemente, en la atribución de ciertas dificultades en los bilingües cuando éstas realmente no son suyas.

Precisamente, la investigación realizada sobre todo en el ámbito educativo muestra que, parte de los resultados negativos en rendimiento académico tanto en la primera lengua (L<sub>1</sub>) como en la segunda (L<sub>2</sub>), así como

en CI verbal de los grupos minoritarios podrían explicarse por una *utilización inadecuada* de los instrumentos de medida empleados (De Jong, 1987; Resing, Bleichrodt y Drenth, 1986; van de Vijver y Poortinga, 1991; Helms-Lorens y Van de Vijver, 1995). En estos casos, es muy probable que las escalas utilizadas para evaluar la competencia lingüística o rendimiento académico de los sujetos bilingües estuvieran diseñadas únicamente para medir el desarrollo del aprendizaje de la lengua mayoritaria (L<sub>2</sub>). Además, también es posible que los propios profesionales no tuvieran en cuenta la posible influencia que pueden causar la presencia de variables diferenciadores en los sujetos bilingües evaluados, como son el empleo de una lengua familiar distinta a la lengua de enseñanza, la escolarización en una lengua de enseñanza diferente a la lengua mayoritaria, y la aplicación de una prueba en una lengua distinta a la lengua familiar y/o académica.

Desgraciadamente, este tipo de situaciones suele ocurrir con bastante frecuencia en algunas situaciones de contacto donde coexisten dos o más lenguas y, además, se imparten programas de enseñanza bilingüe. Así pues, los resultados de cualquier evaluación sobre la competencia lingüística de los grupos bilingües deben interpretarse con sumo cuidado pues, si se utilizan tests desarrollados con normas monolingües, como suele ser habitual, se puede correr el riesgo de menospreciar el nivel de competencia lingüística alcanzado por los bilingües y, en consecuencia, ser seriamente perjudicados en sus futuras oportunidades educativas.



## **1.2. PERSPECTIVA HISTORICA, CONCEPTUALIZACION Y PROGRAMAS DE EDUCACION BILINGÜE**

### **1.2.1. Introducción**

En este segundo apartado, se analizará una de las variables situacionales más estudiadas en relación a las diferencias que se producen entre distintos grupos bilingües con respecto a la adquisición de una segunda lengua: *la Educación Bilingüe*. La aparición de estas diferencias juegan un papel decisivo en el aprovechamiento académico o desarrollo cognitivo de la persona bilingüe. En concreto, se abordarán los factores socio-históricos que hacen surgir la Educación Bilingüe y la condicionan, su conceptualización y una clasificación de los diferentes tipos de programas de educación bilingüe.



### **1.2.2. Factores socio-históricos en la Educación Bilingüe.**

En sus comienzos, el bilingüismo era considerado un fenómeno excepcional que afectaba únicamente a algunos individuos dentro de algunas comunidades o países bilingües en el mundo. Ante tal diversidad étnica y cultural, la manera de proceder más usual y extendida ha sido la implantación de la *asimilación a la lengua o cultura mayoritaria*. Esta actuación ha estado unida con frecuencia a la pérdida de los lazos lingüísticos y personales con los familiares y país de origen y, en consecuencia, a la oportunidad de profundizar en la propia lengua y cultura.

Un claro ejemplo de esta práctica se encuentra en algunos acontecimientos históricos de los Estados Unidos. Tras su independencia, los ingleses se mantuvieron como grupo dominante y el resto de los grupos no anglosajones como franceses, irlandeses, alemanes, centroeuropeos, asiáticos, africanos y los verdaderamente americanos, los indios, fueron en mayor o menor medida víctimas de la *discriminación*. Un elemento clave en este proceso fue el sistema educativo. La escuela se constituye como el elemento principal en el proceso de aculturación de los hijos de inmigrantes o minorías étnicas. De esta forma, la enseñanza se convierte en el instrumento integrador de ideales políticos, económicos y sociales comunes a todos los ciudadanos, especialmente de las minorías con lengua indígena, las cuales eran percibidas como un obstáculo para el bien común del grupo dominante o mayoritario.

Sin embargo, a principios del siglo XX, el tema del bilingüismo comienza a adquirir un matiz diferente a causa de una serie de cambios políticos, económicos, ideológicos y educativos acaecidos en todo el mundo. Por citar algunos ejemplos, diremos que la progresiva independencia de las antiguas colonias originó una multitud de países multibilingües. La mayoría de estos países adoptaban una lengua común de comunicación y, en ocasiones, la de los propios colonizadores, como en algunos países de Africa (en algunas zonas, además de la lengua tribal, se habla el swahili y el inglés) o la India (diversas variedades del hindi y el inglés, además de la lengua local). Otro acontecimiento relevante fueron los movimientos migratorios sucedidos en todo el Planeta, origen de los éxodos causados por revoluciones y guerras, como la de Cuba y Vietman, o las necesidades de mano de obra de los países occidentales, entre otros. En concreto, estos desplazamientos provocaron la creación de nuevos grupos minoritarios, lo que repercutió enormemente en la aparición del bilingüismo, especialmente si el número de hablantes de una lengua era una minoría en el país de origen. Un ejemplo a citar podría ser la emigración masiva de los irlandeses e hispanos a los Estados Unidos. En el caso de los irlandeses, éstos se asimilaron a la lengua inglesa conservando únicamente el valor de su religión, mientras que la inmigración en las comunidades hispanas de EEUU ha sido totalmente diferente; ésta ha contribuido a que la lengua minoritaria no se pierda, y mantenga su vitalidad durante generaciones (Baker, 1985).

Otro factor a destacar, tuvo lugar en la década de los 60. Al tiempo que se producía un rápido cambio económico y social, la sociedad occidental industrializada comenzó a manifestar una cierta disconformidad con las

normas establecidas hasta el momento, así lo demuestran algunos de los movimientos más significativos de aquella época, como son varias revueltas por universitarios la emancipación de los jóvenes, la libertad sexual, etc.. De igual manera, Estados Unidos experimentaba un nuevo proceso de lucha a favor del reconocimiento de las aportaciones de las minorías lingüísticas y/o culturales a la sociedad americana. En este sentido, el movimiento social en defensa de los derechos humanos focaliza todos sus esfuerzos en el desarrollo de nuevos programas educativos que reflejen la naturaleza multiétnica y multicultural de la sociedad americana y, en consecuencia, aseguren la igualdad de oportunidades para toda la población americana.

Las soluciones adoptadas ante la ocurrencia de estas circunstancias para la mayoría de los países bilingües fue la implantación de multitud de programas de *Educación Bilingüe* destinados fundamentalmente a las minorías lingüísticas, con el fin de llegar a un compromiso entre la enseñanza de la propia lengua y la necesidad de aprender la lengua de la mayoría. En concreto, el desarrollo en el campo del bilingüismo culmina en 1967 con la proclamación del Acta de la Educación Bilingüe, donde se reconoce "*el derecho de los/as niños/as de las minorías lingüísticas a recibir una educación en su lengua familiar, al mismo tiempo que aprenden el inglés como lengua académica*". De esta manera, un aumento de los derechos y libertades de las minorías lingüísticas junto con la necesidad de formar ciudadanos conocedores de lenguas y culturas diferentes son el impulso necesario para que los modelos o programas de enseñanza bilingüe y biculturales se conviertan en uno de los fenómenos educativos más importantes de nuestro tiempo. Este hecho se constata en países como

Canadá, Escandinavia o Nueva Zelanda, entre otros, donde se empieza a adoptar un enfoque más multicultural del sistema educativo.

En la otra cara del mundo, estas medidas de actuación dan origen a los primeros programas de *Educación Compensatoria* implantados en los Estados Unidos. La aparición de este tipo de programas supone un paso decisivo para la Psicología, ya que implica que la Psicología de la Educación va a adquirir un importante papel desconocido hasta entonces. Precisamente, su principal objetivo era transformar una realidad educativa que distaba mucho del principio de igualdad de oportunidades que la sociedad identifica como objetivo básico de todo sistema educativo.

Sin embargo, los países defensores de los programas de enseñanza en la lengua mayoritaria, como son Estados Unidos e Inglaterra, entre otros, apuestan por una educación basada en programas de submersión y formas transitorias de enseñanza bilingüe con el fin de asegurar una rápida conversión a la lengua mayoritaria y, por tanto, compensar las limitaciones de los niños o familias que se encontraban en situación de desventaja (Cummins, 1980). Su principal argumento era: "los/as niños/as de un país necesitan funcionar en la sociedad por medio de la lengua mayoritaria". Por tanto, la razón pedagógica que defendían era que si los alumnos de los grupos minoritarios no establecían rápidamente la competencia en la lengua mayoritaria podía suceder que éstos pudieran *ir por detrás* de sus compañeros de lengua mayoritaria, lo cual quedaría reflejado en un rendimiento académico o desarrollo cognitivo inferior de los grupos minoritarios.

Pero, a pesar de ello el sistema educativo vigente en esos países continuaba mostrando las características propias de una baja calidad educativa, especialmente para los grupos minoritarios, los cuales en comparación con el grupo mayoritario presentaban ciertos "handicaps" y/o retrasos en su desarrollo cognitivo o rendimiento académico, debido fundamentalmente a su procedencia social y/o étnica (Coleman, 1966). Lo que realmente provocó una transformación de la filosofía imperante ("lo mismo para todos") en la educación estadounidense fue una decisión tomada por la Corte Suprema. Ésta llegó a afirmar que 1800 estudiantes de San Francisco, cuya lengua familiar era el chino y cuya lengua de escolarización era el inglés, no estaban recibiendo las mismas oportunidades educativas, dado que *".... los estudiantes que no comprenden el inglés están efectivamente excluidos de toda comunicación con significado"* (Lau contra Nichols, 1974). Así pues, por primera vez el elevado fracaso escolar de los alumnos de familias inmigrantes y demás minorías lingüísticas se atribuye a la falta de consideración por parte del sistema educativo de algunas variables vinculadas a dichos sujetos, por ejemplo tener una lengua familiar distinta a la lengua académica o mayoritaria, y la falta de continuidad en la lengua empleada en el ámbito familiar y escolar.

Más próximo a nuestro entorno, aquí en España en la década de los 60 es el momento de la aparición de los primeros movimientos nacionalistas en Galicia, Cataluña y País Vasco, principalmente. En aquel momento, nuestro país defendía una política lingüística de unificación sobre la base de la lengua castellana. No obstante, la presencia de estos movimientos junto con la implantación de la democracia en el Estado Español hacen posible el surgimiento de los primeros procesos de normalización lingüística, basados

en los derechos lingüísticos de las personas, de las lenguas autóctonas de distintas comunidades. Y todo ello desemboca en la creación de un gran número de programas de educación bilingüe. En ellos se contempla la obligación de impartir parte del currículum escolar en las distintas lenguas autóctonas (Arnau, Comet, Serra y Vila, 1992).

### **1.2.3. Concepto de Educación Bilingüe.**

Una consulta de la literatura especializada muestra que el término *Educación Bilingüe*, al igual que el de bilingüismo es un concepto que cubre realidades educativas muy diversas. Por ello, en numerosas ocasiones esta misma expresión se ha aplicado a situaciones y contextos bilingües bastante diferentes entre sí (Sánchez y Rodríguez Tembleque, 1986). Por consiguiente, se hace necesario una delimitación conceptual del término Educación Bilingüe para distinguir qué sistemas educativos vamos a considerar bilingües.

Desde hace tiempo y coincidiendo con numerosos autores como Fishman (1976), Siguan y Mackey (1986) y Hamers y Blanc (1983), se viene aceptando que la definición de Educación Bilingüe propuesta por Fishman (1976) es la más acertada. Esta definición califica como bilingüe a "*Todo sistema de enseñanza en el cual, en proporciones variables, simultánea o consecutivamente, se imparte la instrucción en al menos dos lenguas, una de las cuales, normalmente, es la primera lengua del alumno*" (Fishman, 1976).

Su principal característica es que al menos dos lenguas se utilizan como vehículos de enseñanza de otros contenidos del currículum escolar, y no sólo como materia de enseñanza. En definitiva, *se enseña en dos lenguas*.

Como se puede apreciar, esta definición excluye los programas de enseñanza de lengua extranjera impartidos por profesores nativos, si la lengua es únicamente materia de enseñanza y no medio de instrucción de otros conocimientos curriculares. En ese caso, el contenido de la enseñanza es el aprendizaje de idiomas y no una práctica educativa bilingüe. De la misma manera, tampoco podemos considerar educación bilingüe los programas de enseñanza cuyos contenidos curriculares son enseñados en una lengua diferente a la del alumno, si exclusivamente utilizan esa lengua. Los alumnos de estos programas probablemente llegarán a ser personas bilingües, pero formalmente, de acuerdo con la definición anterior, estos programas no pueden ser considerados de educación bilingüe. Estos programas de cambio de lengua en el hogar y en la escuela inicialmente desarrollados con gran éxito en Canadá, se han convertido en un elemento fundamental para impulsar el estudio del bilingüismo educativo en todo el mundo, e incluso, para algunos autores, éstos son considerados programas de educación bilingüe.

#### **1.2.4. Clasificación de los programas de Educación Bilingüe.**

A pesar de las acotaciones anteriores, el concepto de Educación Bilingüe abarca realidades educativas muy diversas, a las cuales no se pueden aplicar sistemas educativos idénticos. Por tanto, en la propia realidad socioeducativa nos encontramos con una gran diversidad de programas de enseñanza bilingüe diferentes.

Partiendo de esta pluralidad, en muchas ocasiones se ha intentado establecer una agrupación de los diferentes modelos bilingües, pero aportar una clasificación con un consenso generalizado no es nada fácil. En este sentido, la literatura ofrece las clasificaciones de Mckey (1976, 1982), Fishman y Lowas (1970), Fishman (1977), Hamers y Blanch (1983), Skutnabb-Kangas (1978) y Siguán y Mackey (1986). Entre esta multitud de tipologías, aquí vamos presentar únicamente la clasificación de Fishman y Lovas (1970), citada en muchas referencias bibliográficas. Ésta agrupa los programas de educación bilingüe en tres grandes categorías, teniendo en cuenta la intensidad, el objetivo del programa y el status de cada una de las lenguas.

Según la **intensidad** con que se utilicen las lenguas, se pueden identificar cuatro tipos de educación bilingüe:

a) *Programas bilingües de transición*, en los que la primera lengua (L<sub>1</sub>) del bilingüe, que suele coincidir con la lengua materna, se utiliza sólo para facilitar el paso a la instrucción en la L<sub>2</sub>, que habitualmente es la lengua mayoritaria o de mayor prestigio económico y social.

b) *Programas bilingües asimétricos*, donde se utilizan las dos lenguas, pero todo el proceso de aprendizaje de la lectoescritura se realiza en la lengua mayoritaria.

c) *Programas bilingües complementarios*, donde se utiliza la lectoescritura en ambas lenguas, pero la lengua minoritaria se utiliza para las asignaturas consideradas menos importantes o más fáciles en el programa de enseñanza, mientras que la L<sub>2</sub> se utiliza para las asignaturas científicas y técnicas.



Los programas b y c son un ejemplo de diglosia escolar donde la lengua de la minoría (L<sub>1</sub>) se reserva para algunas funciones menos valoradas y no compartidas por la comunidad en su totalidad, mientras que la L<sub>2</sub> se reserva para las funciones más valoradas en la escuela y compartidas por toda la comunidad.

d) *Programas bilingües totales*, donde los alumnos desarrollan las cuatro destrezas (entender, hablar, leer y escribir) en las dos lenguas y para todos los ámbitos del currículum escolar.

El segundo criterio que considera Fishman y Lovas (1970) es el **objetivo** que pretende alcanzar la educación bilingüe, y permite distinguir tres programas diferentes:

a) *Programas de transición o compensación*, que son bastante parecidos a los que en el apartado anterior hemos calificado de bilingües de transición. Son programas que se utilizan para "compensar el déficit" de las minorías lingüísticas en la lengua oficial del Estado. El objetivo de este programa es permitir que el niño se escolarice en una lengua que le es familiar, lo cual le permitirá progresar más rápidamente en el aprendizaje de la L<sub>2</sub> mayoritaria, y así con posterioridad continuar la enseñanza en esa lengua.

b) *Programas de mantenimiento*. Son programas dirigidos a las minorías lingüísticas, pero a diferencia de los programas compensatorios buscan preservar la lengua y la cultura de la minoría, dándoles a la vez la posibilidad de participar en las actividades del grupo dominante.

c) *Programas de enriquecimiento lingüístico*. Su característica más notable es que son programas diseñados para el conjunto de la comunidad y no sólo para las minorías. Su propósito es ofrecer una experiencia

educativa adicional, unida al conocimiento de la lengua minoritaria de su sociedad.

La tercera variable que condiciona la educación bilingüe es el **status de las lenguas**. Según Fishman y Lovas (1970), la importancia de esta variable estriba en que puede ser una variable predictora del éxito de un programa bilingüe. En función de ella, se establecen cuatro dicotomías:

a) Lengua de importancia primaria versus lengua de importancia secundaria en la enseñanza bilingüe.

b) Lengua del hogar con uso coloquial versus lengua de la escuela, lengua vehicular del proceso de aprendizaje.

c) Lengua de prestigio internacional versus lengua de menor importancia en el mundo.

d) Lengua institucionalizada en ámbitos sociales fuera de la escuela versus lengua no institucionalizada.

En este sentido, si la lengua de importancia primaria es también la que se habla en casa y, además, es una lengua institucionalizada, como es el caso de los niños de lengua familiar catalana en los programas bilingües en Cataluña, se puede predecir el éxito del programa bilingüe con mayor seguridad que en el caso de que una lengua sea de importancia secundaria y no institucionalizada, aunque ésta sea una lengua internacional o de importancia en el mundo y, además, la hablada en casa, como sería el caso de los hispanos en algunos Estados de Norteamérica.



### **1.3. LA INVESTIGACION EMPIRICA EN LA RELACION BILINGÜISMO E INTELIGENCIA**

#### **1.3.1. Introducción**

A continuación se revisarán los conocimientos actuales sobre ciertos aspectos del desarrollo cognitivo de los sujetos bilingües, en comparación con el de los monolingües. Para conseguir este objetivo, dentro de este apartado se presentará cómo se ha abordado el tema de las relaciones entre Bilingüismo e Inteligencia, desde los primeros momentos de la investigación científica sobre bilingüismo hasta nuestros días. Esta revisión permitirá conocer sobre qué aspectos cognitivos la presencia del bilingüismo supone una ventaja o desventaja en comparación con los monolingües.

Uno de los aspectos más polémicos en la historia del bilingüismo y la educación bilingüe ha sido el tema de las *consecuencias o efectos* del bilingüismo. Esta cuestión adquiere una gran relevancia, sobre todo, cuando los resultados aportados acerca de los efectos producidos por el bilingüismo en los individuos son tan decisivos en aspectos tan importantes como la promoción de las lenguas minoritarias, la presencia de lenguas minoritarias en el sistema educativo, y los modelos y tipos de enseñanza bilingüe implantados en el ámbito escolar, por citar algunos ejemplos.

Es conveniente destacar que el tema del bilingüismo y la educación bilingüe desde sus comienzos siempre han estado estrechamente vinculados con el concepto de inteligencia. Esta situación se debe fundamentalmente a dos razones: (a) a motivos *históricos*, dado que los primeros trabajos científicos sobre bilingüismo coinciden con una etapa de estudio del desarrollo humano en la que predominaba el interés por la inteligencia y su forma de medirla, y (b) a motivos *sociales*, debido al papel tan destacado, y desmesurado a veces, que el rendimiento intelectual o académico adquiere en la sociedad. Por todo ello, no es de extrañar que las relaciones entre bilingüismo e inteligencia hayan sido objeto de estudio en numerosos estudios e investigaciones.

Teniendo en cuenta estas aportaciones, los estudios y trabajos presentados en este ámbito de estudio pueden clasificarse en tres etapas: un primer momento donde existe un predominio de las comparaciones cuantitativas, las cuales determinan una primera visión negativa del bilingüismo; una segunda etapa, centrada en el estudio de diferencias de organización o estructura de las capacidades cognitivas, y que dan paso a

las posibles ventajas cognitivas de los bilingües sobre los monolingües; y una tercera etapa cuyo interés se centra en las relaciones de ciertos procesos cognitivos y la conducta bilingüe.

### **1.3.2. Primera Etapa: Comparaciones cuantitativas**

Durante la primera etapa (desde finales del s.XIX hasta la década de los sesenta), la creencia dominante era que el bilingüismo ejercía un efecto perjudicial sobre el desarrollo cognitivo del individuo. Esta convicción queda patente en un consejo de larga tradición en el ámbito educativo:

"No eduque a su hijo en el bilingüismo o se creará problemas".

En esa época, los problemas asociados al bilingüismo podían ser desde una carga para el cerebro hasta la confusión mental, la inhibición en la adquisición de la lengua mayoritaria, e incluso la división de la personalidad. De esta manera, en aquellos momentos se aconsejaba a padres y profesores que usasen sólo una lengua con los niños. Incluso, en ocasiones cuando los alumnos persistían en hablar dos lenguas en la escuela, los educadores les ofrecían como remedio lavarles la boca con agua y jabón (Isaacs, 1976) o pegarles con una vara por hablar su propia lengua. Esta postura resulta bastante clara en una cita de un profesor de la Universidad de Cambridge de la época:

"Si fuera posible que un niño viviera en dos lenguas a la vez e igualmente bien, tanto peor. Su desarrollo intelectual y espiritual no

por ello se vería doblado, sino reducido a la mitad. La unidad de la mente y del carácter tendrían gran dificultad en afirmarse en tales circunstancias (Laurie, 1890, pp.15)".

Como se aprecia, la preocupación de que dos lenguas puedan tener un efecto negativo en las destrezas cognitivas del individuo tiende a expresarse de dos modos diferentes. Mientras algunos tienden a creer que cuanto más se aprenda y se use una segunda lengua, menos destreza se tendrá en la primera lengua, para otros está la preocupación de que la capacidad de hablar dos lenguas puede ser a costa de la eficiencia cognitiva, es decir, si residen dos lenguas en el pensamiento esto significa menos espacio para desarrollar otras áreas de aprendizaje.

Precisamente, este punto de vista que defiende que el bilingüismo tiene más desventajas que ventajas en términos de desarrollo cognitivo, se mantiene entre la sociedad británica y norteamericana hasta el siglo XX. Por tanto, las primeras investigaciones realizadas sobre bilingüismo e inteligencia surgen con el objetivo de demostrar las consecuencias negativas del bilingüismo.

Prueba de ello son los estudios realizados por Pintner y Keller (1922) o Haught (1931), entre otros, que aportan un "handicap lingüístico" para los sujetos bilingües, ó la investigación del galés Saer (1923) que describe el funcionamiento cognitivo de los bilingües como una especie de "confusión mental". Este último autor, tras encontrar una diferencia de 10 puntos en inteligencia entre monolingües y bilingües, procedentes de áreas rurales de Gales, concluyó que los monolingües eran superiores a los

bilingües, afirmando que los bilingües eran mentalmente confusos y, por tanto, estaban en desventaja en cociente intelectual, riqueza lingüística y momento cronológico de ampliación de vocabulario, es decir, mientras que en el grupo monolingüe ese último hecho se daba en torno a los 8-9 años, en los bilingües no aparecía hasta aproximadamente los 10-11 años. Además, un estudio adicional realizado por los mismos autores señaló que:

"la diferencia en la capacidad mental revelada por los tests de inteligencia es de naturaleza permanente, puesto que persiste en los estudiantes a través de su carrera universitaria" (Saer, Smith y Hughes, 1924, pp. 53).

En la misma época, los datos ofrecidos por Mead (1927) aportan resultados similares. Un poco más tarde, los estudios de Keston y Jimenez (1954) son los que ponen de manifiesto las dificultades lingüísticas de los emigrantes hispanohablantes en Estados Unidos, así como los problemas de fracaso escolar derivados de su escolarización en inglés. Éstos últimos se centran en la presencia de interferencias lingüísticas, pobreza de vocabulario y expresión, y retroceso al idioma familiar. Por su parte, Jones y Steward (1951) no aportan datos concluyentes en diversos estudios, mientras que en algunos trabajos los monolingües sólo superan a los bilingües en inteligencia verbal, en otros los bilingües siempre presentan puntuaciones más bajas en cualquier test de inteligencia.

En definitiva, podemos observar que los primeros resultados tienden a confirmar los efectos negativos del bilingüismo, al encontrar que los sujetos bilingües son inferiores a los monolingües en todas las pruebas de



inteligencia, tanto verbales como no verbales (Darcy, 1953; Pichon, 1947). Estos primeros trabajos suelen utilizar una **metodología comparativa tradicional**. Su interés son las "comparaciones cuantitativas" en los tests de inteligencia entre monolingües y bilingües, con el fin de averiguar quién puntuaba *más* alto, en un intento de decidir quién es *más inteligente*.

Sin embargo, un análisis de la metodología utilizada en estas primeras investigaciones permiten la detección de serios y graves problemas metodológicos. Por tanto, estos primeros estudios sobre bilingüismo e inteligencia no pueden utilizarse para defender las desventajas cognitivas del bilingüismo. A continuación, se destacan los problemas más relevantes:

1) *Problemas en la definición de inteligencia*. El concepto de la inteligencia y la utilización de tests para medirla ha sido un tema bastante controvertido y debatido, debido fundamentalmente a los problemas de definición y medida asociados a la inteligencia. Principalmente, la exposición de su definición ha estado ligada a tres controversias diferentes: la primera hace referencia a los efectos de la herencia y del ambiente en el desarrollo de la inteligencia; la segunda se posiciona en torno a una composición monofactorial o multifactorial de la inteligencia, y la tercera alude a la relatividad cultural de los tests de inteligencia sugiriendo que las generalizaciones interculturales o intergrupales pueden ser limitadas y en ocasiones arriesgadas.

2) *Inadecuación de los instrumentos de medida utilizados*. La principal causa de una utilización inadecuada de los tests en la medida de

la inteligencia reside primordialmente en el instrumento de medida administrado a los sujetos bilingües. Habitualmente, los tests de inteligencia más comunes en el mercado son instrumentos de medida desarrollados en la norma monolingüe, o sea, creados y baremados para individuos pertenecientes a la lengua y cultura mayoritaria o dominante. Así, cuando éstos son aplicados a un grupo de sujetos cuyas características lingüísticas son distintas a las de aquéllos para los que fue creado, se puede producir un serio problema de validez en esa aplicación. Por tanto, la opción preferible será contrastar la inteligencia de los bilingües en su lengua más fuerte o en ambas.

Sin embargo, las primeras investigaciones realizadas en Reino Unido, Estados Unidos y Canadá se realizaron aplicando únicamente tests de inteligencia en versión inglesa. De esta forma, si la inteligencia del grupo bilingüe era examinada en su lengua más débil, ello podía significar una gran desventaja para dichos sujetos. Esta comparación podía derivar en un bajo rendimiento o desarrollo cognitivo en los bilingües comparado con el de los monolingües, cuando en realidad éste no existía.

3) *Problemas de la definición y medida del bilingüismo.* La gran mayoría de los estudios realizados en esta primera etapa no son capaces de ofrecer ni una definición ni una medición operativa del fenómeno estudiado: el bilingüismo. Destacar aquí, que en la actualidad este problema todavía sigue vigente en las investigaciones presentes.

4) *Selección inadecuada de la muestra de sujetos bilingües.* Una clasificación acertada entre sujetos monolingües y bilingües está llena de

dificultades. La decisión correcta en determinar quién es o no bilingüe es un tema bastante complejo. Sin embargo, en las primeras investigaciones realizadas sobre bilingüismo se tendía a considerar la distinción entre monolingües y bilingües como algo no problemático. De este modo, en estos estudios podía darse el caso de que la muestra de bilingües utilizados, difiriera de la de los monolingües, además de la variable mono/bilingüismo, en otras características también relevantes para el objetivo de la investigación.

En aquella época, la mayoría de los sujetos bilingües utilizados bien eran sujetos procedentes de zonas rurales, que tenían un conocimiento mínimo o nulo de la lengua mayoritaria, o bien eran inmigrantes en el país de adopción, que además del problema de la lengua poseían las características propias de la emigración, como son la posibilidad de desarraigo, bajo nivel económico, etc. La situación actual es bien diferente. Hoy en día, la selección de la muestra bilingüe se aborda igualando a ambos grupos monolingüe y bilingüe, en aquellas variables sociodemográficas que puedan interferir significativamente en los resultados obtenidos en la variable objeto de estudio. Dependiendo de la realidad bilingüe, las variables a igualar en ambos grupos pueden ser el sexo, la edad, la lengua familiar, la clase socioeconómica, el modelo de educación bilingüe implantado, el ambiente urbano/rural de la zona, etc...

5) *Utilización del bilingüismo como variable añadida a posteriori.*

Una revisión de los estudios realizados en esta primera etapa permite revelar que éstos ni siquiera estuvieron diseñados para investigar las posibles diferencias del desarrollo cognitivo en sujetos monolingües y

bilingües. Frecuentemente, el tema de investigación era otro, pero con posterioridad se realizaba una división de la muestra en monolingües y bilingües, y se comparaban resultados. Por tanto, queda claro que las conclusiones obtenidas a partir de estos primeros trabajos son ambiguas y no poseen ningún valor concluyente.

6) *Empleo de promedios o estadísticos simples en las comparaciones realizadas.* Las primeras investigaciones realizadas en el campo del bilingüismo suelen aplicar una metodología muy elemental, basada fundamentalmente en el cálculo de *promedios o estadísticos simples*, al establecer comparaciones entre monolingües y bilingües. Este hecho queda claramente demostrado cuando Jones (1966) reanalizó la investigación realizada por Saer (1923), con una metodología diferente, y encontró que no habían tales diferencias significativas entre los grupos monolingüe y bilingüe.

7) *Problemas de Generalización.* La mayoría de los estudios realizados sobre bilingüismo e inteligencia poseen una característica común: el empleo de tamaños muestrales reducidos. Además, dichas muestras abarcan multitud de situaciones, como son escuelas urbanas/rurales, clases sociales distintas, edades diferentes, niveles de motivación variados, así como una amplia combinación de grupos bilingües distintos, por ejemplo grupos de indígenas de lengua minoritaria, personas inmigrantes, niños de lengua mayoritaria con enseñanza minoritaria, y niños con lengua familiar diferente a la académica. Evidentemente, esta manera de proceder en el estudio del bilingüismo ha provocado que no sea nada sorprendente la gran cantidad de diferencias halladas en el desarrollo cognitivo de los

bilingües respecto a los monolingües. Por otra parte, tampoco hay que olvidar que cada país e incluso diferentes comunidades de un mismo país poseen unos objetivos, unas necesidades y un contexto psicosocial muy diferente ante el mismo fenómeno: el bilingüismo. Por lo tanto, los resultados obtenidos en un grupo no pueden generalizarse a otros, ni los obtenidos en un país o comunidad pueden aplicarse a los sujetos bilingües del resto del mundo.

### **1.3.3. Segunda Etapa: Diferencias en la estructura de las capacidades cognitivas**

A medida que avanza la investigación en el campo del bilingüismo la marcada superioridad de los sujetos monolingües va desapareciendo, y las personas bilingües empiezan a igualarse a ellos en algunas áreas, sobre todo en pruebas no verbales de inteligencia.

La investigación de Pintner y Arsenian (1937), tras el empleo de una medida de bilingüismo y el control de algunas variables adicionales en ambos grupos, como son edad, sexo y status socioeconómico, provocó la aparición de una correlación no significativa entre el coeficiente de inteligencia verbal y la variable mono/bilingüismo yidish-inglés. También, otro estudio (Jones, 1959) ofreció los mismos resultados, aportando que la falta de control de la variable *status socioeconómico* era la principal causa de los resultados hallados en gran parte de la investigación previa, permitiendo afianzar la idea defendida por la mayoría de ellos: "la inferioridad de los bilingües".

Pero el momento clave en la relación entre bilingüismo e inteligencia, aparece con la publicación del trabajo de Elizabeth Peal y William Lambert (1962). Este estudio provocó un gran impacto en el ámbito del bilingüismo, por: (a) *su cuidadoso diseño y metodología*, superando muchas de las deficiencias metodológicas del periodo anterior. Este estudio mostraba un especial cuidado en la definición y medida del término bilingüismo, y en el control de aquellas variables que podían influir indebidamente en la medida de la inteligencia de monolingües y bilingües; y (b) *los resultados contrarios a sus propias hipótesis*. El grupo de bilingües analizado (edad:10 años, clase media y sistema escolar idéntico al resto) presentó cierta superioridad en algunas pruebas no verbales y en casi todas las de inteligencia verbal. Además, los comunicados anuales realizados demostraron que los niños bilingües después de cuatro o cinco años de escolarización bilingüe, podían hablar, leer, escribir y utilizar el inglés tan correctamente como aquellos que habían sido instruidos en inglés de manera convencional.

Sin embargo, algunos autores (Arnberg, 1987; Lambert y Anisfeld, 1969) muestran un cierto descontento con estos resultados, señalando que la existencia de algunos fallos y errores metodológicos son los causantes de la supuesta superioridad de los bilingües en las pruebas de inteligencia utilizadas. En concreto, estos autores manifiestan que, Peal y Lambert (1962) en un intento de evitar y corregir las limitaciones de las investigaciones anteriores, posiblemente utilizaran definiciones y medidas acerca del bilingüismo demasiado estrictas, lo cual provocó que la muestra de bilingües empleada estuviera compuesta principalmente por bilingües equilibrados en ambas lenguas. De esta forma, ello provocó que las

diferencias a favor de los bilingües en los tests de inteligencia tuvieran su origen en el hecho de que sólo los sujetos más inteligentes hubieran alcanzado el nivel de bilingüismo exigido por el estudio y, por tanto, las mayores puntuaciones en los tests de inteligencia no fueran consecuencia del hecho de ser bilingüe.

Durante esta segunda etapa, el objetivo principal es el estudio de la estructura o organización de los rasgos cognitivos en grupos de monolingües y bilingües. Por tanto, la orientación dominante es un **enfoque cualitativo** del bilingüismo, es decir, un estudio del bilingüismo en términos de amplitud de estilos, estrategias y destrezas cognitivas.

Dentro de este nuevo periodo, comienzan a producirse los primeros resultados contradictorios acerca de las consecuencias del bilingüismo en el desarrollo cognitivo de los bilingües. Algunos estudios (Macnamara, 1966; Skutnabb-Kangas y Toukomaa, 1976; Torrance y cols, 1970; Tushima y Hogan, 1975, entre otros) informan de la presencia de efectos negativos en el funcionamiento cognitivo de los bilingües debido al dominio simultáneo de dos o más lenguas. Precisamente, Tushima y Hogan (1975) encuentra para los sujetos bilingües, Inglés-Japonés, de 10-11 años puntuaciones más bajas en un test de habilidad verbal que en los monolingües. Los resultados encontrados son descritos defendiendo el siguiente argumento: "Dado que la persona bilingüe posee dos etiquetas o signos para cada concepto, ello provoca ciertas confusiones e interferencias en ambas lenguas y, por tanto, como resultado aparece un retraso en el desarrollo conceptual del bilingüe" (Macnamara, 1966).

Sin embargo, la línea predominante se localiza en numerosos estudios procedentes de países e idiomas diferentes, que continuamente revelan nuevos conocimientos acerca de las ventajas cognitivas encontradas de los sujetos bilingües (Balkan, 1970; Bain, 1975; Bain y Yu, 1978; Cummins y Gulutsan, 1974; Cummins, 1979a, 1979b; Genesee, Toker y Lambert, 1975; Ianco-Worrall, 1972; Liedtke y Nelson, 1968; Swain, 1975, 1978). Entre todos ellos, destacamos el trabajo de Ruiz (1974), quien tras realizar un estudio sobre bilingüismo y tests de inteligencia, obtuvo las siguientes conclusiones:

a) Respecto a algunos factores verbales relativos a la comprensión verbal, los bilingües presentaban puntuaciones más bajas que los monolingües. Este resultado fue interpretado indicando que este hecho no implicaba que los bilingües fueran menos inteligentes sino que éstos poseían una estructuración mental diferente.

b) En relación a la fluidez verbal, los datos no mostraron diferencias significativas entre bilingües y monolingües.

c) Asimismo, tampoco aparecieron diferencias en los factores de tipo no verbal, como son los espaciales y numéricos.

Por último, una tercera alternativa aparece con el estudio de Sandra Ben-Zeev (1977), quien halló en los bilingües tanto efectos cognitivos positivos como negativos en función del tipo de tarea a realizar. Concretamente, los sujetos bilingües español-inglés mostraban ciertas deficiencias en vocabulario y estructura gramatical frente a los monolingües ingleses pero, a su vez los bilingües presentaban habilidades superiores en tareas verbales que requerían ciertos análisis de complejidad estructural y perceptual, así como en tareas no verbales.



Respecto a las razones atribuidas para explicar la superioridad de los sujetos bilingües, cada investigación contribuyó con sus propias aportaciones. De todas ellas, destacamos las más significativas.

Peal y Lambert (1962) se plantean la posibilidad de que los distintos factores intelectuales puedan aparecer en edades diferentes según se trate de monolingües o bilingües, ya que las experiencias lingüísticas y culturales de ambos grupos son diferentes. La investigación empírica en este campo (Landar, Ervin y Horowitz, 1960; Ervin-Tripp, 1964; MacNamara, 1970; Segalowitz, 1980) aporta toda la información necesaria para que esta hipótesis quede totalmente confirmada: los análisis factoriales realizados muestran que, a igual CI en ambos grupos, la estructura semántica y cognitiva de los bilingües difiere de la de los monolingües en dos aspectos: (a) los bilingües presentan una estructuración mental mucho más diversificada, es decir, el mismo coeficiente intelectual en ambos grupos es explicado por mayor número de factores en los bilingües, y (b) existe un mayor predominio en los bilingües del factor denominado "flexibilidad cognitiva" (McLaughlin, 1984). Aquí, la *flexibilidad cognitiva* se refiere a la capacidad de enfocar de distintas formas un problema que pone en juego diferentes capacidades cognitivas hasta conseguir la mejor solución. Estas conclusiones parecen suficientemente generalizables, puesto que se confirman en estudios posteriores (Balkan, 1970; Ruiz, 1974; Sánchez-López, 1980, con muestra española; Groosjean y Miller, 1994) con niños de diversas edades, contextos culturales distintos, y países y lengua diferentes.

Por su parte, Lambert y Tucker (1972) informan que la experiencia de un programa de inmersión, como es el caso de Canadá, anima a los alumnos a una "lingüística de contraste incipiente", la capacidad de cambiar de una lengua otra rápidamente ayuda a la aparición de efectos positivos en el aprendizaje de la primera lengua -el inglés- y también a construir y comprender funciones lingüísticas complejas. Ventajas de esta práctica se han localizado en el área de uso gramatical, puntuación, vocabulario y habilidades metalingüísticas. Igualmente, las aportaciones de algunos autores (Balkan, 1979; Ben-Zeev, 1977, 1981; Ianco-Worrall, 1972) también ayudan a consolidar la postura favorable hacia el bilingüismo, al afirmar que los sujetos bilingües tienen más conciencia de la naturaleza arbitraria de las relaciones entre forma y significado que los monolingües.

#### **1.3.4. Tercera Etapa: Relaciones entre procesos cognitivos específicos y bilingüismo.**

En la década de los 70, con la dominación de la "Psicología cognitiva" se pone de manifiesto la complejidad que conlleva un estudio adecuado de la conducta bilingüe. Este nuevo enfoque aporta una concepción del bilingüismo como un **fenómeno multidimensional**. De tal forma que, el estudio de cada una de las dimensiones del bilingüismo desde perspectivas diferentes implicará el manejo y empleo de nuevas metodologías, y ello por supuesto repercutirá en una investigación con mayor rigor metodológico.

En esta tercera etapa, el interés se centra en el estudio de **cómo afecta el bilingüismo a los procesos cognitivos del individuo**, por ejemplo: ¿Difieren bilingües y monolingües en estilos de pensamiento?

¿Poseen diferencias en el procesamiento de la información? ¿Poseer dos lenguas crea diferencias al pensar sobre el lenguaje?. Este enfoque supone un nuevo intento de establecer una mayor fundamentación teórica en la relación bilingüismo e inteligencia.

A partir de aquí, se derivan una gran cantidad de trabajos cuyas líneas de investigación fueron el estudio de los siguientes procesos cognitivos: razonamiento abstracto, aptitudes aritméticas, aptitud espacial y aspectos relacionados con ella, estilos cognitivos, procesamiento de la información, etc.. Los principales resultados obtenidos pueden resumirse como sigue:

(1) Diferentes estudios realizados acerca de los programas de educación bilingüe, fundamentalmente el canadiense, ponen de manifiesto que, en el peor de los casos, los bilingües ofrecen un desarrollo cognitivo similar al de los monolingües cuando se les aplica baterías de inteligencia general (Lambert, Tucker y D'Anglejan, 1973; Duncan y De Avila, 1979; Hakuta y Diaz, 1984), al igual que otras ventajas cognitivas (Harris, 1992; Bochner, 1996). Por ejemplo, ciertos estudios basados en modelos explicativos avanzados enfatizan el papel que desempeña la "especial sensibilidad hacia el lenguaje" en los bilingües, es decir, los sujetos bilingües que manejan dos códigos lingüísticos diferentes poseen un desarrollo diferente en las habilidades cognitivas no lingüísticas (Diaz y Klinger, 1991; González, 1994) en comparación con los monolingües;

(2) Respecto al cálculo aritmético, los primeros estudios mostraban una clara desventaja de los sujetos bilingües, sobre todo si éste se realiza en la lengua académica; sin embargo, otros presentan resultados contradictorios al respecto (Cohen, 1974). Los datos más recientes señalan

que el grado de bilingüismo adquirido es una variable decisiva en la aparición de diferencias en las habilidades aritméticas de monolingües y bilingües (Mägiste, 1982; Chincotta y Rumjahn, 1995). Una mayor competencia lingüística en el individuo bilingüe va asociada a una clara tendencia de superioridad de los bilingües sobre los monolingües en tareas de habilidad aritmética.

(3) En relación a las medidas de aptitud espacial, los resultados obtenidos muestran que los bilingües alcanzan puntuaciones iguales o mejores que los monolingües (Cummins y Gulutsan, 1974);

(4) Otras investigaciones concluyen que los bilingües, frente a los monolingües, utilizan estrategias perceptivas distintas y más elaboradas en algunas tareas verbales, como por ejemplo mayor atención a la estructura global, y mayor facilidad para reorganizar el material (Sandra Ben-Zeev, 1975, 1977; Hakuta, 1987; Torrance, Gowan, Wu y Aliotti, 1970);

(5) Otros trabajos señalan que los bilingües presentan ventajas significativas en medidas de pensamiento divergente (Scott, 1973; Torrance y cols, 1970), creatividad (Atucha, 1978; Ortí Testillano, 1988), formación de conceptos (Bain, 1975; Liedtke y Nelson, 1968), independencia de campo (Diaz, 1986; Genesse y Hamayan, 1978) y conciencia metalingüística (Bialystok y Ryan, 1985; Bialystok, 1987; Galambos y Hakuta, 1988), siendo esta última la capacidad para localizar errores en una oración escrita en una de las dos lenguas del bilingüe. Sin embargo, estos últimos resultados vienen determinados por el nivel de bilingüismo adquirido en ambas lenguas. Cuánto más bilingüe es el sujeto, esto es, cuando ambas lenguas están relativamente bien desarrolladas mejor serán los resultados en rendimiento en el test o las ventajas cognitivas del bilingüe (Galambos y Hakuta, 1988).

Por contra, estas investigaciones todavía comparten algunas deficiencias metodológicas, siendo las más importantes: la no igualación en muestras monolingües y bilingües de aquellas variables que puedan interferir en los resultados sobre los efectos del bilingüismo y, por tanto, invalidar las interpretaciones procedentes acerca del mismo (Grosjean, 1982), la utilización de muestras muy pequeñas en algunos estudios, lo cual provoca que la realización de generalizaciones sea muy restrictiva o prácticamente inexistente y, por último, la falta de una definición operativa del bilingüismo adquirido por los sujetos participantes de la muestra. La mayoría de estudios que si han evaluado el grado de bilingüismo alcanzado en sus colaboradores han realizado sus análisis básicamente con un tipo de bilingües: los *bilingües equilibrados*.

En resumen, atendiendo a todas las aportaciones realizadas hasta el momento, se puede afirmar que *la presencia del bilingüismo no tiene por qué suponer ningún déficit cognitivo para el individuo*, siempre que exista un nivel de competencia lingüística adecuada en ambas lenguas. Pese a la no existencia de pruebas del todo concluyentes en este campo debido a la presencia de ciertos problemas y limitaciones de tipo metodológico la mayor parte de los estudios y trabajos revisados en los últimos tiempos parecen apoyar que, en algunos aspectos cognitivos específicos, *los sujetos bilingües son capaces de mostrar algunas características positivas y ciertas ventajas en el desarrollo intelectual respecto a sus homólogos monolingües*.

#### **1.4. MARCO TEORICO DEL BILINGÜISMO EN EL AMBITO EDUCATIVO**

La tradición psicológica y pedagógica sobre el bilingüismo se ha modificado notablemente en los últimos 25 años. Prácticamente hasta los años 70, éste era considerado un fenómeno negativo. La Conferencia del "Bureau International de l'education pour l'étude des problèmes du bilingüisme à l'age scolaire" celebrada en Luxemburgo en 1928 reúne a numerosos especialistas de diferentes países europeos en los cuales coexisten diversas lenguas (País de Gales, Luxemburgo, Cataluña, etc). Las conclusiones de esta reunión fueron unánimes en el sentido de que la mayoría de sus participantes consideraron el bilingüismo como un fenómeno perjudicial para el desarrollo intelectual y personal del alumno, lo cual repercutió notablemente en el enfoque científico adoptado para el estudio del bilingüismo (Etxeberria, 1992).

Durante las siguientes décadas, numerosas investigaciones en este ámbito afirman que las personas bilingües por norma general poseen puntuaciones más bajas en rendimiento o desarrollo cognitivo que los monolingües, señalando que la presencia de ciertos handicaps, como la confusión cognitiva y algunos conflictos emocionales eran los principales causantes de estos resultados. Evidentemente, esta concepción provocó que el bilingüismo no fuera bien considerado y se intentó por todos los medios su eliminación del ámbito educativo (Cummins, 1988).

Sin embargo, a principios de la segunda mitad de este siglo la UNESCO publica unas declaraciones afirmando que la lengua materna es el mejor instrumento para enseñar a un/a niño/a. Esta decisión provoca un cambio de tendencia en la forma de abordar el estudio del bilingüismo y, en consecuencia, se plantea la necesidad de una revisión del tema del estudio del bilingüismo y su metodología, con el fin de controlar posibles fallos o problemas metodológicos presentes en los estudios realizados hasta el momento. Sin embargo, en la actualidad algunas de las dificultades metodológicas encontradas en aquel momento todavía persisten, y posiblemente son el principal motivo de la actual fundamentación teórica del tema del bilingüismo. Por ejemplo, la ausencia de una definición y medida operativa sobre la competencia lingüística de los bilingües, junto con el hecho de que el bilingüismo es un proceso complejo que implica algo más que el uso simultáneo de dos lenguas, dado que un estudio adecuado incluye tanto al individuo como su contexto sociolingüístico más inmediato, hacen que la comparación entre los distintos trabajos realizados en este campo a veces resulte muy difícil y, por tanto, que la posibilidad de

generalización sea muy limitada. Además, hay que añadir la gran diversidad de tipos y situaciones bilingües estudiadas.

La preocupación por la falta de un marco teórico capaz de proporcionar orden y coherencia a la acumulación de conocimientos existentes en el área del bilingüismo ha sido manifestada por diversos autores (Cummins, 1979a, 1979b, 1983, 1986, 1988; Etxebarria, 1992; Vila, 1983; Siguán y Mackey, 1986; Hamers y Blanc, 1983; Tabouret-Keller, 1985; Cziko, 1992). En un intento de encontrar un marco teórico, Etxebarria (1992) afirma que a partir de los resultados y conclusiones encontradas en algunas de las investigaciones realizadas, se pueden extraer una serie de *principios generales* los cuales permitan elaborar los enunciados teóricos necesarios para aportar una base más sólida a los datos resultantes de las investigaciones realizadas sobre los programas de enseñanza bilingüe en comunidades y países bilingües. A continuación, siguiendo a Etxebarria (1992), se presentan los seis principios más consolidados y aceptados en el campo del bilingüismo.

#### **1.4.1. Principio o Hipótesis del umbral**

Una de las cuestiones que más interés ha suscitado de este campo ha sido la inclusión de la lengua materna de grupos lingüísticos minoritarios como lengua de enseñanza de ciertos contenidos curriculares en el sistema educativo. La preocupación sobre esta cuestión ha generado en los últimos años un gran número de estudios y trabajos, cuyo objetivo principal era



evaluar las habilidades cognitivas y el rendimiento académico en los grupos lingüísticos minoritarios escolarizados en programas de educación bilingüe (Cohen, 1975; Jones, 1960; Lambert y Tucker, 1972; Peal y Lambert, 1962). A pesar de que, en un principio, numerosas investigaciones aportaran resultados y conclusiones contradictorias, los resultados de estos primeros estudios tuvieron importantes efectos clarificadores dentro de la educación bilingüe. Uno de los más relevantes ha sido la posibilidad de demostrar que el aprendizaje de una segunda lengua depende, en buena parte, de la competencia que el niño tenga en su lengua materna o familiar (Cummins, 1983; Engle, 1975; Epstein, 1977; González, 1977).

En una de las investigaciones más importantes sobre esta cuestión llevadas a cabo con niños finlandeses en Suecia, se comprobó que los niños que habían llegado a Suecia a los 10 años eran los que obtenían mejores resultados en ambas lenguas.

"Sus habilidades en la lengua materna ya se habían desarrollado a nivel abstracto. Por este motivo, dominaban mejor y antes los conceptos en la lengua sueca que aquéllos que habían emigrado más pequeños o al inicio de la escolaridad, y sobrepasaban en algo incluso a los niños inmigrantes que habían nacido en Suecia (...). El desarrollo verbal de los niños que emigraron justo al empezar la escolaridad experimentó una seria perturbación (...). En este grupo, y en los que emigraron antes de empezar la escolaridad, el riesgo de llegar a ser semilingües era máximo" (Skutnabb-Kangas y Toukomaa, 1976, págs.74-76, citado por Cummins, 1983, págs.48-49).

Skutnabb-Kangas y Toukoma (1976) afirman que la condición necesaria para que la segunda lengua se desarrolle bien en el bilingüe es que la competencia lingüística en la lengua familiar sea la adecuada a la edad del sujeto. La aparición de estos resultados en este y otros países lleva a los investigadores de este campo a formular la *Hipótesis del umbral* (Cummins, 1981, 1984; Hatch y Hawkins, 1987; Skutnabb-Kangas y Toukoma, 1976).

Esta hipótesis afirma que el nivel de competencia en la segunda lengua es parcialmente una función de la competencia adquirida en la lengua materna en el momento de aprender la segunda lengua. Esto es, el sujeto debe adquirir un nivel mínimo de competencia lingüística en su lengua para que la introducción de la segunda lengua sea útil y eficaz. De esta manera, cuando se han desarrollado ciertas funciones del lenguaje en la primera lengua es probable que una fuerte exposición a la segunda lengua permita llegar a una buena competencia lingüística en esta última, sin detrimento de la lengua materna. Por lo tanto, un alto nivel de competencia en la lengua materna llevaría a un alto nivel de competencia en la segunda lengua. Si el sujeto no ha adquirido dicho nivel de desarrollo en la primera lengua por falta de apoyo, la segunda lengua se establecerá inadecuadamente de tal manera que, a su vez, obstaculizará el desarrollo de la primera lengua.

De esta manera, desde la formulación de esta hipótesis se plantea la existencia de diferentes niveles o umbrales de competencia lingüística en la primera y segunda lengua, tanto para evitar déficits de tipo cognitivo como para obtener beneficios cognitivos asociados al hecho de ser bilingüe.

Así pues, en el *primer nivel* referido al bilingüismo limitado, se encuentran aquellos sujetos bilingües con un bajo nivel de competencia en ambas lenguas, sobre todo cuando se los compara con el grupo monolingüe de su edad, lo cual deriva en efectos cognitivos negativos o perjudiciales para el sujeto. Mayoritariamente, este caso sería típico de aquellos sujetos bilingües escolarizados en programas de submersión o enseñanza de lengua mayoritaria. Los programas de submersión se caracterizan por ser programas vehiculados a través de la lengua y la cultura dominante y tienen como objetivo social y lingüístico la asimilación y la pérdida de la propia lengua y cultura. En el *nivel medio*, bilingüismo parcial, están los que tienen una competencia en una de sus lenguas apropiada a su edad, pero no en la otra. En este caso se encuentran los sujetos que reciben una enseñanza centrada en su primera lengua (sujetos parcialmente bilingües). Estos sujetos a lo largo de su escolarización elemental pueden mostrar algunas diferencias en algunos procesos cognitivos en comparación con los sujetos monolingües, pero es improbable que sean estadísticamente significativas (Cummins, 1983). El bilingüismo que exhibe un alto nivel de competencia lingüística sólo en una lengua y baja en la otra no comporta efectos cognitivos ni positivos ni negativos. Y en el *nivel superior*, se sitúan los niños que se aproximan a los bilingües equilibrados, los cuales tienen una competencia lingüística en dos o más lenguas apropiada a su edad, lo cual puede repercutir en poseer ciertas ventajas cognitivas sobre los monolingües. Una representación gráfica de esta hipótesis se presenta en la figura 1.

FIGURA 1. Representación gráfica de la Hipótesis del umbral.

TIPO DE BILINGÜISMO	EFFECTOS COGNITIVOS	
<hr/>		
A. BILINGÜISMO TOTAL:		
Alto nivel en las dos lenguas	Efectos cognitivos positivos	<i>Umbral elevado de competencia bilingüe</i>
<hr/>		
B. BILINGÜISMO PARCIAL:		
Nivel nativo en una de las dos lenguas	Efectos cognitivos ni positivos ni negativos	<i>Umbral bajo de competencia bilingüe</i>
<hr/>		
C. BILINGÜISMO LIMITADO:		
Bajo nivel en las dos lenguas	Efectos cognitivos negativos	
<hr/>		

La formulación de esta hipótesis tiene su base en afirmaciones como la siguiente:

"La base para poder llegar al nivel del umbral en la competencia en L<sub>2</sub>, parece ser el nivel al que se ha llegado en la lengua materna. Si en una fase temprana de su desarrollo un niño de la minoría se encuentra en un ambiente de aprendizaje en el cual se utiliza una lengua que no conoce, sin que al mismo tiempo se le proporcione un apoyo adecuado en su lengua materna, el desarrollo de sus habilidades en la lengua materna se frenará o incluso se parará, dejando al niño sin una base para aprender la segunda lengua lo suficientemente bien como para llegar al nivel del umbral" (Cummins, 1983).

Por lo tanto, diremos que la hipótesis planteada prevé dos umbrales de competencia lingüística: un umbral de competencia lingüística en L<sub>1</sub> o L<sub>2</sub> más bajo que la persona bilingüe debe alcanzar para evitar cualquier efecto cognitivo negativo, y un nivel más elevado, esta vez en L<sub>2</sub>, para conseguir los beneficios potenciales que el hecho de ser bilingüe tiene en su desarrollo cognitivo.

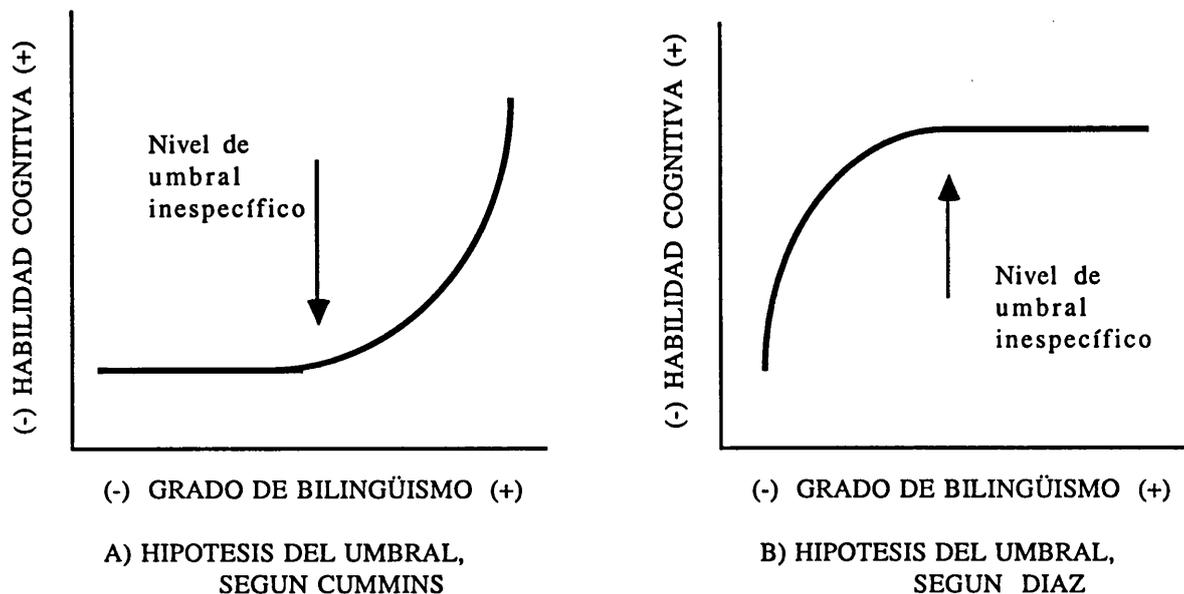
La hipótesis de los umbrales ayudó a establecer el por qué en Estados Unidos los niños con una lengua minoritaria a los que se les enseñaba en una segunda lengua (inglés), a veces no llegaban a desarrollar una competencia lingüística suficiente en dicha lengua. En estas situaciones, los alumnos de minorías lingüísticas con una lengua X se incorporan a programas de educación general cuya lengua de instrucción es la lengua mayoritaria Y, y, además, predominan los escolares con lengua materna Y. En la medida en que no todos los alumnos parten de la misma situación respecto al conocimiento de la lengua Y, lo más probable es que el esfuerzo por contextualizar el lenguaje e intentar incorporar al alumno a situaciones donde éste pueda entender el sentido y significado de la lengua académica, desaparezca ya que un gran número de alumnos no llegan a comprender adecuadamente aquello se les está enseñando en clase. De esta forma, el alumno de lengua minoritaria X no desarrolla una competencia lingüística adecuada ni en la segunda ni en la primera lengua lo cual desemboca en bajos resultados de rendimiento y desarrollo cognitivo. Por tanto, es importante tener presente que el ingreso de grupos de lengua minoritaria en la institución escolar no suponga un corte o cambio de lengua, ya que esta situación puede establecer ciertos retrasos en el desarrollo lingüístico y

cognitivo del alumno y, en consecuencia, limitar su capacidad para afrontar adecuadamente la educación obligatoria. En este sentido, los programas de Lengua Patrimonial implantados en los Estados Unidos han permitido que los niños de algunas minorías lingüísticas aprendan en su lengua más desarrollada, su lengua materna o familiar, lo cual ha llevado, como lo confirman datos empíricos, a un desarrollo satisfactorio de sus destrezas académicas en la lengua mayoritaria hacia el final de la escolarización (Gerard, 1983; Secada, 1991).

Sin embargo, la hipótesis del umbral presenta algunas debilidades metodológicas. Una de ellas reside en su formulación tan general (Cummins y Swain, 1986). Aunque, en un principio el carácter generalista de esta hipótesis hubiera podido resultar bastante útil en la aportación de un sentido y justificación para algunos resultados y conclusiones aparentemente contradictorias, cada vez más la propia realidad bilingüe sugiere la necesidad de una formulación más específica dentro de los distintos tipos y situaciones bilingües. Otra debilidad reside en la imposibilidad de definir los niveles o umbrales de competencia lingüística que un/a niño/a tiene que obtener para, en primer lugar, evitar los efectos negativos del bilingüismo y, en segundo, obtener las ventajas positivas del mismo (Vila, 1983), debido principalmente a que éstos varían en función del desarrollo del niño y de las exigencias curriculares que la escuela plantea en cada momento.

Dentro de esta hipótesis, quizás podamos ubicar las aportaciones de los estudios de Diaz y Hakuta (1981; Diaz, 1985).

FIGURA 2. Dos alternativas de la hipótesis del umbral.



Los resultados de sus estudios ofrecen apoyo empírico a la hipótesis del umbral pero, además, introduce una interpretación alternativa sobre cómo actúan los efectos del bilingüismo en las habilidades cognitivas de los sujetos. Como puede observarse en la figura 2, mientras que la hipótesis de Cummins (1981) supone que los beneficios del bilingüismo sólo se observan después de que el individuo haya conseguido un determinado nivel o umbral de aprovechamiento lingüístico en la segunda lengua ( $L_2$ ), la formulación de Diaz (1985) sugiere que "el grado de bilingüismo alcanzado por el sujeto permitirá predecir partes significativas de sus habilidades cognitivas, pero antes de conseguir un determinado nivel de aprovechamiento lingüístico de la  $L_2$ ".

En definitiva, independientemente de qué hipótesis del umbral se ajuste más a los hechos empíricos, queda claro que las aportaciones realizadas dentro de esta hipótesis señalan que un grado de bilingüismo equilibrado en ambas lenguas deriva en un efecto positivo tanto en desarrollo cognitivo como rendimiento académico del alumno (Cummins y Mulcahy, 1978; Cummins, 1983; Holmstrand, 1979; Husen y Opper, 1984).

#### **1.4.2. Principio de Interdependencia o Transferencia Lingüística**

El principio de la *interdependencia o transferencia lingüística* se encuentra muy relacionado con el anterior. Este principio trata de explicar el desarrollo de las habilidades académicas en relación con la competencia adquirida en la primera y en la segunda lengua, asumiendo que hay una interdependencia o transferencia de ambas lenguas en el niño. Por tanto, si la transferencia en ambas lenguas se da, los beneficios cognitivo-académicos del bilingüismo pueden lograrse mediante un desarrollo adecuado de la L<sub>1</sub> en sujetos bilingües. Su propio autor, Cummins (1986) afirma que:

"en la medida en que la enseñanza en la lengua minoritaria es efectiva para desarrollar la competencia académica en la lengua minoritaria, la transferencia de esta competencia a la lengua mayoritaria se dará dada una adecuada exposición y motivación para aprender la lengua" (Cummins, 1986, pp.20).



La formulación de este principio permite aportar una explicación al por qué los sujetos que reciben la educación elemental en su primera lengua, a pesar de estar menos expuestos a la enseñanza en la segunda lengua, desarrollan buenos niveles de competencia en la misma. En base a este principio, si los sujetos bilingües se incorporan al ámbito escolar con programas de enseñanza que desarrollan inicialmente la competencia lingüística en su lengua familiar y de manera gradual se introduce la segunda lengua, existirá una transferencia de los niveles alcanzados en la lengua familiar, en este caso la lengua académica, a la segunda lengua, consiguiendo buenos resultados cognitivo-académicos en ambas lenguas. Así pues, cuanto mayor sea la competencia lingüística en la primera lengua, mayores serán los efectos beneficiosos en la segunda lengua. Por el contrario, si los alumnos en el sistema escolar operan en una primera lengua insuficientemente desarrollada, la calidad y la cantidad de lo que aprendan sobre los complejos contenidos del plan de estudios puede ser relativamente débil y pobre. Este último caso suele ocurrir en contextos donde el prestigio o poder social de ambas lenguas es desigual, por ejemplo a causa de una actitud desfavorable hacia el aprendizaje en la primera o segunda lengua, o cierta presión para reemplazar la lengua del hogar por la lengua mayoritaria. Todo ello provoca que el funcionamiento cognitivo y el rendimiento académico de dichos sujetos pueda estar negativamente afectado.

Este principio ha sido puesto a prueba en numerosos contextos bilingües, y los datos aportados realmente confirman la existencia de una interdependencia entre el desarrollo de una lengua y los resultados obtenidos en la otra. Estudios en países como Estados Unidos, Canadá y diversos países europeos con niños de lengua minoritaria a los que se

permite utilizar su lengua materna durante parte o casi toda su educación elemental, demuestran que el mantenimiento de la lengua materna no sólo no produce ninguna pérdida en la adquisición de la lengua mayoritaria sino que incluso parece mejorarla. Por lo tanto, la aparición del fenómeno "interdependencia o transferencia lingüística" en la evaluación de programas bilingües permite afirmar que los niños que aprenden a leer y escribir en su lengua materna no sólo están desarrollando habilidades específicas en su propia lengua, sino que también están desarrollando destrezas cognitivas y lingüísticas de orden superior que les ayudarán en su futuro aprendizaje y desarrollo de la lengua mayoritaria, así como en su desarrollo intelectual general (Gerard, 1983; Schwarzwald y Cohen, 1982; Secada, 1991).

Sin embargo, a pesar de la aparición de resultados satisfactorios a favor de esta hipótesis, en ocasiones la evaluación de algunos programas bilingües aportan información sobre la presencia de altas tasas de fracaso escolar en alumnos bilingües escolarizados en programas de enseñanza bilingüe. Para aportar una posible explicación a esta situación, Cummins (1984) presentó una conceptualización doble de la competencia lingüística, distinguiendo entre competencia "*comunicativa interpersonal básica*" (BICS) y competencia "*cognitivo-lingüística*" (CALP). Realizada esta distinción, Cummins (1984a, 1984b) señala que el éxito de un programa bilingüe está determinado fundamentalmente por las capacidades cognitivo-lingüísticas de los niños, las cuales necesitan estar bien desarrolladas para afrontar adecuadamente el proceso curricular de la clase. Esta diferenciación es la que permite explicar por qué bastantes niños bilingües que aparentemente

demuestran un buen nivel conversacional en la segunda lengua ("competencia comunicativa interpersonal") no logran desarrollar una adecuada "competencia cognitivo-lingüística", capacidad relacionada con la lectoescritura, que deriva en un bajo rendimiento académico.

A pesar de la utilidad de este principio (Cummins, 1981, 1984, 1986), éste ha recibido numerosas críticas (Edelsky y cols, 1983; Martin-Jones y Romaine, 1986; Rivera, 1984; Frederickson y Cline, 1990, entre otros). A continuación, se resumen las más significativas:

1) La primera de ellas establece que la relación entre el desarrollo del lenguaje y el desarrollo cognitivo o aprovechamiento escolar del alumno no es una relación simple. Los expertos en este campo señalan que el estudio del bilingüismo y la enseñanza bilingüe, además de incluir el estudio de los factores personales del alumno, necesita considerar otro tipo de variables, como factores culturales, sociales y políticos de la propia comunidad, las expectativas del profesor y de los propios padres, etc., las cuales pueden ayudar a estudiar y comprender el bilingüismo como un fenómeno social e individual.

2) La segunda crítica se refiere a la distinción entre BICS y CALP. Aunque, esta clasificación posibilita una interpretación del por qué algunos niños bilingües que aparentemente son fluidos en su segunda lengua no son capaces de afrontar adecuadamente el plan de estudios en dicha lengua, esta distinción carece de apoyo empírico debido a que no se indica cómo obtener una definición y medida operativa en ambas competencias. De este modo, el empleo de ambos términos podría ser impreciso y derivar en

una simplificación de la propia realidad, siendo mal utilizados para clasificar y catalogar a los distintos alumnos bilingües.

3) La última crítica se centra en el hecho de que este principio surge como una explicación *post hoc* a partir de una gran variedad de hallazgos encontrados en la investigación sobre la educación bilingüe. En este sentido, los críticos sugieren que la elaboración adecuada de un marco teórico acerca del bilingüismo requiere investigación empírica de la realidad educativa, replicación y confirmación de resultados en comparaciones intracultural e intranacional, en el tiempo y en la tradición educativa.

#### **1.4.3. Principio del Bilingüismo Aditivo y Sustractivo**

Los estudios de Lambert y Tucker (1972) y Lambert (1974) ponen de manifiesto la imposibilidad de un único modelo de bilingüismo. A partir de los datos conseguidos en sus estudios, se plantea la distinción de diferentes tipos de bilingüismo, puesto que las consecuencias del bilingüismo pueden ocasionar resultados diferentes en el desarrollo cognitivo o académico del individuo. Un mismo programa de educación bilingüe puede tener, en algunos casos, efectos cognitivos positivos o aditivos sobre el alumno y, en otros casos, efectos negativos o sustractivos a causa de la influencia de algunos factores socioculturales propios de la situación bilingüe. En este sentido, a partir de los

resultados obtenidos tras la evaluación de programas de educación bilingüe, algunos investigadores se plantean la siguiente cuestión:

"¿Por qué el hecho de que sean diferentes la lengua familiar y la lengua de escolarización proporciona altos niveles de bilingüismo funcional y de logros académicos en los niños de clase media de lengua mayoritaria (Cohen, 1974; Davis, 1967; Lambert y Tucker, 1972; Swain, 1978); y sin embargo, conduce a un dominio inadecuado de ambas lenguas, lengua familiar y académica, junto con un resultado académico pobre en muchos niños de lenguas minoritarias?" (Cummins, 1983, pp.37).

La aportación de una explicación a los resultados contradictorios del párrafo anterior debe tener en consideración, para una situación específica, la relación de dominancia o prestigio existente entre la primera y segunda lengua, así como las dos culturas a las que cada lengua representa. Recordar que esta distinción se ha señalado anteriormente, en el apartado 1.1.3 del capítulo 1, cuando se intentaba establecer una clasificación de los diferentes tipos y situaciones bilingües. Teniendo en cuenta esta variable, Lambert (1974) diferencia entre bilingüismo aditivo y bilingüismo sustractivo.

1) El *bilingüismo aditivo* se da cuando el aprendizaje de una segunda lengua se produce sin perjuicio para el desarrollo de la lengua materna, como sucede en los programas de inmersión implantados en Canadá. En estos programas, todos los alumnos tienen un nivel de dificultad lingüística similar, puesto que todos hablan en su ámbito familiar una lengua diferente a la que se utiliza en la escuela, la lengua familiar es valorada y

conocida por la institución escolar, y la enseñanza de la segunda lengua se relaciona con el éxito y la autoeficacia del bilingüe en el dominio lingüístico de ambas lenguas, no planteando ningún conflicto con su cultura e identidad familiar. Así pues, en un contexto aditivo la primera lengua y cultura del sujeto no se reemplaza sino que se añade a la adquisición de la segunda lengua, lo cual generalmente va asociado a efectos lingüísticos y culturales positivos en dichos sujetos.

2) El bilingüismo *sustractivo* ocurre cuando el aprendizaje de una segunda lengua obstaculiza el desarrollo de la primera lengua, como sucede con frecuencia con las minorías lingüísticas al recibir una enseñanza en su mayoría, sino toda, en la lengua mayoritaria. Esta situación es característica de los denominados programas de *submersión* (Swain, 1978), donde los alumnos con una lengua minoritaria necesitan adaptarse en su proceso de enseñanza-aprendizaje a una segunda lengua, la lengua mayoritaria. Además, en este tipo de programas los profesores desconocen la lengua y cultura de los grupos minoritarios, lo que suele provocar que los mismos profesionales de la enseñanza lleguen incluso a infravalorar la competencia académica de dichos sujetos y, por tanto, a atribuir a la lengua familiar la principal causa de sus dificultades escolares. Estas situaciones propias de un bilingüismo *sustractivo*, se han relacionado generalmente con efectos cognitivos-académicos negativos de manera que, como señala Cummins (1984), es muy probable que el desarrollo lingüístico escolar de dichos niños sea inferior en cada una de sus lenguas que el de los hablantes nativos no bilingües.



En resumen, podemos apreciar cómo este principio combina los elementos individuales y sociales del bilingüismo. El bilingüismo aditivo y sustractivo son conceptos que están muy relacionados con las circunstancias socioculturales de la situación bilingüe y, por consiguiente, tienen mucho que ver con el prestigio que en la sociedad tiene la lengua que se quiere introducir en la institución escolar. De esta forma, si la incorporación de una lengua en el ámbito académico no es vista como una amenaza para el mantenimiento de la lengua y cultura propias, el individuo puede convertirse perfectamente en un bilingüe aditivo, pero si dicha lengua es vista como un peligro para la propia cultura y/o lengua seguramente tendrá lugar el llamado bilingüismo sustractivo. La importancia de estos factores han provocado que autores como Vila (1983), Cummins (1983) y Tucker (1977), se hayan pronunciado a favor de los factores socioculturales más que de los propiamente pedagógicos, como los más importantes para el éxito y progreso de la educación bilingüe. De este modo, la presencia de ciertas circunstancias socio-culturales en una determinada situación de contacto de lenguas como es una fuerte motivación, una actitud positiva y un suficiente prestigio de la lengua minoritaria que se imparte en la escuela, hacen que los efectos del bilingüismo puedan ser aditivos y, por tanto, enriquecedores para las dos lenguas del individuo bilingüe.

#### **1.4.4. Principio de la Significatividad del Aprendizaje.**

La poca eficacia de los métodos tradicionales en el terreno de la enseñanza de un idioma han sido ampliamente criticados por la mayoría de sus profesionales. Los expertos en el tema del bilingüismo educativo

insisten en que un aprendizaje sólido en el alumno suele producirse cuando éste tiene lugar en situaciones significativas para el alumno, es decir, cuando existe cierta implicación o actividad por parte del alumno y se establece una interacción o feedback entre profesor y alumno. En este sentido, numerosos autores defienden como principio teórico de la educación bilingüe el que el aprendizaje de una lengua se realice en un contexto significativo para el alumno (Cummins, 1983, 1987, 1993; Titone, 1976; Vives, 1979; Fillmore, 1985; Genesse, 1992; Sánchez y Rodríguez Tembleque, 1986; Etxeberria-Baxok, 1990; Domenech, 1990; Torró y Brotons, 1990; Artigal, 1991; Arnau, 1995; entre otros).

Numerosos trabajos centrados en el proceso enseñanza-aprendizaje de una segunda lengua proponen el siguiente principio: *"una lengua no se enseña, una lengua se aprende en un contexto comunicativo"*. Esta afirmación significa que todos los intentos de enseñar una lengua como un código, es decir, enseñar el vocabulario, las estructuras, las reglas morfológicas de una determinada lengua es un procedimiento muy poco efectivo. Estudios realizados en el campo de la educación bilingüe muestran que las situaciones pedagógicas más adecuadas para enseñar una lengua son aquellas donde la lengua que se aprende es utilizada de manera natural como medio de comunicación. Por tanto, a nivel de resultados, es más efectivo realizar una enseñanza en una lengua, que no enseñar una lengua (Arnau, 1995).

En definitiva, los resultados satisfactorios en el proceso de enseñanza-aprendizaje de una lengua procedentes de la enseñanza bilingüe destacan dos aspectos fundamentales: 1) la enorme importancia que tiene el



método de enseñanza utilizado, ya que la cualidad de la enseñanza y su grado de adaptación a las características del sujeto determinarán en buena parte los resultados obtenidos en rendimiento escolar, y 2) la necesidad de que el aprendizaje sea interactivo y significativo para el propio alumno.

#### **1.5.5. Principios de Complejidad, Contextualización y Continuidad**

Estos principios han sido elaborados a partir de las aportaciones teóricas de Hamers y Blanc (1983). En base a éstas, Etxebarria (1992) propone las "3-CON": *Complejidad, Contextualización y Continuidad*, que presentamos a continuación:

##### 1) COMPLEJIDAD.

En la actualidad, el estudio científico de muchos fenómenos sociales requiere cada vez más un acercamiento o perspectiva interdisciplinar. Esta nueva tendencia se manifiesta claramente en el estudio del lenguaje, el cual se presenta como un proceso activo, dinámico y en constante interacción con un amplio conjunto de factores de tipo individual, sociológico, histórico y cultural. Precisamente, el bilingüismo es un fenómeno complejo que abarca multitud de aspectos tanto individuales, como los referentes a la adquisición y aprendizaje del lenguaje, como colectivos o sociales, como es la propia situación de contacto de lenguas.

En este sentido, dado que la complejidad es una característica fundamental del bilingüismo, disciplinas como la "sociolingüística" o la

"psicolingüística" empiezan a ocupar un puesto relevante en su estudio. Desde esta pluralidad de disciplinas se insiste cada vez más en la necesidad de enmarcar los aspectos sociales, culturales y lingüísticos propios de cada situación bilingüe en el estudio del bilingüismo.

## 2) CONTEXTUALIZACION.

Por otra parte, los planteamientos teóricos de Hamers y Blanc (1983) señalan que *"cada situación de contacto de lenguas presenta una serie de características propias y específicas, siendo éstas y la forma concreta que una comunidad tiene de afrontar el contacto entre lenguas lo que realmente define a una comunidad y su bilingüismo"*. Así pues, otro aspecto relevante para abordar adecuadamente el estudio del bilingüismo es una correcta contextualización del mismo, de tal forma que el contexto sociolingüístico de una situación bilingüe determinada es lo que realmente define el tipo de relaciones que existen entre las lenguas de una determinada comunidad o país bilingüe.

Un claro ejemplo es la situación de contacto de lenguas entre el castellano y el catalán en el Estado Español, situación bilingüe organizada y afrontada a nivel legal, organizativo y de relaciones sociales de forma muy diferente en cada una de las tres Comunidades Autónomas (Cataluña, Comunidad Valenciana e Islas Baleares) donde está presente esta lengua. En este sentido, cada comunidad ha contextualizado de manera muy diferente su propia realidad bilingüe, de modo que los resultados y conclusiones

adoptadas en el estudio del bilingüismo han de interpretarse y adecuarse a la realidad de cada situación bilingüe.

### 3) CONTINUIDAD.

Dentro de este principio, Hamers y Blanc (1983) manifiestan su preocupación por la falta de un cuerpo teórico en el campo del bilingüismo. En este sentido, con el propósito de elaborar una fundamentación teórica sobre este fenómeno estos autores afirman que existe una necesidad de buscar aquellos elementos comunes al fenómeno bilingüe en todas y cada una de las situaciones bilingües estudiadas y, por tanto, evitar los exclusivismos y las contradicciones del mismo. Dicho de otra manera, la complejidad de la realidad bilingüe debe abordarse situando las diferentes conclusiones señaladas en cada situación lingüística a lo largo de un continuo y no recurriendo a simples explicaciones dicotómicas y categóricas. De este modo, este tipo de actuaciones centradas en la búsqueda de características y cualidades universales del bilingüismo permitirán dar un orden y coherencia a la cantidad de conocimientos existentes y, a veces, contradictorios acerca del bilingüismo y la educación bilingüe.

#### **1.4.6.Principio del Valor Simbólico de la Lengua.**

Atendiendo a todo lo expuesto hasta el momento, resulta bastante evidente que la lengua no es una entidad independiente de la sociedad sino todo lo contrario. En numerosas ocasiones, la lengua de una comunidad o

país comporta el símbolo del poder social, político y/o económico de dicha comunidad o país.

Sin embargo, muchas veces el estudio acerca de los efectos cognitivos del bilingüismo han sido analizados como si éstos estuvieran totalmente desligados de la estructura o contexto social de la propia situación bilingüe, cuando en realidad las circunstancias sociolingüísticas que envuelven a la situación bilingüe determinan en gran medida los resultados encontrados en los programas educativos bilingües implantados en dichos territorios. Las lenguas o idiomas son aprendidos por individuos que viven en un contexto sociolingüístico determinado, donde la propia sociedad les atribuyen un valor y una utilidad simbólica. Ello determinará las actitudes y las motivaciones que dichas personas tengan para aprenderla y, en cierta medida, determinará los resultados de los programas de enseñanza bilingüe.

En esta línea, la importancia del contexto social, político y/o cultural de una determinada situación lingüística para el estudio del bilingüismo educativo es de tal relevancia, que algunos autores han llegado incluso a afirmar que éstos pueden ser tan decisivos o más que las aptitudes lingüísticas de cada uno de los alumnos para determinar el éxito en el aprendizaje de una lengua (Gadner, Smythe, Clement, Glikzman, 1976; Fernandez Mauro, 1981; Arnau, 1985; Sanchez y Tembleque, 1986).

A modo de ejemplo, presentamos la situación de enseñanza bilingüe castellano-catalán en dos Comunidades Autónomas: Cataluña y Comunidad Valenciana. En Cataluña, el prestigio y el poder social de la lengua

minoritaria es fuerte, lo que se traduce en una gran presencia de la lengua minoritaria en todos los niveles educativos de los programas bilingües implantados en dicha comunidad y, por consiguiente, una elevada motivación y actitud de los estudiantes hacia el estudio del y en catalán. En cambio, en la Comunidad Valenciana esta lengua posee menor prestigio y, además, las clases sociales elevadas y los representantes políticos la utilizan mucho menos. Esta situación se traduce en un valor instrumental de la lengua valenciana mucho más inferior que la del castellano, lo cual queda patente en la propia realidad bilingüe, donde la presencia de la lengua minoritaria en el ámbito escolar es mucho menor y las motivaciones de los estudiantes valencianos para aprenderla son más bajas. La comparación de estas dos situaciones bilingües, donde una misma lengua minoritaria posee valores simbólicos diferentes, sugiere que los resultados sobre los estudios de bilingüismo en la Comunidad Valenciana serán totalmente diferentes que los obtenidos en Cataluña. Además, esta situación se agrava todavía más cuando dentro de una misma Comunidad, como es el caso de la Comunidad Valenciana se distinguen claramente situaciones sociolingüísticas totalmente diferenciadas. Por este motivo, en un capítulo posterior se abordará esta problemática presentando una contextualización sociolingüística de la Comunidad Valenciana.

## **1.5. ESTUDIOS EMPIRICOS: EXPERIENCIAS EN COMUNIDADES Y EN PAISES**

### **1.5.1. Introducción**

Dentro de este apartado, se expondrán las principales conclusiones aportadas tras la evaluación de varios programas de enseñanza bilingüe implantados en países y comunidades diferentes. En primer lugar, se abordarán los programas de inmersión en Canadá para alumnos de mayorías lingüísticas y, en segundo término, se ofrecerán los resultados más significativos tras la evaluación de diversos programas de educación bilingüe implantados en tres Comunidades Autónomas del Estado Español: Cataluña, Galicia y País Vasco.

### 1.5.2. Programas de inmersión en Canadá

Durante muchos años, y quizás todavía hoy, Canadá es el país de referencia básico para los estudios sobre el bilingüismo y la educación bilingüe. Con más de 1000 estudios de investigación en sus programas de inmersión, Canadá ha sido capaz de demostrar que la educación bilingüe es un experimento educativo de gran éxito y desarrollo, principalmente debido a su eficacia para producir resultados satisfactorios en el aprendizaje bilingüe (Genesee, Lambert y Holobow, 1986). Estos resultados han influido de forma decisiva en la enseñanza bilingüe que en la actualidad se imparte en Europa y otros países del mundo.

Desde que por primera vez, en el año 1965 se introducen en Canadá los programas de enseñanza de una segunda lengua mediante la inmersión lingüística, la evaluación de la competencia lingüística adquirida en ambas lenguas, el rendimiento académico y las actitudes de los alumnos escolarizados en programas de naturaleza bilingüe han sido objeto de estudio de numerosas evaluaciones durante más de 20 años. En concreto, los programas de inmersión canadienses son programas de enseñanza de una segunda lengua (L<sub>2</sub>:lengua francesa) dirigidos a alumnos de lengua y cultura mayoritaria (L<sub>1</sub>:lengua inglesa).

Desde las posiciones tradicionales de la Psicología y la Pedagogía se insistía, cada vez más, en la necesidad de aportar apoyo empírico a las siguientes preguntas: Si la enseñanza por inmersión suministra la ruta a la fluidez en la segunda lengua (francés), *¿es a costa de lograrla en la primera lengua? ¿Resulta el bilingüismo en un menor rendimiento en la primera lengua comparado con los alumnos de los programas generales?*.

Respecto a otras áreas del plan de estudios, ¿el bilingüismo se logra a costa del rendimiento en otras asignaturas?.

La evaluación de estos programas se realizó mediante la utilización de *diseños longitudinales clásicos*. El diseño de evaluación empleado consistió en comparar el rendimiento del grupo que seguía un programa bilingüe, con el grupo control que seguía la enseñanza regular en inglés y que tenía el francés como asignatura, y con otro grupo cuya lengua familiar y lengua de enseñanza eran sólo el francés. Además, en estos estudios se controlaron en los tres grupos las variables sexo, edad, cociente de inteligencia y posición socioprofesional de la familia. Además, también se incluyó como variable independiente el medio sociocultural en que vivían los niños, teniendo en cuenta sobre todo las actitudes hacia el aprendizaje de la segunda lengua y las actividades culturales que realizaba la familia de los sujetos participantes en el estudio. Respecto a los instrumentos de medida utilizados, éstos fueron tests estandarizados ya existentes en versión inglesa y se desarrollaron otros para evaluar el rendimiento escolar en francés; en concreto, se midieron las siguientes habilidades: conocimiento del vocabulario, gramática, ortografía y puntuación.

En cuanto a la evaluación de los programas de inmersión en Canadá, acerca de los resultados en el *rendimiento académico en la L<sub>1</sub>* (lengua inglesa) y *la L<sub>2</sub>* (lengua francesa) de los alumnos escolarizados en dichos programas, se obtuvieron las siguientes conclusiones (Lambert y Tucker, 1972; Genesee, Tuler y Lambert, 1975; Genesee, 1986; 1987; Genesee, Lambert y Holobow, 1986):



Por lo que respecta al *rendimiento académico en lengua inglesa*, en los 4 primeros años de los programas de **inmersión total temprana** (inicio en los primeros niveles de escolarización -Jardín de Infancia- y 100% de instrucción en la segunda lengua (L2:lengua francesa), no se apreció ningún retraso en comprensión y expresión oral en la primera lengua (L1:lengua inglesa), pero sí se experimentó un retraso inicial en las habilidades de lectoescritura en inglés comparadas con las de los alumnos monolingües escolarizados en programas de enseñanza general en inglés. Por ejemplo, las habilidades en la lectura, la ortografía y la puntuación no estaban tan desarrolladas en los bilingües a causa de que estos niños no recibían instrucción en lengua inglesa durante uno, dos o tres años después de empezar en la escuela. Sin embargo, este patrón inicial no persistía. Después de seis años de escolarización aproximadamente, los niños de inmersión total temprana habían alcanzado a sus compañeros monolingües ingleses en las destrezas correspondientes a la lengua inglesa (Swain y Lapkin, 1982, 1991) e, incluso les sobrepasan en los primeros años de secundaria. Así lo demuestra el siguiente comentario:

"Los niños que siguen un programa de inmersión en Canadá presentan normalmente una demora en el rendimiento académico, cuando se les enseña el plan de estudios en la segunda lengua (en este caso, en francés), es decir, tienden a ir detrás de sus compañeros monolingües durante un breve periodo de tiempo. Pero, una vez dichos sujetos adquieren competencia suficiente en la segunda lengua, es decir, que el francés se desarrolla adecuadamente para afrontar las tareas conceptuales de la clase, éstos generalmente alcanzan a sus compañeros monolingües. Por tanto, es improbable que este tipo de

enseñanza tenga consecuencias perjudiciales para el rendimiento académico o desarrollo cognitivo de estos niños (Cummins, 1987)".

Este hallazgo se enlaza con las posibles ventajas cognitivas que se derivan del hecho de ser bilingüe, y quizás tenga su explicación en la hipótesis sobre la interdependencia lingüística formulada por Cummins (1986), la cual establece una serie de relaciones interdependientes entre la competencia lingüística de las dos lenguas, de tal forma que la competencia lingüística alcanzada en una lengua se puede transferir a la otra, consiguiendo un grado de bilingüismo aceptable. Sin embargo, aunque todavía se conocen poco los procesos psicolingüísticos implicados en la "transferencia", se sabe que este proceso no es el mismo en todos los programas y que depende de las características de las lenguas de enseñanza: de tal modo que entre inglés y francés es mayor que entre inglés y hebreo o inglés y persa.

Por su parte, los alumnos de **inmersión parcial temprana** (inicio en los primeros niveles de escolarización y 50% de instrucción en la L2:lengua francesa) también tendieron a rezagarse durante tres o cuatro años en las destrezas lectoescritoras de su primera lengua: lengua inglesa. Su rendimiento era poco diferente del de los alumnos de **inmersión total temprana**, lo que fue bastante sorprendente ya que la enseñanza por **inmersión parcial** tenía más contenido en la lengua inglesa. Al finalizar la escuela elemental, los niños de **inmersión parcial temprana** alcanzaron en habilidades de lengua inglesa a sus compañeros escolarizados en programas convencionales de enseñanza inglesa. A diferencia de los alumnos de **inmersión total temprana**, los de **inmersión parcial** no tendían a sobrepasar a los grupos de los programas regulares en inglés.

Por último, los programas de **inmersión tardía**, donde el 100% de la instrucción se realiza en la L<sub>1</sub> (lengua inglesa) y la introducción de la segunda lengua (L<sub>2</sub>:lengua francesa) tiene lugar en los últimos niveles de la escolarización elemental. En ocasiones, se suelen incorporar con posterioridad cursos de enseñanza de la L<sub>2</sub> e incluso enseñanza en L<sub>2</sub> de alguna materia en secundaria. Los resultados en estos programas mostraron que éstos no tenían ningún efecto perjudicial en las destrezas de lengua inglesa adquiridas (Genesee, 1983, 1986).

Por lo que se refiere a la evaluación del *rendimiento académico en la L<sub>2</sub>* (lengua francesa), los alumnos de cualquier programa de inmersión comparados con los monolingües francófonos obtuvieron resultados similares en tests de comprensión lectora, pero presentaron puntuaciones más bajas en tests de expresión oral y escrita en francés. Así pues, aunque los alumnos de inmersión tienen una buena competencia funcional en la L<sub>2</sub>, cometen más errores al expresarse. Todo significa que aprender una segunda lengua igual que un nativo lleva siempre mucho tiempo y la escuela siempre es un modelo limitado de uso lingüístico.

Por lo tanto, resumiendo los datos aportados por la investigación canadiense sugieren que los niños escolarizados en programas de inmersión aprenden francés sin costo adicional para su competencia en inglés. Además, no sólo hay una ganancia de una segunda lengua, el francés, sino que también hay ciertas pruebas que sugieren que la inmersión puede resultar en posibles beneficios extras para la competencia lingüística adquirida en inglés.

En relación a *otras áreas del plan de estudios*, ¿cómo avanzan los niños de programas de inmersión en áreas como ciencias naturales, matemáticas, historia y geografía, comparados con los niños que cursan la enseñanza general? Los resultados aportados muestran que los alumnos de **inmersión total temprana** rinden en general en tales asignaturas tan bien como lo hacen los niños de programas generales. Esto es, el rendimiento escolar no se ve adversamente afectado por la enseñanza bilingüe de inmersión total. Por su parte, las evaluaciones de los alumnos de enseñanza por **inmersión parcial temprana e inmersión tardía** señalan que cuando éstos aprenden matemáticas y ciencias en francés, su rendimiento empeora al principio, pero posteriormente éste tiende a igualarse al obtenido por los de la escuela de enseñanza convencional tanto en inglés como francés. Este resultado puede deberse a que sus destrezas en francés no están lo suficientemente desarrolladas para aprender correctamente contenidos curriculares en ese idioma. Los resultados globales sugieren que la mayor parte de los alumnos ganan una segunda lengua sin costos para su rendimiento en todas las asignaturas del plan de estudios. Ahora bien, la clave para el éxito parece que reside en que las destrezas lingüísticas deben estar bien desarrolladas para trabajar adecuadamente el plan de estudios con una segunda lengua y, en consecuencia, obtener resultados satisfactorios en el rendimiento escolar. Este último resultado podría tener su interpretación en la hipótesis del umbral formulada por Cummins (1981, 1984).

En definitiva, las dos conclusiones más importantes que se derivan de las numerosas evaluaciones realizadas en los programas bilingües canadienses (Genesee, 1976, 1987) son que:

a) los alumnos escolarizados en programas de inmersión obtienen niveles elevados de competencia lingüística en la segunda lengua sin ser perjudicada la competencia en la propia lengua, y

b) los resultados en rendimiento académico, no sólo no se ven afectados por el programa de enseñanza bilingüe sino que, en algunos casos como en los programas de inmersión total temprana, los resultados fueron más positivos que los obtenidos por sus homólogos monolingües ingleses.

Antes de finalizar este apartado, es conveniente recalcar que la presencia de resultados tan satisfactorios en el ámbito del bilingüismo educativo puede provocar un peligro de generalización de la experiencia canadiense a otros programas bilingües implantados en el resto del mundo. Es importante destacar que el bilingüismo canadiense tiene una serie de características peculiares que hacen muy difícil, sino imposible que los resultados de estos programas sean comparables a otras situaciones o países donde el bilingüismo también está presente. Pero si todavía pensamos que la enseñanza por inmersión en Canadá merece la pena generalizarla a otros países, hay ciertas condiciones de la realidad bilingüe canadiense que deben ser consideradas: 1) La enseñanza por inmersión en Canadá afecta a dos lenguas de estatus internacional, el francés y el inglés; situación poco frecuente en el resto de países; 2) La enseñanza bilingüe es opcional y no obligatoria. Por tanto, este modelo de enseñanza bilingüe funciona cuando hay convicción y no una conformidad forzada por parte de profesores y padres de niños, lo cual afecta al sentido ético de la escuela, la motivación y el rendimiento de los niños; 3) La enseñanza de inmersión canadiense empieza con niños que están en un nivel similar en sus destrezas lingüísticas, lo cual puede repercutir en una enseñanza más eficiente e

igualdad de oportunidades para todos; 4) Los programas de inmersión lingüística en Canadá aseguran que hay un respeto para la lengua y cultura maternas del niño, lo que deriva a una situación bilingüe aditiva, donde profesores junto con padres de alumnos, funcionarios e investigadores son colaboradores en la implantación de dichos programas de inmersión; 5) La implicación de los profesores tiene importantes repercusiones en el éxito alcanzado en la inmersión canadiense. En un estudio, Roberts (1985) señala que el compromiso del profesor con la enseñanza bilingüe lleva a un buen rendimiento del sujeto en la escuela canadiense; y 6) por último, un aspecto fundamental de la situación bilingüe canadiense es que detrás de los programas de inmersión existe toda una ideología política, social y cultural que defiende con fuerza estas formas de enseñanza bilingüe.

#### **1.5.2. Programas de Educación Bilingüe en el Estado Español**

En los últimos años, la situación lingüística en el sistema educativo del Estado Español ha experimentado un gran cambio. La implantación de un "Estado de Autonomías" ha supuesto el traspaso de distintas competencias a las diferentes Comunidades Autónomas, lo cual ha permitido la existencia de Comunidades Autónomas con lengua propia y competencias exclusivas en la enseñanza y, en consecuencia, la aparición de un número importante de programas de educación bilingüe en dichos territorios.

A partir del año 1978, cuando se aprueba el "Decreto de Bilingüismo", comienzan a implantarse las Leyes de Normalización Lingüística en las Comunidades de Cataluña, País Vasco, Islas Baleares, Comunidad Valenciana y Galicia. El objetivo común a todas ellas era extender el conocimiento, a

nivel oral y escrito, de la lengua mayoritaria oficial en todo el territorio español y la lengua minoritaria propia de cada Comunidad Autónoma. La puesta en marcha de esta nueva política educativa deriva en la obligación de impartir parte del curriculum escolar en las distintas lenguas autóctonas propias de cada comunidad, sin que el conocimiento y utilización de una lengua distinta al castellano perjudique los progresos de la competencia de los individuos en la lengua castellana.

En la actualidad, dentro del territorio Español se hablan además del castellano, la lengua catalana en Cataluña y Andorra, y sus variantes, el valenciano en la Comunidad Valenciana, y el mallorquín, ibicenco y menorquín en las Islas Baleares (siendo la segunda lengua del Estado en número de hablantes). El gallego en Galicia, y el vascuence en el País Vasco y algunas zonas de Navarra. En el Valle de Arán se habla y se enseña en aranés junto al catalán y el castellano; y, en Asturias se intenta revitalizar el bable mediante la aplicación de programas experimentales de enseñanza en bable a alumnos de preescolar. Por último, la etnia gitana utiliza el caló, pero no existe información de programas de enseñanza impartidos en esta lengua. En la gran mayoría de estas regiones han surgido programas de enseñanza bilingüe.

Sin embargo, en el Estado Español no existe una tradición evaluadora de los resultados del sistema educativo y, por tanto, es muy difícil hacerse una idea global de los resultados de la educación escolar en este país. No obstante, recientemente algunos programas de educación bilingüe han sido objeto de evaluación, lo cual nos puede ayudar a analizar las ventajas e inconvenientes del sistema educativo actual. Teniendo en cuenta estas limitaciones, en los apartados siguientes se ofrece un resumen de los

estudios realizados sobre los modelos de educación bilingüe implantados en Cataluña, Galicia y País Vasco.

#### **1.5.3.1. Evaluación de programas en Cataluña.**

De acuerdo con la Ley de Normalización Lingüística, aprobada por el Parlamento catalán en abril de 1983, *"los alumnos de Cataluña al finalizar los estudios de Educación General deben poseer un dominio equilibrado de las dos lenguas oficiales, el castellano y el catalán, permitiendo utilizar indistintivamente una u otra lengua en cualquier tipo de situación comunicativa"*. Durante los años transcurridos desde la aprobación de esta ley, los trabajos de evaluación realizados sobre los diferentes modelos de enseñanza bilingüe en Cataluña se han centrado en dos objetivos distintos: 1) Evaluar cómo las distintas tipologías lingüísticas escolares inciden en el rendimiento académico global de los escolares, y 2) Conocer el valor de los distintos modelos lingüísticos en la adquisición del dominio lingüístico en la lengua castellana y en la catalana a lo largo de todo el proceso de escolarización.

Una gran parte de estos estudios ha tenido como objeto de estudio los programas de inmersión lingüística (PIL) dadas las dificultades psicopedagógicas que se plantean en su diseño (Generalitat de Catalunya, 1986, 1989; Arenas, 1986; Serra, 1992). El PIL es un programa educativo que plantea el aprendizaje precoz de la lengua minoritaria en la escuela y que tiene por objetivo que los alumnos escolarizados en ellos, en su mayoría alumnos de ambientes no-catalanoparlantes, dominen de forma semejante, al final de la enseñanza obligatoria, la lengua mayoritaria oficial en todo el territorio y la lengua minoritaria propia de cada comunidad.



Respecto al *primer objetivo*, la evaluación del rendimiento académico global del alumno se ha centrado, principalmente, en el estudio de las habilidades de lectura y escritura adquiridas al final del ciclo inicial en los programas de inmersión ya que, en definitiva, el éxito implicado en el dominio de dichas habilidades está en relación con el posterior éxito o fracaso escolar.

Arenas (1986) cita un trabajo realizado por el SEDEC (Servei d'Ensenyament del Català) en colaboración con el Departament de Psicologia de la Universitat de Barcelona, donde se comparan los rendimientos escolares globales en alumnos de Preescolar (4 años) que siguen o no un programa de inmersión (habilidades lectoescritoras en catalán). En este estudio, se utilizaron tres grupos: dos programas de inmersión y uno en castellano, controlando que los 3 grupos tuvieran el mismo nivel sociocultural. La prueba utilizada fue el *McCarthy*, un test que permite medir el nivel verbal, perceptivo-motor, memorístico, numérico y cognitivo general de los sujetos. Los resultados mostraron que, al menos al inicio de la escolaridad, no existían diferencias significativas entre los tres grupos, es decir, que la inmersión lingüística no perjudicaba el rendimiento o desarrollo global de los alumnos. Otro estudio realizado por el SEDEC también manifestó que los alumnos de 3º curso de Primaria escolarizados en programas de inmersión mostraban los mismos resultados en rendimiento académico que sus respectivos escolarizados en castellano.

Los próximos estudios a considerar son los realizados por Forns-Santacana y Gómez-Benito (1990, 1993, 1994, 1995). Dichos estudios también se realizaron en colaboración con el SEDEC. Su principal objetivo fue

evaluar la incidencia de los programas de inmersión lingüística en el desarrollo cognitivo, lingüístico, de adaptación y rendimiento académico en alumnos de Pre-escolar (4-6 años). Para ello, se comparan las variables indicadas en alumnos de Preescolar que siguen o bien programas de inmersión o bien programas de enseñanza monolingüe en castellano. Los instrumentos de medida utilizados para evaluar el desarrollo cognitivo fueron el Test de Columbia (Burgemeister, Blum y Lorge, 1979), el Test de McCarthy (McCarthy, 1972) en ambas versiones y el Test de GEFT (Witkin, 1981); para el desarrollo verbal y las habilidades comunicativas se aplicaron el Test Bankson (Bankson, 1977) en versión castellana y catalana (Forns y Triadó, 1987), la lámina de Boada (Boada y Forns, 1989) y una técnica comunicativa (Boada y Forns, 1989, 1994; Forns y Boada, 1993); para la adaptación socioemocional se aplicó la Escala de DESB (Spivack y Swift, 1967, 1975) y el Test de TOESD (Hresko y Brown, 1984), y por último para el rendimiento académico se utilizó el Test de actividad gráfomotora (prueba CS-72 de Opération Jevnes Prèmièrs, 1972), Test de conocimiento aritmético (Automatismos Aritméticos de Rey, 1967), Prueba de lectura y escritura en catalán (SEDEC, Departament d'Ensenyament del Català, 1989). La mayoría de los tests fueron aplicados en versión castellana, dado que la primera lengua de los alumnos objeto de estudio eran escolares castellanoparlantes. Los resultados aportados mostraron la no existencia de diferencias significativas en las habilidades cognitivas adquiridas por ambos grupos y una relación entre dichas habilidades y el conocimiento lingüístico alcanzado por los escolares alta y estadísticamente significativa en ambas lenguas para el grupo de inmersión y en castellano para el grupo monolingüe. Este último hecho confirma la relación que existe entre la habilidad de aprender una segunda lengua y el desarrollo lingüístico adquirido en la primera lengua. Sin embargo, la producción verbal en

catalán para el grupo de inmersión fue inferior respecto a la realizada en castellano su primera lengua. Respecto al rendimiento escolar, los sujetos del grupo de inmersión mostraron un peor rendimiento en habilidades lectoescritoras en catalán que el grupo control, sin embargo éstos no presentaron diferencias significativas en las dos pruebas aplicadas de habilidad gráfica y aritmética. Una posible explicación reside en la distinción aportada por Cummins (1984a, 1984b) entre BICS y CALP. De forma que, los grupos de inmersión centran su objetivo en la comprensión oral de una segunda lengua y por tanto desarrollan la competencia comunicativa interpersonal básica (BICS) del sujeto, y los grupos de enseñanza monolingüe en el aprendizaje de las habilidades lectoescritoras, esto es, la competencia cognitivo-lingüística del alumno (CALP). Por último, en relación a la adaptabilidad social y emocional de los alumnos analizados, los resultados conseguidos no mostraron diferencias significativas en variables como sociabilidad, ansiedad, comprensión, atención y velocidad de realización, entre otras, pero sí mostraron diferencias significativas en participación, locus de control, iniciativa, obediencia, ofreciendo los alumnos del grupo de inmersión respecto al grupo monolingüe de castellano mayor obediencia, locus de control interno, participación pero menor iniciativa y ciertas dificultades para establecer contacto personal con los profesores en su segunda lengua.

Por su parte, Serra (1989) realizó un estudio comparativo sobre diversos aspectos de la competencia lectora a final del Ciclo Inicial en programas de inmersión y de enseñanza monolingüe en castellano. Sus resultados mostraron que los alumnos escolarizados en programas de inmersión eran más lentos en habilidades de lectura, pero ofrecían una mejor comprensión lectora que los escolarizados en programas monolingües de

castellano. Sin embargo, este hecho no determinó la existencia de diferencias significativas en rendimiento escolar en ambos grupos. Estos mismos resultados se confirmaron en parte por un estudio realizado por el SEDEC. Así, pues la comparación entre una muestra amplia de alumnos de la escuela pública que seguían programas de inmersión y otros que seguían programas en castellano en Segundo, Quinto y Octavo de EGB mostraron que la comprensión escrita de los alumnos de los programas de inmersión era más alta que la de los que aprendían a leer y escribir en castellano.

Sin embargo, en un estudio más reciente realizado en Badalona durante el curso 1990-1991 (Boixaderas, Canal y Fernández, 1991), tras comparar el rendimiento en matemáticas, lengua castellana y catalana en alumnos de 3ª curso de Primaria escolarizados en programas de inmersión y monolingüe en castellano, los resultados mostraron que los grupos de inmersión obtenían peores resultados en cálculo numérico que los escolarizados en castellano; sin embargo, éstos no presentaban diferencias significativas en relación a estrategias de resolución de problemas lógicos. Una posible interpretación de este hallazgo puede ser la mayor cantidad de tiempo empleado por los programas de inmersión lingüística en el dominio de habilidades lectoescritoras, lo que los aspectos relativos al cálculo numérico quedan relegados para cursos posteriores. Así pues, estos resultados permiten concluir que los programas de inmersión catalanes no perjudican el desarrollo lógico de sus alumnos, pero dado el énfasis en el dominio lingüístico de la lengua en los primeros niveles de la enseñanza, el modelo de inmersión evaluado no permite trabajar suficientemente bien habilidades relacionadas con el cálculo matemático.

Respecto al *segundo objetivo*, basado en conocer el valor instrumental que poseen los diferentes modelos educativos en la adquisición lingüística del castellano y el catalán en todo el proceso de escolarización, el primer estudio realizado fue durante el curso 1981-82 (Gabinet d'Estudis del SEDEC: Alsina, Arnau, Bel, Garolera, Grijalvo, Perera y Vila, 1983). Su principal objetivo fue analizar los conocimientos lingüísticos adquiridos en lengua catalana y castellana en alumnos de 4º curso de Primaria de todos los modelos educativos implantados en Cataluña. Sus resultados manifestaron que los alumnos con lengua familiar castellana adquirirían un dominio mucho más pobre del catalán que los de lengua familiar catalana respecto a su dominio del castellano. Además, se observó que la dispersión sobre la competencia lingüística en lengua catalana por parte de escolares castellanoparlantes era mucho más alta que la dispersión del dominio en castellano por parte de los alumnos con lengua familiar catalana. Por lo tanto, los resultados conseguidos evidenciaron que los únicos alumnos capaces de mostrar un conocimiento equilibrado en ambas lenguas eran los escolarizados en catalán y cuya lengua familiar era también dicha lengua. Por otra parte, este mismo estudio también destacó que el factor que tenía más peso para explicar la competencia lingüística en lengua castellana y catalana era la "habilidad general para aprender". Otros factores como "modelo lingüístico escolar", "tipología lingüística del aula", "lengua familiar" y "tipo de centro" (público/privado) aparecieron como aspectos significativos en el conocimiento de la lengua catalana, mientras que el factor "condiciones lingüísticas del entorno" se revelaron como muy importantes para explicar los resultados en lengua castellana.

El siguiente estudio (Serra y Vila, 1991; Bel, Serra y Vila, 1993) se diseñó con el objetivo de evaluar el dominio lingüístico en castellano y

catalán de alumnos que estaban cursando el final de la enseñanza obligatoria y, además, estudiar los factores determinantes en su adquisición. Los resultados obtenidos revelaron que, tras 7 años de escolarización, los escolares de todos los modelos educativos implantados en Cataluña seguían dominando mejor el castellano que el catalán. Además, un resultado relevante fue que la dispersión de la competencia de cada una de las lenguas era mucho menor a medida que aumentaban los años de escolarización. Respecto a los factores explicativos del nivel de bilingüismo adquirido, se observó que mientras el dominio lingüístico del catalán, a nivel oral y escrito, dependía en buena medida de factores como su presencia en el curriculum escolar, y la situación lingüística familiar; el conocimiento del castellano era totalmente independiente de dichos factores, lo cual evidenciaba la importante presencia ambiental de dicha lengua en el territorio catalán (Arenas y Abeyà, 1984; SEDEC, 1985).

En resumen, el seguimiento de los programas de educación bilingüe implantados en Cataluña permiten extraer dos conclusiones principales: 1) En general, que las distintas tipologías lingüísticas analizadas, incluso cuando éstas comportan un cambio de lengua hogar-escuela, no revelan diferencias significativas respecto al rendimiento académico de los sujetos catalanes; y 2) Que tras 10 años de programas de enseñanza bilingüe, se puede apreciar la existencia de un cambio en los factores determinantes del nivel de competencia adquirida en la lengua catalana. Mientras que en 1982 el dominio lingüístico del catalán dependía de factores como el modelo lingüístico, la zona lingüística, la lengua familiar y el tipo de centro; en cambio, en 1990 la competencia lingüística del catalán a nivel oral y escrito dependía únicamente del modelo lingüístico escolar. Respecto al dominio lingüístico en castellano, los resultados ofrecidos mostraron que

los escolares de Cataluña, independientemente de su lengua familiar y del tipo de programa al que asisten, son capaces de desarrollar un buen conocimiento de la lengua castellana. Quizás la justificación de este resultado se encuentre en la enorme presencia social que el castellano tiene en Cataluña. Las condiciones lingüísticas del entorno ofrecen al alumno muchas posibilidades de efectuar intercambios sociales en castellano, lo que provoca que esta variable se convierta en una fuente importante del desarrollo lingüístico de dicha lengua.

#### **1.5.3.2. Evaluación de programas en Galicia.**

En la actualidad, la lengua autóctona de Galicia, el gallego, ha tomado un importante impulso. Además, de ser la lengua que se hablaba en las zonas rurales que nunca se castellanizaron, comienza a utilizarse en ámbitos cada vez más amplios de la sociedad gallega, desde la Universidad al mundo profesional y de los negocios. En concreto, el panorama lingüístico del sistema escolar gallego ha recorrido un trecho importante a lo largo de los últimos 15 años, pasando de una situación en la que se producían sólo algunas experiencias aisladas de enseñanza en gallego a otra en la que se imparte el gallego como asignatura de estudio en la totalidad de los centros educativos.

Sin embargo, debido a su reciente incorporación en el sistema educativo es pronto para evaluar los resultados de estos programas. Todavía, hoy en día no hay datos disponibles que se puedan citar. Únicamente, señalar que las escasas investigaciones sobre bilingüismo realizadas en Galicia (Pérez Vilarinho, 1979; Pérez Pardo, 1985; González Lorenzo, 1985) se han centrado, principalmente, en el estudio de la

influencia de la lengua familiar gallega en el rendimiento académico de los alumnos escolarizados en castellano. Básicamente, estos trabajos revelan la importancia de desarrollar programas bilingües de mantenimiento para aquellos escolares que habitualmente emplean el gallego en sus relaciones personales, dada la rápida y progresiva integración de su lengua autóctona en la sociedad gallega.

#### **1.5.3.3. Evaluación de programas en el País Vasco.**

Respecto a la enseñanza bilingüe en el País Vasco, en los últimos años se han publicado diversos trabajos que han evaluado el conocimiento del euskera y del castellano de los alumnos escolarizados en los distintos modelos educativos adoptados en este territorio (Garagorri y Eguilior, 1983; Garmendia, 1985; Etxeberria, 1986; Gabiña y cols, 1986; Sierra y Olaziregi, 1989, 1990; Olaziregi y Sierra, 1993).

Uno de los primeros estudios realizados (Garagorri y Eguilior, 1983) estudió la influencia del modelo lingüístico sobre el rendimiento académico y el conocimiento del euskera en los alumnos vascos. Pero sus resultados fueron poco relevantes debido a que este estudio presentó graves problemas metodológicos. Sin embargo, a finales de 1983, el Gobierno Vasco promueve los denominados estudios EIFE (Euskararen Irakaskuntza: Faktoreen Eragina: Estudios ) 1, 2 y 3 que fueron publicados en 1986, 1989 y 1990, respectivamente. Éstos tienen como objetivo principal evaluar el conocimiento del castellano y del euskera de los escolares castellanoparlantes en función del modelo lingüístico al que asisten, y determinar qué factores son determinantes en el aprendizaje de ambas lenguas. La evaluación de programas se realizó en una muestra



representativa de escolares de 2º y 5º curso de Primaria en los modelos A, B y D implantados en el País Vasco (Gabiña y cols., 1986). Las pruebas utilizadas evaluaron los 4 destrezas lingüísticas: comprensión oral, expresión oral, comprensión escrita y expresión escrita, tanto en castellano como en euskera. Los resultados obtenidos mostraron que entre las variables que tenían más peso en el aprendizaje del castellano en los tres modelos considerados, la más relevante era la aptitud personal del alumno (coeficiente de inteligencia y el rendimiento escolar). En el caso del euskera, aparecieron como factores significativos en el modelo D (aprendizaje de habilidades lectoescritoras en euskera) la aptitud personal del individuo y el ambiente lingüístico familiar, y en el resto de modelos A y B (habilidades lectoescritoras en castellano) fueron el ambiente lingüístico familiar y escolar (lengua familiar, tiempo dedicado al euskera en la escuela) los aspectos más significativos. Por otra parte, estos estudios también revelaron que el conocimiento del castellano era relativamente homogéneo en todos los alumnos castellanoparlantes de los programas bilingües implantados en el País Vasco, si bien las puntuaciones más bajas, pero sin llegar a mostrar diferencias estadísticamente significativas, se obtenían en el modelo de escolarización temprana en euskera, donde las habilidades lectoescritoras son aprendidas en euskera (Olaziregi y Sierra, 1993).

En resumen, los datos aportados por estos estudios señalan que la escolarización temprana en euskera en sujetos castellanoparlantes no deteriora el desarrollo lingüístico de la lengua familiar, sino todo lo contrario, permite acceder a un buen dominio del castellano a pesar de que esta lengua no tenga una presencia relevante en el currículum escolar recibido. Por otra parte, estos trabajos también revelan la importancia de

las características individuales del alumno en la obtención de buenos rendimientos en castellano, y el mantenimiento de la lengua autóctona en el ámbito familiar y su presencia en el curriculum escolar como aspectos importantes en el conocimiento del euskera.



## ***CAPITULO 2***

**La situación bilingüe en la Comunidad Valenciana**



## **2.1. INTRODUCCION**

El objetivo de nuestro estudio se contextualiza en la Comunidad Valenciana, donde existe una situación de contacto de lenguas, que posee unas características muy peculiares. La Comunidad Valenciana presenta una realidad bilingüe donde conviven dos lenguas, el castellano y el valenciano, pero donde a la vez la presencia de cada una de ellas es muy diferente según la zona geográfica de que se trate. Por tanto, un análisis de la contextualización sociolingüística de la Comunidad Valenciana será un aspecto importante en el estudio del bilingüismo en la citada comunidad.

Hoy en día, las circunstancias sociolingüísticas de la sociedad valenciana son un determinante indiscutible en el uso del valenciano, como lengua habitual de comunicación en el ámbito familiar y coloquial y, no menos importante, como lengua de escolarización en la enseñanza de determinados contenidos curriculares escolares. Precisamente, la presencia o poder social de cada lengua en la población valenciana ha sido uno de los elementos, quizás el más relevante, en la planificación y puesta en marcha de los programas de enseñanza bilingüe en la Comunidad Valenciana. Por tanto, la configuración sociolingüística de la comunidad, especialmente en dos zonas lingüísticas claramente diferenciadas (zona castellanoparlante y zona valencianoparlante) va a ser un componente importante en la identificación de la lengua familiar y/o los modelos de educación bilingüe valencianos, variables fundamentales en el planteamiento de esta investigación.

De este modo, dentro de este capítulo, en primer lugar vamos a ofrecer una breve perspectiva histórica de la Comunidad Valenciana, con el fin de descubrir las causas de la actual situación bilingüe valenciana. En segundo lugar, se presentará una visión de la situación lingüística de la sociedad valenciana mediante la consulta de algunos estudios realizados por la Consellería de Educación y Ciencia (1990, 1991, 1995) en la Comunidad Valenciana. Y por último, para finalizar se expondrán los modelos de educación bilingüe implantados en la comunidad, producto de la política lingüística defendida por la Generalitat Valenciana, así como un resumen de los principales resultados obtenidos tras su evaluación.

## 2.2. PERSPECTIVA HISTORICA

Una buena comprensión de la situación lingüística en la Comunidad Valenciana requiere remontarnos a los orígenes de la sociedad valenciana.

La actual "Comunidad Valenciana" comienza a existir como un país diferenciado en el siglo XIII, cuando el rey Jaime I conquistó y repobló estos territorios. En ese momento, las tropas conquistadoras estaban formadas principalmente por catalanes y aragoneses. Por tanto, la repoblación de este país se realizó a partir de la gente procedente de estas dos comunidades, siguiendo la siguiente distribución: la zona de la costa y los núcleos urbanos fueron repoblados por los catalanes, mientras que las tierras del interior lo fueron por los aragoneses.

Precisamente, en la duplicidad de esta procedencia se encuentra la razón del bilingüismo actual de los valencianos: "La distribución geográfica de las lenguas está en función de la naturaleza de los colonizadores de cada comarca" (Fuster, 1962). Este hecho hace que aparezcan dos zonas lingüísticas claramente diferenciadas según la procedencia de los repobladores: las zonas interiores (Viver-Sogorb, Aiora, Bunyol, etc., y las que se incorporan posteriormente, como es el caso de Utiel) que hablarán la lengua castellana, y la zona litoral que adoptará la lengua de los repobladores catalanes. Con el paso del tiempo, ésta última se convertirá en la lengua dominante y, con posterioridad, en el idioma oficial del Reino, siendo utilizada por la Corte Real, el Gobierno y la Administración.



Los siglos siguientes se caracterizaron por una gran prosperidad del Antiguo Reino de Valencia alcanzando su máximo esplendor en el siglo XV, considerado por muchos autores como el "siglo de oro" valenciano. En ese siglo, la lengua valenciana consigue un notable esplendor y cultivo literario, con autores como Ausias March, Rois de Corella, Jaume Roig y Joanot Martorell, entre otros.

Sin embargo, en el siglo XVI desaparece dicho esplendor y comienza el declive. La "guerra de les Germanies", junto con el fracaso de los "agermanats" provoca que la nobleza valenciana comience su desvinculación con la lengua propia. Comienza en ese momento histórico, una creciente castellanización de las clases dirigentes del País Valenciano siendo, además, un hecho defendido por la monarquía castellana (Fuster, 1956). El final de la guerra de Sucesión, con la derrota de la batalla de Almansa y el Decreto de "Nova Planta", hacen que el castellano se imponga como lengua oficial en todo el territorio valenciano.

A lo largo del siglo XVIII, diversas cédulas y numerosas disposiciones reales van reduciendo el ámbito del uso del valenciano hasta llegar a prohibir su presencia en el contexto escolar. El país sufre un fuerte proceso de castellanización y el valenciano sólo se conserva en los ámbitos familiares y coloquiales, especialmente de las zonas rurales. Por su parte, las clases sociales superiores continúan utilizando el castellano como medio de comunicación. De esta forma, "castellanizarse" se constituye como el modelo a imitar por la gran mayoría de las clases inferiores, ya que esta opción va vinculada al símbolo de la ascensión social. Todo ello provoca en la clase trabajadora un cierto sentimiento de rechazo hacia la

propia lengua, ya que el progreso social estaba ligado a la pérdida de la identidad cultural, la asimilación al grupo dominante y, posiblemente, un cierto menosprecio de su grupo de origen.

Sin embargo, en el siglo XX en la década de los sesenta tienen lugar dos hechos aparentemente contradictorios en la sociedad valenciana. El primero se refiere a que la lengua valenciana pierde mucho terreno de actuación. La aparición de una serie de factores, como la extensión de la educación obligatoria en castellano, la creciente influencia de los medios de comunicación en castellano y el cambio de una sociedad agraria a una sociedad industrial junto con la llegada de una gran cantidad de inmigrantes castellanoparlantes, no favorecen en absoluto la presencia y utilización del valenciano como lengua habitual de comunicación. El segundo suceso tiene lugar cuando empiezan a aparecer en algunos sectores, sobre todo intelectuales, una serie de iniciativas con una orientación claramente nacionalista ligadas a la recuperación de la propia lengua. La aparición de estas manifestaciones, ligadas ideológicamente al antifranquismo y en defensa de la lengua valenciana, son consideradas como las primeras generaciones nacionalistas del País Valenciano (Mollà, 1994).

La importancia de estos hechos reside en que para algunos autores valencianos la actual situación lingüística de la Comunidad Valenciana tiene su origen en dichos incidentes (Ninyoles, 1986; Domenech, 1990; Pasqual y Sala, 1991). Un aumento de la industrialización y de la cultura urbana provocada por los fuertes movimientos migratorios hacia los núcleos industriales más castellanizados, tanto de las zonas rurales valencianoparlantes como de las comarcas interiores castellanoparlantes y

otras zonas del Estado Español, exigían una población alfabetizada y con un mínimo dominio del castellano, lo que provoca una alteración de las redes sociales, las relaciones entre las personas y las pautas en el uso lingüístico de la comunidad. Además, la aparición de ciertos cambios políticos también originan importantes transformaciones en la estructura sociolingüística de la Comunidad Valenciana, pasando de un conflicto latente de dos idiomas a un conflicto manifiesto, con todas las tensiones sociales que comporta (Ninyoles, 1986). En aquel momento, se origina un movimiento de recuperación lingüística de la lengua valenciana en los ámbitos universitarios. Y en los círculos intelectuales y artísticos del País son muy frecuentes las actitudes que demuestran tener conciencia del valor que puede aportar la lengua autóctona del País (Lluch, 1976; Cucó, 1989; Sanchis, 1978; Mollà, 1994).

A grandes rasgos, ésta es la situación sociolingüística de la Comunidad Valenciana cuando llega la transición democrática del post-franquismo y posteriormente el "Estatut d'Autonomia" (1983).

Tras la aprobación de la "Llei 4/1983 d'ús i Ensenyament del Valencià" en noviembre de 1983, la lengua valenciana se incorpora de forma obligatoria en el sistema educativo de la Comunidad Valenciana. A partir de este momento, la aplicación de este decreto junto con la política educativa defendida por la Generalitat Valenciana son totalmente determinantes en los programas de enseñanza bilingüe implantados en la Comunidad Valenciana, un elemento crucial para nuestro estudio, y que posteriormente abordaremos.

### 2.3. LA SITUACION LINGÜISTICA EN LA SOCIEDAD VALENCIANA

La utilización de padrones municipales de habitantes como fuente de recogida de información sobre la realidad sociolingüística de una sociedad ha sido cuestionada por algunos autores, principalmente por las propias limitaciones que existen en la subjetividad de las respuestas de los ciudadanos al contestar el padrón. Sin embargo, otros autores consideran que ésta es una inestimable fuente de datos referentes al uso de la lengua (Molla y Palanca, 1987), ya que su manejo permite examinar esta variable lingüística junto con otras variables incluidas en el censo municipal, tales como sexo, edad, residencia, etc... Por tanto, en este apartado se analizará la situación lingüística de la población valenciana tras la consulta de algunos estudios realizados por la Consellería de Cultura, Educación y Ciencia (1990, 1991, 1995) en la Comunidad Valenciana.

Una **primera clasificación** de las zonas lingüísticas de la Comunidad Valenciana procede de un estudio oficial (Consellería de Cultura, Educación y Ciencia, 1992) realizado con los datos del censo municipal de 1991. Dada la existencia de dos zonas lingüísticas claramente definidas en la Comunidad Valenciana, según predomina el castellano o el valenciano, podemos distinguir las siguientes zonas o sectores:

(1) *Zona lingüística castellanoparlante*, que incluye once comarcas de habla castellana: L'Alt Millars, L'Alt Palància, El Racó d'Ademús, Els Serrans, La Foia de Bunyol, La Plana d'Utiel, La Vall d'Aiora, La Canal de Navarrés, El Baix Segura, L'Alt Vinalopó y una parte del Vinalopó Mitjà. No obstante, a pesar de que la extensión del territorio castellanoparlante es

amplia, el 41,6% del territorio, es importante destacar que su participación en número de habitantes es muy escasa, constituyendo sólo el 9,3% de la población valenciana (Pasqual y Sala, 1991) y, además, predominantemente localizada en zonas rurales y poco pobladas.

(2) El resto de las comarcas son de *predominio lingüístico valenciano*. En ellas, podemos distinguir cuatro variedades lingüísticas tradicionalmente aceptadas: el habla de transición entre el catalán meridional y el nortoccidental en las comarcas del norte de Castellón, el catalán meridional en la Plana y la mayor parte de las comarcas al sur de la Ribera, el valenciano "apitxat" de las comarcas centrales del País (Camp del Túria, Camp de Morvedre, L'Horta y La Ribera) y el subdialecto "salat" en algunos lugares de las comarcas de La Marina y El Comptat. Precisamente, esta variación lingüística ha provocado una cierta desorientación entre muchos valencianos debido a las grandes diferencias léxicas entre una variedad y otra, y en consecuencia, una falta de conciencia de la unidad lingüística de la lengua valenciana. Dentro de esta zona, podemos distinguir entre:

a) *Zona lingüística con tasa alta de valencianización*. Comprende las comarcas con un porcentaje superior al 80% de valencianoparlantes: L'Alt Maestrat (93.9%), Els Ports (92.3%), El Comptat (91%), La Vall d'Albaida (90.7%), L'Alcalatén (86.4%), La Safor (86.3%), El Baix Maestrat (85.2%), La Costera (87.2%) y La Ribera Alta (86.4%).

b) *Zona lingüística con tasa media de valencianización*. Incluye las poblaciones con una proporción de valencianoparlantes comprendidos entre el 41% y el 80%: El Camp de Túria (76.4%), La Marina Alta (75.4%), La

Plana Baixa (72.1%), L'Alcoià (68.6%), La Plana Alta (62.4%), L'Horta Sud (59.9%), L'Horta Nord (57.3%), La Marina Baixa (48.6%), El Camp de Morvedre (44.9%) y El Baix Vinalopó (39.9%).

c) *Zona lingüística con valencianización baja.* Engloba las localidades de comarcas con un porcentaje de valencianoparlantes igual o inferior al 40%: València ciutat (39.7%), L'Horta Oest (32.9%) y L'Alacantí (26.2%).

Una **segunda clasificación** se establece a partir de una "Encuesta sobre l'ús del valencià" realizada, en mayo de 1995, por la Dirección General de Política Lingüística de la Consellería de Educación y Ciencia. Los datos extraídos recogen información sobre el "*índex general de domini del valencià*" en la población valenciana. Esta medida integra la información procedente de cuatro categorías: 1) "*entende el valenciano*", que equivale a comprender una conversación en valenciano, pero no ser capaz de interactuar en ese idioma; 2) "*sabe hablar en valenciano*", significa la capacidad de mantener una conversación sobre un tema cotidiano en valenciano; 3) "*sabe leer en valenciano*", indica la capacidad de leer textos habituales; y 4) "*sabe escribir en valenciano*", mide la capacidad de redactar textos y notas habituales en valenciano con una corrección suficiente, aunque no sea perfecta. Los resultados obtenidos se ofrecen por comarcas, al igual que en el caso anterior. De este modo, podemos agrupar la Comunidad Valenciana en cinco niveles:

a) Comarcas con **alto dominio** de valenciano: El Comptat, La Marina Alta, Els Ports, L'Alt Maestrat, El Baix Maestrat, L'Alcalatén, La

Plana Baixa, El Camp del Túria, La Ribera Alta, La Ribera Baixa, La Costera, La Vall d'Albaida y La Safor.

b) Comarcas con nivel **medio-alto** de valenciano: L'Alcoià, La Plana Alta, El Camp de Morvedre, L'Horta Nord, La Ciutat de València y L'Horta Sud.

c) Comarcas con nivel **medio-bajo**: La Marina Baixa, El Baix Vinalopó y L'Horta Oest.

d) Comarcas con nivel **bajo**: El Vinalopó Medio, L'Alacantí, La Canal de Navarrés, La Foia de Bunyol, Els Serrans, L'Alt Palància, L'Alt Millars y L'Alt Vinalopó.

e) Comarcas con nivel **muy bajo** de dominio del valenciano: El Racó d'Ademús, La Plana d'Utiel-Requena, La Vall d'Aiora y El Baix Segura, comarcas que corresponden a la zona castellanoparlante de la Comunidad Valenciana, aunque no todas las comarcas del territorio castellanoparlante se incluyen en este nivel.

Para una mejor comprensión, hemos aportado un mapa geográfico de la Comunidad Valenciana (ver figura 3.1).





Por último, para finalizar este apartado queremos hacer referencia a algunos factores demográficos que, siguiendo los estudios poblacionales valencianos (Conselleria de Cultura, Educació i Ciència, CCEC, 1990, 1995), permiten explicar mejor el uso y dominio de la lengua valenciana en la sociedad actual. Estas variables, por orden de importancia, son:

1) el *lugar de nacimiento del sujeto*. El hecho de nacer en la Comunidad Valenciana, Cataluña y Baleares o no en cualquier otra ciudad española o en el extranjero configura la variable con mayor nivel de discriminación en el dominio lingüístico valenciano;

2) la *comarca de residencia*. La población que más domina el valenciano son aquellos nacidos en la Comunidad Valenciana, Cataluña y Baleares y residentes en las siguientes comarcas: El Comtat, L'Alcoià, La Marina Baixa, Els Ports, L'Alt Maestrat, El Baix Maestrat, L'Alcalatén, La Plana Baixa, La Costera, La Vall d'Albaida y la Safor;

3) el *hábitat*. En las comarcas valencianoparlantes se domina más el valenciano en los municipios pequeños hasta 50.000 habitantes, que en las grandes ciudades como València, Castelló, Xàtiva, Gandia, etc. Sin embargo, en las comarcas castellanoparlantes son, justamente, las ciudades de más de 20.000 habitantes las que obtienen un dominio del valenciano mayor que las ciudades pequeñas de esas comarcas; y por último,

4) la *edad* de la persona. Dada la reciente incorporación de la educación bilingüe a la comunidad valenciana, el colectivo que presenta una mayor competencia lingüística valenciana son los jóvenes en edad escolar.

En definitiva, como se puede observar la situación sociolingüística de la Comunidad Valenciana es uno de los elementos más relevantes en el estudio del dominio lingüístico del valenciano como lengua de comunicación en el ámbito familiar y coloquial, así como lengua de enseñanza en los modelos de educación bilingüe implantados en la Comunidad Valenciana, como veremos a continuación. Dada la importancia de estos estudios basados en el contexto sociolingüístico valenciano, para la localización e identificación de las anteriores variables las cuales serán determinantes para el propósito de nuestra investigación, algunos factores mencionados en estos estudios fueron incluidos en una encuesta sociolingüística administrada a cada sujeto participante de este estudio, con el fin de llevar a cabo una selección de los grupos lingüísticos componentes de nuestra muestra lo más rigurosa posible. La encuesta sociolingüística utilizada será presentada en un capítulo posterior.

#### **2.4. LA EDUCACION BILINGÜE EN LA COMUNIDAD VALENCIANA**

El 23 de noviembre de 1983, las Cortes Valencianas aprobaron la "*Llei d'Us i Ensenyament del valencià*". En dicha ley, el artículo 18 expresa que "*la incorporación del valenciano a la enseñanza es obligatoria en todos los niveles educativos*", y el artículo 19 manifiesta que "*con independencia del programa de enseñanza bilingüe implantado, al finalizar la educación elemental, los alumnos valencianos deben de estar igualmente capacitados, a nivel oral y escrito, en la utilización del valenciano y el castellano*".

Así pues, a partir de ese momento, el bilingüismo educativo se convierte en uno de los objetivos prioritarios del sistema educativo

valenciano. Pero, ¿cuáles fueron las soluciones adoptadas para conseguir este objetivo?. Hasta ahora, la Comunidad Valenciana ha llevado a cabo una política lingüística que contempla tres modelos educativos diferentes: (1) Enseñanza monolingüe en castellano; (2) Enseñanza bilingüe: 2a) Programa de mantenimiento de la lengua y cultura, y 2b) Programa de enriquecimiento. Seguidamente, pasamos a describir algunas de sus principales características.

(1) ENSEÑANZA MONOLINGÜE EN CASTELLANO.

El modelo de "enseñanza monolingüe en castellano" es un programa dirigido principalmente a aquellos alumnos en su mayoría castellanoparlantes y que, además son residentes en territorios valencianos pero con predominio lingüístico del castellano. Así pues, en términos generales este modelo educativo predominará en aquellas zonas donde la presencia ambiental del valenciano es prácticamente inexistente, como son las comarcas de la zona castellanoparlante de la Comunidad Valenciana. Básicamente, el objetivo central de este programa reside en la enseñanza íntegra en castellano, a excepción del valenciano como materia o asignatura de estudio a lo largo de la educación obligatoria, y el idioma extranjero elegido por el centro escolar en cuestión (francés o inglés).

Precisamente, una de las soluciones adoptadas por la Generalitat Valenciana ha sido implantar la enseñanza del valenciano como asignatura, desde el año 1983, en los niveles de Educación Preescolar, Educación General Básica, Bachillerato y Formación Profesional de todos los centros escolares del territorio lingüístico valenciano, quedando eliminados de su

enseñanza aquellos alumnos que lo soliciten por residencia temporal y que presentan las circunstancias determinadas por el artículo 24 de la 'Llei d'Us i Ensenyament del valencià'.

Sin embargo, la enseñanza del valenciano como materia académica no se puede considerar un tipo de enseñanza bilingüe (Fishman, 1976). Por tanto, los programas que sólo incluyen el valenciano como una asignatura más se denominarán "Programas de enseñanza monolingüe en castellano".

## (2) ENSEÑANZA BILINGÜE.

Si se quieren cumplir los objetivos propuestos en la "Llei d'Us i Ensenyament del valencià" es necesario ofertar modelos de enseñanza bilingüe, donde se contemple la posibilidad de la enseñanza de contenidos curriculares escolares en dos lenguas, en castellano y en valenciano. Sin embargo, antes de proceder a la implantación de dichos programas es necesario conocer cuál es la situación sociolingüística de la población a la cual van dirigidos.

Dentro de la Comunidad Valenciana, podemos distinguir dos poblaciones claramente diferenciadas. Un *primer grupo* está compuesto por alumnos valencianoparlantes y residentes en comarcas donde el habla mayoritaria es el valenciano. En este grupo, si el propósito final es el mantenimiento y la recuperación del valenciano es lógico que reciban la enseñanza en su propia lengua, al margen de la presencia que el castellano también tenga a lo largo de dicho programa. El *segundo grupo* está formado por alumnos valencianos que hablan castellano. Dentro de éste, pueden clasificarse tres

subconjuntos diferentes: a) un primer grupo minoritario con un nivel sociocultural elevado de los padres, los cuales tienen contacto con la lengua y la cultura valenciana y están motivados para aprender el valenciano, transmitiendo el deseo de aprenderlo a sus hijos pero sin rechazar el castellano; b) un segundo grupo, el más problemático, hijos de emigrantes que viven en concentraciones urbanas donde no se habla el valenciano, pero su curriculum escolar debe incluir tanto el castellano como el valenciano, y c) un tercer grupo que vive en comarcas donde el valenciano perdió mucho terreno debido a la política de unificación lingüística, pero se sienten ligados a la cultura valenciana.

En definitiva, se puede apreciar que las circunstancias sociolingüísticas para cada grupo son bastante diferentes. Por esta razón, siempre hay que tenerlas presentes en la evaluación de los resultados sobre el bilingüismo educativo valenciano y sus consecuencias en el rendimiento académico o desarrollo cognitivo de los sujetos valencianos. Dentro de este modelo, podemos distinguir dos programas:

*2a) Programa de Mantenimiento de la Lengua y la Cultura.*

Es un tipo de programa dirigido principalmente a los alumnos que viven en territorios de predominio lingüístico valenciano, y su objetivo principal es el mantenimiento de la lengua y cultura valenciana en dichos escolares. Sus principales características se encuentran en un uso predominante del castellano como lengua de instrucción a lo largo de toda la escolarización obligatoria y la implantación del valenciano como lengua de enseñanza únicamente para el *área de experiencias* en el Ciclo Medio, y

las asignaturas *ciencias sociales y/o naturales* en el Ciclo Superior, además claro está de la enseñanza del valenciano como materia de estudio en toda la Educación General Básica.

De hecho, la propuesta acerca del programa "Área de experiencias del medio" para el Ciclo Medio de la Educación General Básica, aprobada por la Orden del 20 de mayo de 1986, determina la aplicación de cinco horas semanales del valenciano como lengua de enseñanza en el área de experiencias: *materia referida a la propia herencia étnica y cultural*. Posteriormente, la actuación del valenciano se amplía al ciclo superior, concretamente a las asignaturas de ciencias sociales y/o naturales en función del plan de estudios aprobado en cada centro escolar.

*2b) Programa de Enriquecimiento.*

El "Programa de Enriquecimiento" está dirigido especialmente a aquellos alumnos en su mayoría valencianoparlantes y, además, residentes en los territorios con predominio lingüístico valenciano. Básicamente, su propósito se centra en favorecer la convivencia de las dos lenguas dentro del contexto social, mostrando la lengua propia como vehículo ideal de comunicación, instrucción y cultura en los valencianoparlantes, así como inculcar un respeto a la lengua y cultura propias del país y su consideración como medio de enriquecimiento cultural e intelectual en los castellanoparlantes. Este programa está en consonancia con las tesis de Lambert (1981) y Cummins (1982) de potenciar la lengua más desfavorecida en el sistema educativo.

Sin embargo, la presencia de este programa en el sistema educativo valenciano ha estado determinada por multitud de factores, tales como la opción solicitada por los padres, la situación sociolingüística concreta del alumno, los medios disponibles para llevarlo a cabo, lo cual no ha sido una buena garantía para su correcta implantación. Concretamente, la propia Consellería de Educación y Ciencia de la Generalitat Valenciana ha reconocido que la implantación de este tipo de programas ha acarreado muchos problemas, como el reciclaje deficiente en valenciano del profesado, la oposición de grupos sociales activos, la propia organización de los centros y la metodología utilizada, la inclusión de dichos programas como una línea más dentro del mismo centro.

En definitiva, queda claro que la presencia del bilingüismo y la educación bilingüe aparecen como aspectos muy significativos en la vida cotidiana de todos los valencianos, especialmente de aquellas personas pertenecientes a ciertas minorías lingüísticas. Partiendo de esta realidad, resulta bastante evidente que el interés de nuestro estudio resida en los tres modelos educativos presentados. Con el fin de realizar un adecuado análisis e interpretación de los resultados obtenidos tras la medición y evaluación de las aptitudes intelectuales de los escolares de dichos programas, a continuación procedemos a presentar un breve resumen de los resultados obtenidos en la evaluación de estos programas.

#### **2.4.1. Evaluación de programas educativos valencianos**

Desde la aprobación de la "Llei d'Us i Ensenyament del Valencià" y su posterior proceso de implantación, el "Servei d'Ensenyament en Valencià" de

la Consellería de Educación y Ciencia Valenciana ha puesto en marcha varios estudios que han evaluado el dominio lingüístico en valenciano y castellano de aquellos alumnos escolarizados en los distintos modelos educativos implantados en la Comunidad Valenciana. Básicamente, la mayoría de estos estudios han intentado establecer ciertas relaciones entre la tipología lingüística y el rendimiento académico de los alumnos, para lo cual han centrando sus investigaciones en la adquisición de las habilidades de lectura y escritura en los distintos programas bilingües implantados, ya que, en definitiva, el éxito implicado en el dominio de dichas habilidades está en relación con el posterior éxito o fracaso escolar.

El primer estudio, realizado por Miralles, Gotor, Cervera, Monterde y Sempere (1989), se inscribió dentro del "Projecte d'Avaluació Lingüística d'Alumnes Escolaritzats en Valencià". Su objetivo era evaluar la competencia lingüística en castellano y valenciano en alumnos de lengua familiar valenciana en el Ciclo Inicial de EGB. Los modelos educativos evaluados fueron: programa monolingüe en castellano y programa de enseñanza bilingüe tipo 2b (ver tipología de modelos educativos valencianos). Los instrumentos de medida empleados se basaron en dos pruebas, una en valenciano y otra paralela en castellano, con el fin de medir el dominio lingüístico de ambas lenguas en los dos grupos a comparar. Los resultados mostraron que no existían diferencias significativas en los conocimientos adquiridos en valenciano y castellano en función del modelo lingüístico o la lengua de la prueba administrada. Sin embargo, no sucedió lo mismo tras analizar los distintos factores que componían las pruebas utilizadas. Los alumnos escolarizados en valenciano manifestaron un equilibrio más sólido en la adquisición de las habilidades propias de cada factor en las dos



lenguas que los que recibían la enseñanza en castellano. En concreto, los resultados obtenidos mostraron que, tras promediar los resultados totales en ambas lenguas, el grupo de alumnos escolarizados en valenciano presentaba un rendimiento escolar significativamente menor en conocimiento lingüístico general, conocimiento léxico y conceptos, mientras que los escolarizados en castellano realizaban peor las tareas de discriminación auditivo-visual.

En un segundo estudio, Gotor, Miralles y Perea (1990) evaluaron la incidencia de la lengua de escolarización en el rendimiento escolar de los escolares valencianos en los primeros años de asistencia a la institución escolar. En concreto, su objetivo era evaluar la influencia del valenciano en la enseñanza del área de experiencias (programa bilingüe 2a) sobre el rendimiento académico en alumnos de 3º curso de Primaria. La evaluación del aprovechamiento lingüístico se realizó mediante la prueba PAL (Prova d'Aprofitament Lingüístic: Gotor, Miralles, Cervera y Perea, 1990) y el rendimiento matemático con una prueba de aprovechamiento en matemáticas (Gotor y cols, 1989). Ambas pruebas fueron aplicadas a ambos grupos en sus dos versiones: castellano y valenciano. Los resultados obtenidos respaldaron claramente la importancia de la lengua familiar como lengua de enseñanza de ciertos contenidos curriculares escolares. Los alumnos escolarizados en valenciano manifestaron un aprovechamiento lingüístico superior, destacando sobre todo en el conocimiento morfológico en comparación con el obtenido por alumnos escolarizados en castellano. Este resultado indica que el mantenimiento de la primera lengua como medio vehicular de instrucción apunta a un buen aprovechamiento de la primera y segunda lengua, en comparación con la situación escolar en la que se recibe

la enseñanza en la segunda lengua. Sin embargo, los alumnos del programa bilingüe presentaron algunas deficiencias significativas en ortografía castellana. Respecto a la lengua de la prueba, ésta no tuvo efectos significativos en el rendimiento escolar. Por otra parte, en relación al aprovechamiento matemático los resultados mostraron un rendimiento significativamente mayor en aquellos sujetos escolarizados en el modelo bilingüe. Precisamente, dichas diferencias fueron significativas en factores donde la presencia de los aspectos lingüísticos eran más evidentes. Estos resultados demuestran que el rendimiento en matemáticas, al final del ciclo inicial no está libre de los efectos de las situaciones contextuales y, en consecuencia, confirman la interrelación entre los procesos lingüísticos y cognitivos de los sujetos.

Otra investigación realizada por Pascual y Domenech (1989) estudió el dominio lingüístico del castellano y el valenciano en una muestra de escolares, donde un grupo recibía el área de experiencias en castellano y el otro en valenciano a lo largo del Ciclo Medio. Los resultados obtenidos mostraron que los alumnos del programa bilingüe (2a) progresaban en las variables lingüísticas de lectura (exactitud, comprensión y velocidad), escritura (dictado) y fluidez verbal asociadas al valenciano de forma más sistemática, positiva y coherente que el grupo escolarizado en castellano. En cuanto a las mismas variables lingüísticas pero en castellano, se observó que ambos modelos no presentaban diferencias significativas en la variable comprensión lectora; pero sí se apreció una cierta tendencia en el grupo escolarizado en castellano a obtener mejores resultados en variables como fluidez verbal, velocidad lectora y dictado. Finalmente, en relación al rendimiento académico del área de experiencias, los resultados mostraron

que el grupo de enseñanza en valenciano superaba significativamente a los de enseñanza en castellano.

Más recientemente, Bueno y Pascual (1995) realizaron un estudio donde se evaluó la competencia lingüística del castellano y valenciano en alumnos de 5º curso de Primaria, en función del modelo lingüístico y el contexto sociolingüístico inmediato. Los tres modelos educativos implantados en la Comunidad Valenciana fueron objeto de análisis. Y la competencia lingüística alcanzada en valenciano y castellano se midió con la aplicación de las pruebas VAL y CAL, respectivamente. Ambas pruebas son dos instrumentos de medida que permiten realizar un análisis lectoescritor del valenciano (VAL) y el castellano (CAL). Su utilización permite analizar y evaluar los principales factores que intervienen en la adquisición de la competencia lingüística de una lengua, desde niveles de morfosintaxis, comprensión lectora, fluidez verbal hasta tareas de producción y riqueza semántica, entre otras. Tras realizar los análisis de varianza pertinentes, los resultados obtenidos permitieron ofrecer las siguientes conclusiones:

(1) Respecto al *modelo educativo*, los alumnos escolarizados en *programas de enriquecimiento* (2b) dominaban mejor el valenciano en tareas de asociación, comprensión, producción y discriminación de palabras. En cambio, sus resultados se invirtieron en cuanto a su dominio del castellano, obteniendo los peores resultados en este tipo de pruebas, comparadas con los programas monolingüe y de mantenimiento. Los escolares de *programas de mantenimiento* comparativamente con el resto, presentaron los mejores resultados en tareas de comprensión en castellano, y destacaron aunque no significativamente en tareas de discriminación en valenciano y de

corrección tanto en castellano como en valenciano. En general, este grupo ofreció resultados razonables en ambas lenguas, destacando sobre todo en castellano. Respecto al modelo *monolingüe*, éste presentó los mejores y peores resultados en castellano y valenciano respectivamente, en comparación con el programa de enriquecimiento, y únicamente destacó en tareas de corrección en lengua castellana respecto al grupo de mantenimiento).

(2) Em cuanto a la influencia del *entorno sociolingüístico* o presencia social del valenciano en el dominio lingüístico de ambas lenguas, los resultados obtenidos presentaron tres segmentos poblacionales diferentes:

a) Los grupos situados en zonas lingüísticas con baja presencia del valenciano (0-40% valencianoparlantes) destacaron únicamente en la corrección escrita en castellano, puntuando más bajo en el resto de las pruebas.

b) Los grupos poblacionales ubicados en zonas lingüísticas con un porcentaje de 41-80% de valencianoparlantes fueron superiores a los grupos situados en zonas donde los valencianoparlantes son minoría (0-40%) en prácticamente casi todos los aspectos de dominio lingüístico valenciano: morfosintaxis, de producción, acceso léxico, discriminación y corrección.

c) Y los alumnos situados en zonas con alta presencia del valenciano (80-100% valencianoparlantes) presentaban los mejores resultados en tareas de morfosintaxis, producción y acceso léxico en valenciano, y también en tareas de acceso léxico y morfosintaxis en castellano.

(3) En relación a la interacción entre *modelo lingüístico* y *situación lingüística*, los datos obtenidos mostraron que:

a) Los grupos de enseñanza monolingüe y mantenimiento de las poblaciones donde los valencianoparlantes son mayoría (Monolingüe 80-100%, y Mantenimiento 80-100%), obtenían las mejores puntuaciones en tareas de morfosintaxis y acceso léxico tanto en castellano como valenciano. Sin embargo, las ventajas del programa monolingüe desaparecían al finalizar la escolarización elemental (Díez, 1995);

b) Respecto a la zona intermedia de valencianoparlantes (40-80% valencianoparlantes), el grupo de enriquecimiento obtuvo mejores puntuaciones en tareas de producción y corrección escrita en valenciano, comparado con los otros dos programas (Monolingüe y Mantenimiento). Asimismo, este grupo mostró las peores puntuaciones en tareas de corrección escrita en castellano, en comparación con el resto. Por su parte, los modelos de enriquecimiento y mantenimiento fueron superiores al monolingüe en la discriminación de palabras valencianas respecto a ciertos castellanismos;

c) Por último, la comparación del modelo de enriquecimiento situado en la zona de alta presencia del valenciano (80-100% valencianoparlantes) con el modelo monolingüe de la zona de baja valencianización (0-40%), mostró que el primero era superior al segundo en tareas cognitivas, morfosintaxis, producción, discriminación y corrección en lengua valenciana, mientras que el segundo respecto al primero era superior en tareas morfosintácticas, de discriminación y corrección en castellano.

En resumen, la evaluación de la competencia lingüística en castellano y valenciano en alumnos de 5º curso de Primaria muestra una clara dicotomía de resultados, donde el modelo de enriquecimiento obtiene las mejores puntuaciones en las pruebas de valenciano y el modelo monolingüe de enseñanza en castellano en la pruebas redactadas en castellano. Por su parte, los alumnos escolarizados en modelos de mantenimiento obtienen resultados intermedios tanto en valenciano como en castellano. Los autores de este estudio concluyen que la presencia curricular del valenciano en el sistema educativo valenciano no es perjudicial, y atribuyen a los pocos años de práctica del modelo de enriquecimiento la presencia de algunos resultados poco satisfactorios en cuanto a la competencia lingüística alcanzada en castellano por los alumnos de dicho modelo. En este sentido, los mismos autores de esta investigación señalan que hacen falta unos cinco años más de dedicación para obtener resultados más favorables en la evaluación de los modelos educativos bilingües implantados en la Comunidad Valenciana.

Para finalizar, el trabajo de Díez (1995) evaluó del conocimiento lingüístico en castellano y valenciano en alumnos a punto de finalizar su escolaridad obligatoria, teniendo en cuenta el modelo educativo cursado y la zona sociolingüística del centro. Se evaluaron los tres modelos educativos implantados en la Comunidad Valenciana. Y la aplicación de las pruebas CAL (versión en castellano) y VAL (versión en valenciano), permitió un análisis lectoescritor de ambas lenguas, abarcando desde niveles de morfosintaxis y comprensión hasta niveles de producción y discriminación de palabras. Los resultados conseguidos permitieron ofrecer las siguientes conclusiones:

(1) Respecto al *modelo lingüístico*, los programas de *enseñanza bilingüe* (2a y ab) demuestran que la presencia de la lengua minoritaria como lengua vehicular de enseñanza no produce ningún detrimento de la lengua castellana, sino todo lo contrario, el dominio lingüístico del castellano en estos dos modelos (2a y 2b) respecto al modelo monolingüe en castellano era superior en algunas habilidades específicas, aunque en la mayoría de ellas esta superioridad no era estadísticamente significativa. Por su parte, el *modelo monolingüe en castellano* apareció como un modelo muy pobre para conseguir niveles mínimos de competencia lingüística en valenciano. Estos últimos resultados están en consonancia con muchas investigaciones realizadas en otros países, las cuales señalan que la enseñanza de una lengua minoritaria como materia de estudio dentro del curriculum escolar no garantiza unos niveles aceptables de competencia en dicha lengua y, sobre todo, si ésta es una lengua que socialmente está infravalorada.

(2) En relación a la influencia del *contexto sociolingüístico*, los resultados mostraron que esta variable posee un peso fundamental en el aprendizaje de una u otra lengua, especialmente en los escolares de la comunidad donde la diversidad sociolingüística es su característica principal. Como ya se comentó, el ambiente sociolingüístico es una de las razones de la disimetría del sistema educativo valenciano. Los alumnos pertenecientes a zonas lingüísticas con alta presencia del valenciano empiezan muy pronto a estar en contacto con las dos lenguas e, incluso, llegan a la escuela con una superioridad en la lengua minoritaria. Por el contrario, los alumnos de las zonas castellanoparlantes con poco contacto

con el valenciano, llegan a la escuela con poca práctica del valenciano y si además el medio social no favorece su uso, su lengua habitual sigue siendo el castellano. Teniendo en consideración esta variable, los mejores resultados en ambas pruebas (VAL y CAL) se produjeron en las zonas con alta presencia del valenciano. Este hecho demuestra que el monolingüismo ambiental no favorece el castellano y, menos aún, los resultados en valenciano.

(3) En cuanto a la interacción entre la *situación sociolingüística* y el *modelo lingüístico*, los datos aportaron que los mejores resultados en ambos idiomas aparecían en programas bilingües (2a y 2b) situados en zonas de elevado número de valencianoparlantes (> 80% de valencianoparlantes). De este modo, el tratamiento escolar de la primera lengua y su uso en el medio social y familiar son la mejor combinación para producir alumnos con un bilingüismo equilibrado entre valenciano y castellano. El segundo grupo con mejores puntuaciones en ambas lenguas fueron los alumnos de programas monolingües en castellano situados en zonas lingüísticas con alta presencia del valenciano. En este caso, parece ser que el alumno no experimenta un cambio de lengua al incorporarse al programa de enseñanza monolingüe, debido a que antes de su incorporación ya ha adquirido algunas habilidades lingüísticas en castellano. Este hallazgo pone de relieve la enorme importancia que tienen los factores sociolingüísticos en la educación bilingüe valenciana, confirmando así que el bilingüismo que no se practica en la escuela está "compensado" por el contacto y la interacción constante de las dos lenguas en el ambiente.



Si comparamos estos resultados con los del estudio anterior (Bueno y Pascual, 1995), podemos observar que los efectos positivos de la escolarización en la primera lengua comienzan a manifestarse más claramente a medida que se avanza hacia una descontextualización del lenguaje, objetivo principal de todo programa de enseñanza una vez finalizada la escolarización elemental. Así pues, el aprovechamiento lingüístico de los alumnos valencianoparlantes de un programa bilingüe, sujetos que desarrollan las habilidades de competencia lingüística necesarias en su contexto ambiental, presentarán mejores patrones de rendimiento escolar que los alumnos valencianoparlantes de programas de enseñanza en castellano que se encuentran expuestos a un corte en el desarrollo evolutivo de su dominio lingüístico, ya que tienen que reestructurar su lenguaje comunicativo para adecuarlo al lenguaje exigido por la institución escolar. Los resultados conseguidos en estos estudios también señalan que el mejor rendimiento en el grupo de enseñanza bilingüe no afecta únicamente a la lengua valenciana sino también a la lengua castellana. Una posible explicación a estos resultados puede argumentarse desde la *hipótesis del umbral* basada en los efectos cognitivos y lingüísticos que las personas obtienen del bilingüismo temprano, y/o la *hipótesis de transferencia lingüística* de una lengua a otra, en este caso de la lengua valenciana a la castellana (Cummins 1981, 1984, 1986).

## **2.5. CONCLUSIONES**

En términos generales, el bilingüismo educativo valenciano no produce ninguna carencia en ninguna de las dos lenguas oficiales presentes en la Comunidad Autónoma Valenciana, sino todo lo contrario. En ocasiones, en

función de las características del modelo bilingüe implantado y el contexto sociolingüístico evaluado, se producen ganancias no sólo en valenciano sino también en algunas variables de dominio lingüístico castellano en relación con la media del conjunto de la población escolar monolingüe de su territorio. Así pues, *los niños y niñas valencianos escolarizados en un programa bilingüe dominarán el valenciano y el castellano igual o mejor que los niños y niñas escolarizados en un programa de enseñanza monolingüe*. Los resultados obtenidos en el sistema educativo valenciano se encuentran en sintonía con los resultados producidos en otras zonas bilingües del Estado Español con más tradición en la evaluación de la educación bilingüe, como es el caso de Cataluña (Bel y cols., 1993) y el País Vasco (Olaziregi y Sierra, 1993). Todos estos resultados permiten afirmar que, al final de la escolarización obligatoria, el bilingüismo educativo implantado en la Comunidad Valenciana lejos de aumentar las diferencias entre la competencia lingüística en castellano y en valenciano en los alumnos escolarizados en programas bilingües lo que ha provocado ha sido, como mínimo, eliminarlas e incluso, en ocasiones, llegan a ser superiores cuando se comparan con alumnos escolarizados en programas monolingües en castellano, especialmente para aquellos alumnos con un mayor grado de bilingüismo adquirido.

Este hecho podría explicarse mediante algunas de las hipótesis ampliamente aceptadas en los estudios científicos sobre bilingüismo: las hipótesis de la "interdependencia lingüística", los "umbrales mínimo y máximo" de Cummins o el concepto de "bilingüismo aditivo" de Lambert. Según estos autores, en las situaciones de contacto de lenguas, como es el caso de la Comunidad Valenciana, un prerrequisito para obtener un nivel alto de competencia bilingüe es mantener las habilidades lingüísticas en la primera

lengua para el caso de sujetos nativos lo cual se consigue con la incorporación de la lengua familiar como lengua de enseñanza en la escuela, o adiestrar las habilidades de la lengua minoritaria en el caso de sujetos con lengua familiar mayoritaria. De esta forma, la utilización vehicular del valenciano en los programas bilingües, al tratar de desarrollar el nivel alcanzado en la primera lengua, produce una especie de transferencia hacia la otra lengua del sistema escolar, aportando también resultados satisfactorios en esta otra lengua. Este hecho ha llevado a los investigadores a pensar en la existencia de una *interdependencia lingüística* en ambas lenguas (Cummins, 1986) que hace posible la transferencia de habilidades cognitivas y académicas de una lengua a la otra, sobre todo en dos lenguas con un corta distancia interlingüística, como son el castellano y el valenciano.

Sin embargo, es importante resaltar que los beneficios lingüístico-académicos derivados de los programas de enseñanza bilingüe son percibidos con mayor claridad al finalizar la enseñanza elemental (Díez, 1995), y no a lo largo del proceso de escolarización del bilingüe (Bueno, 1995). Estos resultados coinciden con algunos estudios realizados (Baker, 1993; Boada, 1986; Cummins, 1984a, 1984b; Genesee, 1985; Lambert y Tucker, 1972, Sanchez y Tembleque, 1986; Swain y Lapkin, 1982) que señalan que un programa bilingüe ha de tener una duración mínima de cuatro a seis años para obtener resultados satisfactorios en rendimiento en ambas lenguas. Además, los últimos estudios (Espí, 1990) muestran que el lenguaje oral continúa desarrollándose hasta los once o trece años (1º y 2º curso de E.S.O.). Por tanto, un programa bilingüe que se aplique durante un tiempo menor del señalado no puede consolidarse como un programa bilingüe eficaz. Dado que

los alumnos que se encuentran en cursos intermedios de su proceso de escolarización están menos expuestos a la presencia y aprendizaje de la lengua mayoritaria, el castellano, es posible que sus habilidades lectoescritoras en dicha lengua no se desarrollen y consoliden suficientemente hasta llegar a los últimos cursos de la enseñanza obligatoria, lo cual puede provocar y así lo demuestran algunos estudios realizados (Bueno, 1995) que algunos de estos alumnos se encuentren con ciertas dificultades o handicaps lingüísticos al ser evaluados en su conocimiento lingüístico castellano, y por tanto, esto derive en resultados poco satisfactorios en rendimiento académico cuando se utiliza el castellano como lengua de evaluación en su proceso de escolarización.

Por otra parte, los resultados obtenidos también señalan que la presencia de las lenguas en el ámbito social es un factor muy importante en la obtención de una competencia lingüística equilibrada en ambas lenguas. Dentro de este ámbito, la variable que recibe mayor peso es la lengua hablada por el alumno en el ámbito familiar, escolar y coloquial con los amigos más próximos. Precisamente, la importancia de la presencia y uso social del valenciano ha demostrado ser en algunas ocasiones (Diez, 1995) superior a la influencia que puede ejercer la propia institución escolar en el aprendizaje de ambas lenguas. Por lo general, los resultados aportados en la evaluación de un programa bilingüe al final de la enseñanza obligatoria señalan que la interrelación entre el uso social de la lengua con su enseñanza en el ámbito académico posee un peso fundamental en el dominio lingüístico conseguido en la población escolar valenciana.



## ***CAPITULO 3***

**Evaluación del funcionamiento diferencial en items**



### **3.1. INTRODUCCION**

En este tercer capítulo, nuestro objetivo se centra en el problema del funcionamiento diferencial de los items en la medida. El sesgo en los items, y por extensión en los tests, es un tema de estudio que aparece con cierta profusión a partir de los años setenta en la literatura psicométrica especializada. Precisamente, la aparición de los nuevos modelos de Teoría de Respuesta al Ítem (TRI) ofrecen la posibilidad de un marco teórico más adecuado para el estudio del sesgo que la Teoría Clásica de los Tests (TCT). No obstante, a pesar de su aparición tardía su implicación en el campo de la medición y la evaluación de los tests psicológicos ha originado una gran controversia social, especialmente cuando el problema del sesgo ha sido detectado con frecuencia en situaciones de contacto entre lenguas, donde coexisten dos o más lenguas tanto a nivel familiar, social, cultural, etc.. y, además, se imparten programas de enseñanza bilingüe.



Con el fin de aportar una visión global de la problemática del sesgo, primero se presentarán los orígenes históricos de la aparición y posterior estudio del tema del sesgo, término que como veremos posteriormente es reemplazado por el término de *Funcionamiento Diferencial de los Items* (*Differential Item Functioning*, DIF, en adelante). Segundo, se dedicará una pequeña sección del capítulo a establecer la diferencia que existe entre los conceptos de funcionamiento diferencial del ítem (DIF) e impacto. A continuación, se expondrá una revisión crítica de los distintos métodos estadísticos empleados en la detección del DIF, y para finalizar se ofrecerá una revisión de la literatura empírica centrada en el estudio del funcionamiento diferencial de los ítems de un test en el campo de la medición y evaluación psicológica. En concreto, esta última se abordará con especial interés en los contextos de naturaleza bilingüe.

### 3.2. ANTECEDENTES HISTORICOS.

Los primeros estudios centrados en el sesgo de los ítems comienzan en 1910 con la investigación de Alfred Binet (Binet y Simon, 1916). El mencionado autor estudiaba a niños con bajo estatus socioeconómico, y observó que algunos ítems de los tests, además de la capacidad mental del individuo, podían estar midiendo los efectos de su formación cultural, desarrollada y adquirida tanto en el ámbito familiar como escolar. La aparición de estos resultados llevaron a este autor a la eliminación de algunos ítems de los tests empleados en sus investigaciones. En 1912, William Stern investigó la presencia e influencia de las diferencias de clase en la evaluación de la inteligencia en sujetos alemanes. Y sus resultados, al igual que a Binet, mostraron la existencia de diferencias

significativas en el coeficiente intelectual (CI) en función de las clases sociales evaluadas, lo que llevó a afirmar que los tests de inteligencia favorecían a una determinada clase social sobre otra (Stern, 1914).

Sin embargo, el comienzo de la investigación moderna sobre el sesgo de los tests se sitúa en un estudio realizado por Eells, Davis, Havighurst, Herrick y Tyler (1951). En sus investigaciones, estos expertos son los primeros que se plantean la posibilidad de que algunas de las diferencias encontradas en puntuaciones sobre inteligencia en diferentes grupos socioeconómicos, culturales o étnicos, pueden ser una consecuencia directa del contenido específico de los items en dichos tests y no de diferencias reales en los sujetos en la habilidad medida por el test. En este sentido, los citados autores analizaron más de 650 items de 8 tests de inteligencia diferentes, intentando detectar aquellos items que por su familiaridad, bien por su aspecto formal o de contenido, para ciertos grupos con un determinado nivel socioeconómico pudieran estar reflejando, en lugar de la variable que se pretendía medir, la presencia de distintas oportunidades educativas en los grupos sometidos a comparación. En aquel momento, a este tipo de sesgo se le denominó *sesgo cultural*.

Durante este mismo periodo, la controversia en torno al sesgo de los tests se agudiza con un artículo publicado por Jensen (1969), en el que se defendía la herencia de la inteligencia. Su tesis afirma que el coeficiente intelectual (CI) está determinado más por la herencia que el ambiente. Por tanto, se asume que las diferencias existentes entre blancos y negros respecto al CI son de origen genético, debido a que las diferencias

encontradas entre ambos grupos son demasiado grandes para poder atribuir las sólo al contexto social del individuo. Resultan bastante evidentes las importantes repercusiones sociales, educativas y, sobre todo, a nivel político de este tipo de afirmaciones.

Pero el auténtico impulso para este campo surge en la década de los 60 y 70, con la aparición del movimiento de los derechos humanos en EEUU, cuyo máximo interés se focaliza en el desarrollo de nuevos programas educativos de carácter multicultural con el fin de asegurar la igualdad de oportunidades a toda la población americana. Precisamente, este tipo de actuaciones provocan el origen de los primeros programas de *Educación Compensatoria*, dirigidos principalmente a las minorías lingüísticas y/o culturales residentes en Estados Unidos. Sin embargo, las primeras evaluaciones de estos programas no mostraron los resultados esperados, debido a un aumento del fracaso escolar de los alumnos pertenecientes a dichos grupos minoritarios.

Esta situación llevó a la creación de nuevas iniciativas en la educación y enseñanza impartida a las minorías tanto en Norteamérica como en Europa. Respecto al tema del sesgo, en aquellos momentos éste se focalizó con cierta exclusividad en los tests de inteligencia, los cuales eran utilizados habitualmente como principal fuente de información en la toma de decisiones educativas y laborales, por ejemplo, para la admisión en la enseñanza superior, para el logro de nuevas oportunidades de empleo o promoción en el empleo, etc. En concreto, los investigadores centrados en el estudio de la inteligencia en los diferentes grupos minoritarios lingüísticos y/o culturales defienden que la principal causa de las

diferencias encontradas en el desarrollo cognitivo de estos sujetos es el *sesgo cultural*, procedente de la aplicación de cuestionarios o instrumentos de medida originariamente elaborados y baremados para grupos con lengua y cultura mayoritaria. Así pues, a partir de ese momento, los tests dejan de ser vistos como instrumentos de medida neutrales. La posibilidad de que éstos presenten algún sesgo que provoque una desigualdad de oportunidades en la sociedad, recibe una amplia atención por parte del público y de los profesionales de la medida, en particular. Estos empiezan a cuestionarse hasta qué punto el distinto bagaje cultural, social, económico y lingüístico que se encuentra presente en los sujetos de los grupos sometidos a comparación afecta a las puntuaciones obtenidas en los tests de rendimiento y aptitud aplicados a dichos grupos minoritarios y, por tanto, las diferencias encontradas reflejan o bien diferencias reales en la variable medida por el test o son simplemente un artefacto de los tests utilizados, debido principalmente a que éstos son construidos por la clase económica y políticamente dominante.

Sin embargo, en las dos últimas décadas esta situación todavía se ha agravado más a causa del importante incremento de estudiantes pertenecientes a otras culturas en los colegios americanos y europeos; estudiantes cuya primera lengua no es la lengua mayoritaria. Dentro del ámbito educativo, algunos estudios han señalado que un gran número de estos sujetos pertenecientes a minorías lingüísticas y/o culturales presentan ciertas dificultades académicas en comparación con los alumnos del grupo mayoritario, cuya primera lengua es el inglés (Spence, Mishra y Ghozeil, 1971). En éstos, una de las variables que resultó más relevante para la

aparición de estas dificultades fue la falta de una competencia lingüística inglesa apropiada. Por su parte, otros estudios (Hernández, 1973; Solomon y Orenstein, 1991) demostraron que los estudiantes que poseían el inglés como segunda lengua presentaban ciertas desventajas frente a sus homólogos monolingües, especialmente cuando el material de lectura, las instrucciones y/o los exámenes a realizar estaban elaborados en dicha lengua. En concreto, los escolares de los grupos minoritarios necesitaban más tiempo y realizaban mayor esfuerzo en la decodificación e interpretación del material redactado en inglés que los monolingües ingleses, los cuales solían centrar toda su atención en la comprensión e integración de conceptos.

Dentro de este ámbito, el término *sesgo* (bias) se equiparaba al concepto frecuentemente utilizado de *injusticia* (unfair). Sin embargo, en un intento de clarificar conceptos y situar cada cosa en su sitio, Jensen (1980) en su libro '*Bias in mental testing*' aporta algunas definiciones bastante precisas de lo que se entiende por sesgo, intentando desligarlo del concepto de injusticia. Para Jensen (1980) el *sesgo* es un término referido a cómo un instrumento de medida o test específico debe ser correctamente utilizado, y no limitado al "sesgo cultural", mientras que los conceptos de *injusticia*, *justicia social* e "*igual protección de la ley*" son ideas morales, legales y filosóficas, y como tales deben ser valoradas en esos términos (pp. 376). A su vez, el presente autor en su libro contribuye con otra definición aportando la siguiente distinción:

"En estadística matemática, el "sesgo" se refiere a una sobre o infraestimación sistemática de un parámetro poblacional por un

estadístico basado en muestras extraídas de la población. En *psicometría*, el "sesgo" se refiere a errores sistemáticos en la validez predictiva o en la validez de constructo de las puntuaciones en el test de individuos pertenecientes a grupos diferentes. Por tanto, el sesgo psicométrico es un serie de atributos estadísticos aplicados conjuntamente a un test determinado y dos o más subpoblaciones específicas" (Jensen, 1980; pp. 375).

Sin embargo, a pesar de todas estas aclaraciones las connotaciones sociales ligadas al término sesgo continuaban presentes en la gran mayoría de los temas de debate. De esta manera, Angoff (1982, 1993) y Camilli (1993) y Camilli y Shepard (1994) insisten en sustituir el término análisis del sesgo de aquellas situaciones que sólo intentan dilucidar las propiedades estadísticas de los items de un test aplicado a distintos grupos sin proporcionar valoraciones acerca de la existencia de sesgo. Pero no hasta 1988 con la expresión **Funcionamiento Diferencial del ítem** (*Differential Item Functioning*, DIF, a partir de ahora) propuesta por Holland y Thayer, cuando finalmente es reemplazado el término sesgo de los items. Este último término engloba a los diferentes procedimientos estadísticos para la detección de un posible funcionamiento diferencial y no es sinónimo de sesgo. Por tanto, los estadísticos de DIF por sí solo no ofrecerán ningún tipo de información acerca de por qué los grupos se comportan de forma diferente con respecto a un ítem determinado, mientras que un estudio sobre el sesgo lo que ofrecerá será toda explicación posible acerca del por qué de este comportamiento diferencial de los items en el contexto de la validez de constructo de los items, determinando cuáles de

éstos están sesgados para eliminarlos del test. En otras palabras, los métodos DIF serán procedimientos estadísticos para detectar el funcionamiento diferencial en el ítem y el análisis del sesgo se enmarcará en el contexto más general de la validez del test, utilizando los resultados obtenidos en la aplicación de los primeros.

En definitiva, afirmar que un ítem de un determinado test presenta *Funcionamiento Diferencial del Ítem (DIF)* significará que el citado ítem presenta propiedades estadísticas diferentes en distintos grupos, cuando en realidad éstos no poseen diferencias reales en la habilidad medida por el test, y si es el test, en su conjunto, el que presenta propiedades estadísticas distintas en cada grupo, entonces se dirá que existe *Funcionamiento Diferencial del Test (FDT)*. Para una mejor comprensión de este término, a continuación vamos a ofrecer la diferencia que existe entre los conceptos funcionamiento diferencial del ítem e impacto.

### **3.3. FUNCIONAMIENTO DIFERENCIAL DEL ÍTEM (DIF) E IMPACTO.**

La psicometría se ha ocupado del DIF tal y como éste ha sido definido por los diferentes psicómetras expertos en el área. Así pues, atendiendo a las distintas definiciones presentadas en la literatura especializada (Hambleton y Swaminathan, 1985; Hambleton, Swaminathan y Rogers, 1991; Holland y Thayer, 1988; Lord, 1980; Mellenberg, 1982), podemos afirmar que *un ítem o un test presentará DIF si al ser aplicado a miembros pertenecientes a distintos grupos, definidos por sus características sociodemográficas como son su propia lengua, cultura, raza, etc., pero con igual nivel de competencia o habilidad en la variable medida por el test,*

poseen diferentes probabilidades de acertar o superar con éxito el ítem o test en cuestión.

Por ejemplo, si un test que mide la competencia matemática está formulado de tal modo que exige un alto nivel de comprensión verbal y, además, es aplicado en su forma estándar a individuos con diferentes propiedades lingüísticas en cuanto a su lengua familiar y/o académica, con cierta seguridad el citado test estará sesgado contra aquellos lectores menos eficientes. En este caso, existirá una confusión entre el efecto de la comprensión verbal y el de la competencia matemática del sujeto evaluado, por lo que si un sujeto puntúa bajo en el test no se sabrá con certeza si este resultado debe atribuirse a su bajo rendimiento en matemáticas o a que su competencia verbal se encuentra limitada por el hecho poseer una lengua familiar y/o escolar diferente a la del resto del grupo normativo, y por consiguiente no ha llegado a entender correctamente los problemas planteados en el test administrado. De este modo, en algunas ocasiones, en sentido contrario a lo esperado, algunos ítems de un test pueden mostrar un funcionamiento diferencial entre los dos grupos sometidos a comparación; por tanto, la presencia de *DIF* en un test implica que un grupo determinado posee una "injusta" ventaja sobre otro simplemente por poseer algunas características sociodemográficas distintas.

Sin embargo, muchas veces se recomienda cierta prudencia a la hora de interpretar las diferencias halladas en un determinado test tras aplicarlo a poblaciones diferentes, dado que la posibilidad de que distintos grupos o poblaciones puedan presentar diferencias entre ellos, en sus respuestas



ante un ítem, puede ser debido a razones diferentes a las del funcionamiento diferencial (DIF). Por ejemplo, la presencia de diferencias en los resultados de un test de rendimiento académico aplicado a grupos distintos, ¿son resultado de diferencias reales en la habilidad medida por el test, o son producto de un funcionamiento diferencial de los ítems que forman parte del test?, es decir, ¿existe Impacto o DIF?

La necesidad de clarificar estos dos términos nos obliga a presentar una distinción entre ambos conceptos. Siguiendo la definición de Ackerman (1992), existirá *impacto* cuando una diferencia real entre los distintos grupos sometidos a comparación en la variable medida por el test viene causada por una diferencia real de la actuación en el test en dichos grupos. Esto es, si un ítem presenta *impacto*, la probabilidad de responderlo correctamente será mayor para un grupo que para otro, reflejando de esta manera las diferencias entre ambos grupos en la habilidad medida, y la probabilidad de responder correctamente a ese ítem será la misma para sujetos con el mismo nivel de habilidad con independencia del grupo al que pertenezcan. Por el contrario, habrá *DIF* si algunos de los ítems del test miden de forma diferencial a los sujetos de distintos grupos de tal manera que para sujetos con idéntico nivel de habilidad en la variable medida, la probabilidad de responder correctamente a un ítem difiere, no es la misma, sino que depende del grupo al que se pertenece.

De este modo, si ambos conceptos, DIF e IMPACTO son dos cosas distintas, el primer requisito de los métodos y técnicas estadísticas empleadas para detectar correctamente el DIF en un determinado test será

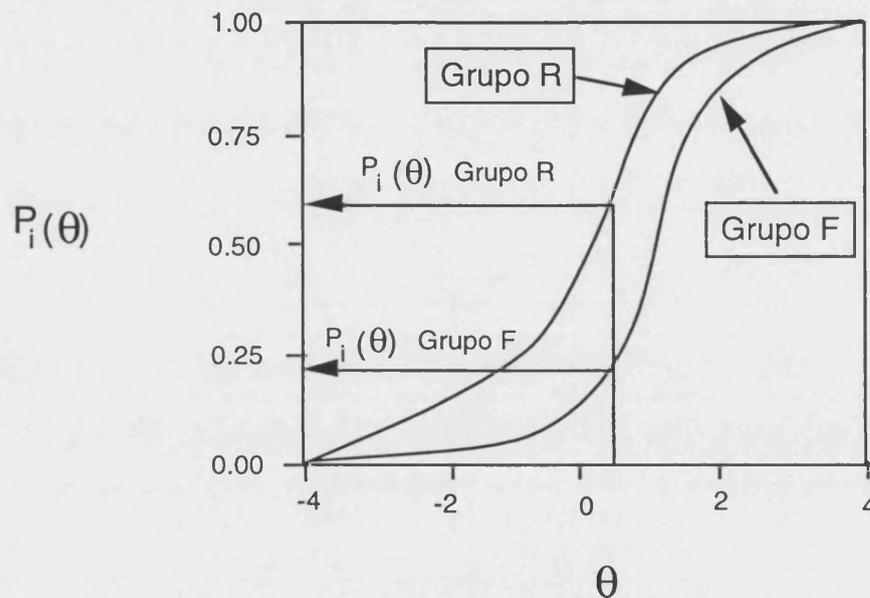
distinguir claramente las diferencias reales entre grupos (IMPACTO) de las artificiales (DIF), esto es, averiguar si realmente las diferencias encontradas en la habilidad en grupos distintos son diferencias reales debidas a la habilidad medida por el test o son simplemente un artefacto implícito en el proceso de medición y evaluación mediante los tests psicológicos.

En los estudios basados en la detección del DIF, normalmente se suele clasificar a los miembros objeto de interés en dos grupos diferentes: el grupo focal (F) y el grupo de referencia (R). Se suele denominar grupo focal al grupo objeto de análisis, casi siempre un grupo minoritario. Mientras que el grupo de referencia es el grupo que sirve como criterio de comparación y que casi siempre es un grupo mayoritario. Por su parte, la variable de agrupamiento de los individuos sometidos a comparación puede ser cualquier variable sociodemográfica (sexo, edad, nivel educativo, raza, lengua materna...) en la que se intuya que las propiedades psicométricas de los items o el test pueden variar.

Dentro de este ámbito, podemos distinguir entre *DIF uniforme* y *DIF no uniforme*. Los citados términos fueron acuñados por Mellenbergh (1982), considerando que existe *DIF uniforme* cuando no existe interacción entre el nivel de aptitud o habilidad y la pertenencia a un grupo, es decir, cuando la probabilidad de responder correctamente a un ítem es uniformemente superior en un grupo respecto a otro a lo largo de los diversos niveles de habilidad considerados. El *DIF no uniforme* aparece cuando sí existe interacción entre nivel de aptitud y pertenencia al grupo, es decir, cuando

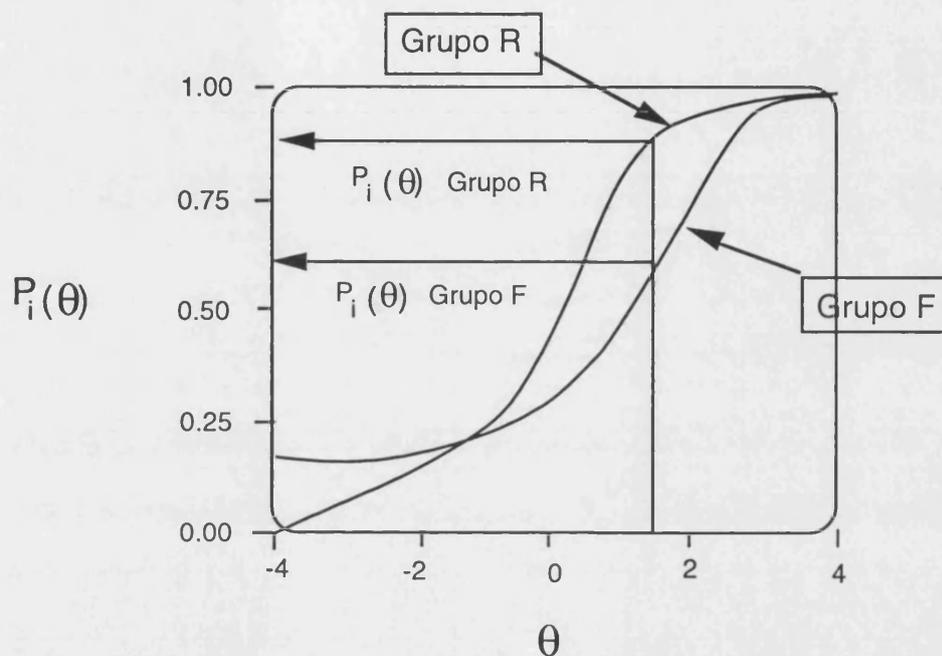
la probabilidad relativa de responder correctamente a un ítem varía a través de los niveles de habilidad o aptitud considerados. Para una mejor apreciación de estos tipos de DIF, en las Figuras 1a y 1b podemos apreciar una representación gráfica de los mismos.

**Figura 1a.** DIF uniforme.



A su vez, para el último caso Rogers y Swaminathan (1993) distinguen entre DIF no uniforme y no uniforme mixto. El *DIF no uniforme propiamente dicho*, ocurre en aquellos casos en que las diferencias en probabilidad entre los dos grupos sometidos a análisis se cancelan, esto es, cuando el parámetro  $b$  se mantiene constante y el parámetro  $a$  varía para ambos grupos ( $a_R \neq a_F$ ), y el *DIF mixto* cuando las mencionadas diferencias no se anulan, es decir, cuando los parámetros  $a$  y  $b$  en ambos grupos son desiguales ( $a_R \neq a_F$  y  $b_R \neq b_F$ ).

Figura 1b. DIF no uniforme.



Como hemos señalado anteriormente, el tema del estudio del DIF dirigido a la identificación de aquellos items que provocan una ejecución diferencial en personas con la misma habilidad pero pertenecientes a grupos diferentes, ha sido objeto de interés en los últimos años, convirtiéndose en una importante área de estudio dentro del campo de la aplicación de las pruebas psicométricas. Sin embargo, la masiva aparición de artículos y trabajos relativos a este área de investigación ha puesto prácticamente todo su énfasis en investigar qué métodos o procedimientos estadísticos son los más apropiados para evaluar el DIF de los items, quedando relegada en un segundo plano la investigación empírica aplicada de los mismos. Por lo tanto, dada la importancia de estos primeros estudios para un estudio adecuado del DIF en los items de un determinado test, en el próximo

apartado se ofrecerá una revisión crítica de los principales métodos estadísticos para el estudio del DIF y, con posterioridad, una revisión de la literatura empírica centrada en el estudio del DIF en algunos tests aplicados en contextos bilingües.

#### **3.4. REVISION CRITICA DE MÉTODOS ESTADISTICOS PARA LA DETECCION DEL DIF**

Probablemente el método más eficaz para evitar en lo posible la presencia de sesgo en los items de un test sea un riguroso análisis de su contenido por parte de varios profesionales y/o expertos en el tema objeto de estudio. Sin embargo, la propia experiencia muestra que una vez hecha tal revisión y posterior aplicación de los items del test a una muestra de sujetos, todavía son necesarios algunos análisis estadísticos adicionales que permitan detectar el DIF de aquellos items que, posiblemente, en la etapa anterior no fue detectado. A este tipo de procedimientos o métodos estadísticos aplicados a posteriori son los que vamos a presentar aquí. Es importante señalar que la aplicación de estas técnicas únicamente detectan si un ítem funciona de forma diferencial en dos grupos previamente establecidos, pero en ningún caso ofrece información acerca de si la cantidad o magnitud de DIF encontrado en un determinado ítem es aceptable o no para el constructo objeto de medición. Por tanto, la utilización de estas técnicas estadísticas son sólo una práctica añadida dentro de todo el proceso de estudio y comprobación del sesgo.

En sus comienzos, las *aproximaciones tradicionales* utilizadas para identificar el DIF de los tests se situaron en la comparación de items a

través de la replicación de varias muestras utilizando la Teoría Clásica de Tests (TCT) (Gulliksen, 1950; Lord y Novick, 1968). Concretamente, los primeros métodos propuestos se centraron en el análisis de varianza (ANOVA) (Cleary y Hilton, 1968; Plake y Hoover, 1979) o el método delta (Angoff y Ford, 1973; Angoff, 1982); sin embargo, éstos no poseían un control del nivel de habilidad de los grupos sometidos a análisis, y en consecuencia, existía cierta confusión entre la presencia de impacto y DIF (Camilli y Shepard, 1994; Jensen, 1980; Osterlind, 1983; Schmeiser, 1982; Shepard, Camilli y Averill, 1981). Estos métodos así como algunos métodos basados en el estadístico  $\chi^2$  tradicional por ejemplo el método de  $\chi^2$  de los aciertos (Scheuneman, 1979) y el método de  $\chi^2$  total (Camilli, 1979), no serán tratados aquí por haber sido superados por otras aproximaciones más eficaces en la detección del DIF. No obstante, los lectores interesados en los métodos delta y los métodos basados en el estadístico  $\chi^2$  pueden encontrar una descripción de los mismos en los trabajos de Martínez-Arias (1994) y Muñiz (1990, 1992).

Precisamente, las deficiencias encontradas en la TCT, basadas principalmente en su dependencia a la población de estudio, son la principal causa de la aparición de la *Teoría de Respuesta al Item (TRI)*, la cual es capaz de proporcionar un marco teórico más adecuado para el estudio y la detección del DIF. De hecho, la mayor parte de las investigaciones realizadas sobre el estudio y detección del DIF se han centrado en los métodos basados en la teoría de respuesta a los items (TRI) y en el estadístico  $\chi^2$  (Cole y Moss, 1989). Su principal característica, a diferencia de las primeras aproximaciones, es el estudio del DIF desde una

**perspectiva condicional**, es decir, el estudio del funcionamiento diferencial del ítem a partir de una estimación de las diferencias en la dificultad del ítem en distintos grupos pero con el mismo nivel de habilidad en la variable medida por el test.

Atendiendo a la gran variedad de técnicas y métodos estadísticos elaborados para la detección y evaluación del DIF en ítems de un determinado test, se puede apreciar que los profesionales de este campo han realizado varios intentos en el establecimiento de una clasificación de los diferentes métodos desarrollados. En este sentido, la literatura especializada ofrece las siguientes taxonomías: métodos no paramétricos versus paramétricos (Bock, 1993); puntuación observada versus puntuación verdadera (Holland y Wainer, 1993); ó modelos de invarianza condicional observada (OCI) versus modelos de invarianza condicional no observada (UCI) (Millsap y Everson, 1993). Básicamente, bajo estas clasificaciones se esconde la distinción entre aquellos métodos basados en la TRI (paramétricos, puntuación verdadera y modelos de invarianza condicional no observada) versus métodos que son una aproximación a los métodos basados en la TRI, pero no especifican ningún modelo de medida (no paramétricos, puntuación observada y modelos de invarianza condicional observada). Una revisión metodológica en castellano de los principales métodos estadísticos utilizados en la evaluación del funcionamiento diferencial en ítems dicotómicos puede encontrarse en Fidalgo (1996) y Gómez y Hidalgo (1997).

La clasificación sobre los métodos de detección del DIF presentada en este trabajo de investigación se va a establecer, siguiendo la taxonomía propuesta por Millsap y Everson (1993), en torno a dos criterios: 1) el

modelo de medida que subyace al procedimiento, y dentro de este 2) el tipo de respuesta al ítem (dicotómica, politómica, continua).

Atendiendo al primero de ellos, tenemos en *primer lugar* aquellos métodos que evalúan el DIF con un modelo de medida que relaciona las respuestas al ítem con la variable latente que pretende medir el test. En éstos, el DIF se evalúa comprobando si los parámetros que definen el modelo permanecen invariantes a través de los distintos grupos con el mismo nivel de habilidad. Los modelos presentados dentro de la Teoría de Respuesta al Ítem (TRI) pueden variar en función de la naturaleza de la variable latente: unidimensional o *multidimensional*, y la respuesta al ítem: *dicotómica* (por ejemplo, modelos 1PL 2PL y 3PL) o *politómica* (Samejima, 1969; Thissen y Steinberg, 1984). Dentro de esta categoría, se han propuesto los métodos de comparación de parámetros basados en el estadístico de Lord (Lord, 1980), las medidas de área (Cohen, Kim y Baker, 1993; Kim y Cohen, 1991; Raju, 1988, 1990) y los métodos basados en la comparación de modelos (Thissen, Steinberg y Wainer, 1988, 1993).

En *segundo lugar*, tenemos aquellos procedimientos en los que para evaluar el DIF no es necesario especificar un modelo de medida que relacione la variable medida por el test con la puntuación obtenida en el ítem. En este caso, la puntuación total en el test es el criterio que se va a utilizar para obtener un estimador de la variable que pretende medir el test. El DIF se evalúa verificando si las distribuciones de puntuaciones del ítem entre sujetos con valores iguales en la puntuación total del test son independientes del grupo de pertenencia. Al igual que en el caso



anterior, los dos tipos de formato ampliamente utilizados en los tests, *dicotómico* y *politómico*, son considerados. Dentro de esta categoría, se podrían incluir el estadístico Mantel-Haenszel (MH; Holland y Thayer, 1988); el método de estandarización (Dorans, 1989; Dorans y Kulick, 1986), los modelos loglineales, logit y de clase latente (Fidalgo, 1996; Fidalgo y Mellenbergh, 1995; Fidalgo, Muñiz y Mellenbergh, 1997; Mellenbergh, 1982), los análisis de regresión logística (Swaminathan y Rogers; 1990; Rogers y Swaminathan, 1993) y los más recientes análisis de función logística discriminante (Miller, Spray y Wilson, 1992) entre otros. Recientemente, Shealy y Stout (1993a, 1993b) han elaborado un nuevo método de detección de DIF denominado SIBTEST, que "... es un paso intermedio entre los métodos OCI y UCI" (Millsap y Everson, 1993, p. 324). En concreto, el SIBTEST posee características de ambos métodos.

Sin embargo, cuando la respuesta es *continua* la detección del DIF se ha llevado a cabo mediante la formulación de modelos factoriales. En este caso, los métodos factoriales elaborados para evaluar el DIF se ocuparán de probar la invarianza factorial a través de los distintos grupos sometidos a comparación. La aplicación de esta metodología al problema de la detección del DIF podemos encontrarla en Everson, Millsap y Rodríguez (1991); Millsap y Everson (1991), Oort (1992) y Gómez (1996).

Sin embargo, antes de proceder a la presentación de cada uno de los métodos propuestos en la detección del DIF, es conveniente destacar un problema común a todos ellos: se trata del *problema de la circularidad*, aspecto fundamental que se manifiesta, en mayor o menor medida, en todos los métodos estadísticos expuestos a continuación.

Resulta bastante evidente que para evaluar el DIF de forma adecuada necesitamos tener medidas insesgadas del nivel de competencia de los miembros examinados. Sin embargo si se da el caso de que algunos de los items del test presentan DIF entonces se estará utilizando una medida sesgada de la habilidad para estudiar el posible funcionamiento diferencial de los items. Por tanto, los métodos estadísticos empleados para evaluar el DIF estarán utilizando medidas sesgadas del nivel de competencia de los sujetos examinados. Para evitar este problema, varios son los procedimientos que han sido propuestos, centrados fundamentalmente en purificar la medida de la habilidad utilizada como criterio para formar los distintos grupos de comparación.

Dentro de la aproximación de la Teoría de Respuesta a los items, Marco (1977) y Lord (1980) fueron los primeros en presentar un procedimiento de purificación multietápico de la medida de la habilidad con el fin de obtener una estimación no contaminada del nivel de competencia de los sujetos, sin embargo éste nunca se utilizó en la práctica. Más tarde, Park (1988) y Park y Lautenschlager (1990) proponen una modificación del mencionado método, pero su principal inconveniente, que comparte con el procedimiento de Lord (1980), fue su excesiva complejidad y costo computacional. Una manera de proceder más elemental y eficaz consistió en el empleo de procedimientos iterativos, centrados en la equiparación de las estimaciones obtenidas en la habilidad de los distintos grupos objeto de estudio (Candell y Drasgow, 1988; Candell y Hulin, 1987; Drasgow, 1987; Hulin y Mayer, 1986; Kim y Cohen, 1992; Park y Lautenschlager, 1990;

Segall, 1983). Básicamente, el método propuesto consistía en realizar una equiparación de las estimaciones obtenidas en la habilidad de ambos grupos, eliminando de la ecuación de equivalencia los items que mostraban DIF en la iteración anterior y recalculando para cada nueva iteración los índices de DIF.

Fuera del marco teórico de la Teoría de Respuesta al Item, también se han propuesto algunos métodos que han permitido ofrecer soluciones al problema de la circularidad. Estos son: (1) el método logit iterativo (Fidalgo, 1996; Van der Flier, Mellenbergh, Ader y Wijn, 1984; Kok y cols, 1985), (2) la versión bietápica del estadístico Mantel-Haenszel (Clauser y cols, 1993; Holland y Thayer, 1988) y (3) la purificación simultánea y bietápica del criterio en la regresión logística (Gómez y Navas, 1995; Navas y Gómez, 1994; Rogers y Swaminathan, 1983). La lógica que subyace a todos ellos es muy similar a la de los procedimientos desarrollado en la TRI. Su método consiste en una primera etapa evaluar el DIF en la forma estándar, y a continuación volver a evaluar el DIF, pero utilizando para estimar el nivel de habilidad o competencia de los sujetos examinados sólo aquellos items que no presentaron DIF en el análisis inicial. Todos los estudios de simulación más recientes coinciden en señalar un aumento de las clasificaciones correctas del DIF cuando se utilizan los procedimientos iterativos frente a la evaluación del DIF mediante un sólo análisis (Fidalgo, 1996; Fidalgo y Mellenbergh, 1995; Fidalgo, Muñiz y Mellenbergh, 1997; Fidalgo y Paz, 1995a; Fidalgo y Paz, 1995b; Gómez y Navas, 1996; Hidalgo, 1995; Kok, Mellenbergh y Van der Flier, 1985; Lautenschlager, Flaherty y Park, 1994; Lopez Pina, Hidalgo y Sánchez Meca, 1997; Miller y Oshima, 1992). Por lo tanto, la conclusión a la que se llega es la

siguiente: *la aplicación de un método para detectar la presencia de DIF, sea cual sea, siempre debe ir unido a la aplicación de un procedimiento de purificación.*

A continuación, siguiendo la taxonomía propuesta por Millsap y Everson (1993) vamos a exponer, en primer lugar, los principales métodos estadísticos desarrollados dentro de la Teoría de Respuesta a los Items (TRI), y en segundo lugar, los procedimientos que permiten evaluar el DIF sin necesidad de especificar un modelo de medida previo.

#### **3.4.1. Procedimientos basados en la Teoría de Respuesta a los Items (TRI)**

En general, los modelos basados en la Teoría de Respuesta al Item (TRI) se han propuesto para formatos de respuesta *dicotómica* o *dicotomizada*, siendo *unidimensional* la naturaleza de la variable latente. Esta teoría establece un modelo de medida que relaciona directamente la probabilidad de una puntuación observada (por ej. la respuesta al ítem) con el rasgo o constructo que pretende medir el test. Además, se basa en el supuesto de que la probabilidad de respuesta a un ítem es una función del nivel de competencia adquirida en el rasgo o variable medida por el test. Por ejemplo, la probabilidad de responder un ítem correctamente en una prueba de inteligencia es una función de la competencia alcanzada por dicho sujeto en el rasgo medido por dicho test, en este caso, *lainteligencia*.

La relación funcional entre la variable medida y la probabilidad de dar una determinada respuesta puede ser representada mediante la curva característica del ítem (CCI), la cual viene determinada por el modelo de medida seleccionado. Por ejemplo, el modelo de 3 parámetros (3 PL; Birnbaum, 1968) viene dado por la siguiente expresión:

$$P_i(\theta) = \frac{c_i + (1 - c_i)}{1 + \exp[-Da_i(\theta - b_i)]} \quad (3.1)$$

donde

$P_i(\theta)$ : es la probabilidad de una respuesta correcta en el ítem  $i$  para aquellos sujetos con un nivel de  $\theta$  determinado,

$\theta$ : se refiere al nivel de competencia del sujeto en el rasgo medido por los ítems de la escala,

$b_i$ : es el índice de dificultad del ítem  $i$ ;

$a_i$ : se refiere el índice de discriminación del ítem  $i$ ;

$c_i$ : es el parámetro de pseudo-azar, e indica la probabilidad de una respuesta correcta para individuos con un valor en la variable medida por el test  $\theta$  que tiende a  $-\infty$ ,

$D$ : es una constante escalar igual a 1.702, que se utiliza para aproximar la función logística a la función de ojiva normal

Bajo la óptica de la TRI, la detección del DIF consistirá en calcular la curva característica del ítem (CCI) en cada uno de los grupos sometidos a comparación y posteriormente determinar si éstas coinciden (ausencia de DIF) o no (presencia de DIF). Como se ha visto anteriormente, el cálculo de las CCIs para cada grupo es una función de los parámetros que definen al

ítem, por tanto en una situación de DIF los parámetros del ítem o los ítems de los grupos sometidos a comparación no coincidirán. Por el contrario, si no existe DIF entonces los parámetros de los ítems no presentarán variaciones en ambos grupos, a excepción de las ocasionadas por la presencia del error muestral (Hambleton y Swaminathan, 1985; Hulin y cols, 1983; Lord, 1980; Thissen, Steinberg y Wainer, 1988).

Dentro de esta aproximación, se han propuesto varias estrategias para detectar el DIF. La primera de ellas consiste en comparar los parámetros de los ítems en los distintos grupos sometidos a análisis (Draba, 1977; Lord, 1977, 1980; Wright y Stone, 1979). Si los parámetros del ítem difieren entre ambos grupos entonces las CCIs no coincidirán, existiendo un área entre ellas. Por tanto, una *segunda* estrategia consistirá en determinar si el área existente entre las CCIs de ambos grupos es mayor que la que cabría esperar debida al azar. En este caso, la comparación directa de las CCIs permite una representación visual de la cantidad de funcionamiento diferencial que puede presentar un determinado ítem. Así pues, cuanto mayor sea el área entre ambas curvas mayor DIF presentará el ítem en cuestión (Kim y Cohen, 1991; Linn, Levine, Hastings y Wardrop, 1981; Raju, 1988, 1990; Rudner, 1977; Wainer, 1993). Por último, una *tercera* posibilidad en la evaluación del DIF es la comparación de modelos. Desde esta aproximación, se compara un modelo en que los parámetros del ítem son iguales en los dos grupos frente a otro en el que alguno o algunos de los parámetros del ítem difiere entre ambos grupos. Si el segundo modelo ajusta significativamente mejor a los datos que el primero, entonces podremos

concluir la presencia de DIF (Thissen, Steinberg y Gerrard, 1986; Thissen, Steinberg y Wainer, 1988, 1993).

No obstante, como paso previo a la aplicación de los métodos basados en la TRI es conveniente considerar los supuestos básicos de esta teoría, ya que de su cumplimiento va a depender su posterior aplicación. A su vez, también se presentarán sus principales ventajas y limitaciones.

Respecto a los **supuestos** de la TRI, el primero hace referencia a la *dimensionalidad* del instrumento de medida utilizado. La gran mayoría de los modelos y procedimientos de estimación ligados a la TRI requieren un único rasgo o constructo subyaciendo a las respuestas de los ítems al test. Y el segundo alude a la *independencia local*. Este supuesto afirma que la probabilidad de que un sujeto acierte  $n$  ítems es igual al producto de las probabilidades de acertar la respuesta a cada uno de ellos. En otras palabras, que para un sujeto con un valor  $x$  en la variable medida por el test el valor obtenido en un ítem  $i$  no viene determinado por las respuestas ofrecidas en los otros ítems. Al respecto, Hulin y cols (1983) indican que, en la investigación aplicada, es más conveniente examinar la dimensionalidad del constructo evaluado y, de esta forma, el requisito de la independencia local puede ser aceptado como una consecuencia directa de la unidimensionalidad. Sin embargo, los últimos avances en la medida de la dimensionalidad señalan que la presencia de DIF en un test se produce cuando, bajo ciertas condiciones, se incumple el supuesto de unidimensionalidad del test. Así pues, en la actualidad la multidimensionalidad aparece como una causa de la aparición del funcionamiento diferencial de los ítems en un test. Este tema será abordado

brevemente al final del apartado cuando se presenten los desarrollos o perspectivas actuales en el estudio del DIF.

Las principales **ventajas** de la utilización de la TRI se hallan en la posibilidad de detectar tanto el DIF uniforme como no uniforme, así como el hecho de proporcionar buenas estimaciones de  $\theta$  en tests de longitud corta si el modelo especificado se ajusta bien a los datos. Entre sus *limitaciones*, es importante destacar los grandes tamaños muestrales que se requieren para su análisis, además de ser computacionalmente costosos, aumentando de forma considerable con datos politómicos, y de requerir la aplicación de métodos de equiparación iterativos para que los parámetros estimados no se vean afectados por la presencia de DIF.

A continuación, vamos a examinar con más detalle los métodos más importantes de detección del DIF desarrollados desde la TRI.

#### **3.4.1.1. $\chi^2$ de Lord.**

Una estrategia desarrollada desde la TRI para la detección del DIF ha sido la de comparar directamente las estimaciones obtenidas de los parámetros de los ítems y comprobar la igualdad o desigualdad de los mismos. De tal modo que un ítem presenta un funcionamiento diferencial entre grupos si los parámetros que lo definen varían a través de los grupos considerados. En este ámbito, Lord (1980) propone un estadístico que permite someter a comprobación la hipótesis nula de ausencia de DIF.



En caso de utilizar datos *dicotómicos*, este estadístico de contraste viene dado por (Lord, 1980, pp.223):

$$\text{LORD-}\chi^2 = \hat{\mathbf{v}} \mathbf{S}^{-1} \hat{\mathbf{v}}' \quad (3.2)$$

donde,

$\hat{\mathbf{v}}$  es el vector de diferencias entre los parámetros estimados para un ítem  $i$  en el grupo de referencia y los parámetros estimados para ese mismo ítem en el grupo focal. Así, para el modelo de 3P el vector de diferencias será  $\mathbf{v} = (\hat{a}_F - \hat{a}_R, \hat{b}_F - \hat{b}_R, \hat{c}_F - \hat{c}_R)$ ,

$\hat{\mathbf{v}}'$  es el vector transpuesto de  $\hat{\mathbf{v}}$ , y

$\mathbf{S}^{-1}$  es la inversa de la matriz de varianza-covarianza asintótica de  $\hat{\mathbf{v}}$ , los vectores de diferencias entre los parámetros estimados para ambos grupos. Esta matriz es estimada por  $\mathbf{S}^{-1} = (\mathbf{S}_R + \mathbf{S}_F)^{-1}$ , siendo  $\mathbf{S}_R$  y  $\mathbf{S}_F$  las matrices con las varianzas y covarianzas de las estimaciones de los parámetros del ítem en cuestión en el grupo de referencia y focal, respectivamente.

La hipótesis nula:  $\hat{a}_R = \hat{a}_F$  y  $\hat{b}_R = \hat{b}_F$ , es sometida a prueba frente a la hipótesis alternativa que señala que  $\hat{a}_R \neq \hat{a}_F$  o  $\hat{b}_R \neq \hat{b}_F$ . El estadístico  $\chi^2$  de Lord bajo la hipótesis nula sigue una distribución  $\chi^2$ , con grados de libertad igual al número de parámetros del modelo evaluado. Así pues, un ítem presentará DIF si el valor observado en LORD- $\chi^2$  es mayor que el valor teórico asociado a  $\chi^2$  al nivel de significación establecido por el investigador.

Es importante destacar que la detección del DIF para el modelo logístico de 3P, aunque se puede evaluar utilizando la ecuación (3.2), presenta algunos problemas relacionados con la estimación del parámetro  $c$ . En este sentido, Lord (1980, pp.217) afirma que para este tipo de modelos es conveniente seguir el siguiente procedimiento: 1) ajustar el modelo de 3P tomando los datos de los dos grupos considerados conjuntamente; 2) volver a ajustar el modelo de 3P en cada uno de los grupos por separado, tomando como valor de  $c$  en ambos grupos el obtenido en la muestra total (primer paso), y 3) aplicar el estadístico  $\chi^2$  de Lord utilizando sólo los parámetros  $a$  y  $b$  obtenidos en el segundo paso.

Otro aspecto relevante en los estudios sobre DIF bajo la TRI es la restricción de que los parámetros del ítem en cuestión estén en la misma métrica antes de llegar a cabo cualquier tipo de comparación. Los estudios más actuales en este campo sugieren que el método de equiparación basado en la CCT (Curva característica del test; Stocking y Lord, 1983) es el método más adecuado y preciso, frente a otros (Baker y Al-Karni, 1991; Kim y Cohen, 1992), como son los métodos "Weighted Mean and Sigma" (WMS; Loyd y Hoover, 1980) o "Minimum Chi-Square" (MCS; Divgi, 1985), en especial para tamaños muestrales más pequeños.

Antes de finalizar, es conveniente señalar que la aplicación del estadístico de Lord (1980) no ha resultado ser muy efectiva en algunos estudios comparativos realizados. Precisamente, una de las mayores críticas que se hace al procedimiento  $\chi^2$  de Lord es que la hipótesis nula relativa a la ausencia de DIF puede ser rechazada incluso cuando las CCIs de ambos

grupos son similares pero no iguales, es decir, cuando existe muy poca diferencia en las regiones de las CCI's donde se encuentran la mayoría de los sujetos evaluados (Linn, Levine, Hasting y Wardrop, 1981). Así pues, este estadístico puede indicar la presencia de DIF en aquellas regiones donde apenas existen datos muestrales. Al respecto, algunos autores (Kim y Cohen, 1991; Raju, Drasgow y Slinde, 1993) señalan que siempre que se aplique el estadístico  $\chi^2$  de Lord y se obtengan resultados estadísticamente significativos, el cálculo de alguna medida de área para evaluar la cantidad de DIF existente puede ser de gran ayuda en la detección del DIF. Hidalgo y López Pina (1997) apuntan a que otros métodos como la regresión logística pueden ser igualmente efectivos como complemento al estadístico de Lord.

#### **3.4.1.2. Medidas del área.**

Dentro de la TRI, la segunda estrategia elaborada para detectar el DIF consiste en un conjunto de medidas basadas en el cálculo del área existente entre las Curvas Características del Item (CCI) en el grupo de referencia y en el focal para un intervalo determinado  $\theta[\theta_1, \theta_2]$ . De tal forma que, la cuantía del área obtenida constituye un índice de discrepancia entre las CCI's obtenidas y, en consecuencia, del DIF, ya que si ambas CCI's coincidiesen el área entre ambas CCI's sería cero y entonces habría ausencia de DIF. Aunque ésta es la lógica subyacente a todas las medidas del área, en la literatura consultada podemos apreciar que se han propuesto diferentes índices del área, los cuales difieren en función de los siguientes criterios: 1) Con signo o sin signo; 2) El intervalo de  $\theta$  sea finito o infinito; 3) La aproximación utilizada sea continua

(integrando) o discreta (utilizando el sumatorio); y 4) La ponderación o no de las diferencias entre las probabilidades.

En primer lugar, empezaremos definiendo la forma más general de estas medidas dentro de la aproximación continua, para mostrar a continuación algunas medidas discretas. La expresión general de estas medidas, basadas en la integración continua, viene dada por:

$$A_i = \int_{\theta_2}^{\theta_1} f [P_R(\theta) - P_F(\theta)] d(\theta) \quad (3.3)$$

donde,  $P_R(\theta)$  y  $P_F(\theta)$  representan a las probabilidades de responder correctamente al ítem en cuestión dada  $\theta$  para el grupo de referencia y focal respectivamente, y siendo  $\theta_1$  y  $\theta_2$  los límites inferior y superior elegidos para cubrir un amplio rango de valores de  $\theta$ , por ejemplo entre -3 y 3. La función  $f$  puede especificarse con o sin signo. En el primer caso se obtendría una *medida del área con signo*, y en el segundo una *medida del área sin signo o absoluta*. La expresión general de la ecuación (3.3) puede adoptar distintas formas según el modelo de TRI con el que se esté trabajando, y el signo de los resultados permitirá conocer la dirección del DIF evaluado. Si  $P_R(\theta)$  es habitualmente mayor que  $P_F(\theta)$  el signo será positivo, y entonces el ítem estará sesgado contra el grupo focal. Si  $P_R(\theta)$  es usualmente menor que  $P_F(\theta)$  el signo obtenido será negativo, y por tanto el ítem estará sesgado contra el grupo de referencia. Por lo tanto, la medida del área con signo estará indicada para aquellas situaciones donde las CCIs no se cruzan, como es el caso del DIF no uniforme, y la medida del área sin signo para aquéllas en las que las CCIs

se cruzan, ya que en una situación de DIF no uniforme como la que se muestra en la figura 1b., las diferencias positivas pueden anularse con las negativas, y por tanto llevar a concluir erróneamente la no existencia de DIF. También puede ocurrir que no se supriman completamente unas con las otras y la magnitud del DIF obtenida ser muy pequeña, a pesar de existir grandes discrepancias entre las CCIs obtenidas.

Las primeras medidas del área propuestas (Ironson y Subkoviak, 1979; Linn, Levine, Hastings y Wardrop, 1981; Rudner, 1977; Rudner, Getson y Knight, 1980a, 1980b) calculaban las diferencias entre las CCIs dentro del intervalo definido por dos valores específicos de  $\theta$ , dividiendo el rango de  $\theta$  en pequeños intervalos. Por ejemplo, Rudner (1977; Rudner, Getson y Knight, 1980a) propone el siguiente índice sin signo:

$$R = \sum_{\theta=-3}^{\theta=+3} |P_R(\theta_j) - P_F(\theta_j)| \Delta\theta \quad (3.4)$$

donde,  $\theta_1=-3$ ,  $\theta_2=+3$  y  $\Delta=0.005$ . La ecuación 3.4 especifica que la región de  $\theta$  comprendida entre  $-3$  y  $3$  se dividirá en 600 intervalos de amplitud 0.005, procediendo a la suma de las diferencias en las proporciones entre el grupo de referencia y el focal para cada intervalo. Si se elimina el operador absoluto de la ecuación 3.4, ésta se convierte en la medida del área con signo.

Sin embargo, las medidas del área presentadas hasta el momento poseen algunos *inconvenientes*. En primer lugar, señalar que no tienen en consideración el número de sujetos que existe en cada uno de los intervalos

en los que se divide la habilidad, lo cual implica que se esté dando la misma importancia a las diferencias existentes en aquellas regiones donde hay muy pocos sujetos y en aquéllas donde se concentran la mayoría de ellos. Y en segundo lugar, que las integrales o la naturaleza continua de  $\theta$  no posibilita la obtención de valores finitos cuando los parámetros  $c$  de ambos grupos son desiguales.

En cuanto a la *primera desventaja*, para resolver este problema, Linn y Harnisch (1981) propuso varios índices de área ponderados. Por su parte, Shepard, Camilli y Williams (1984) también proponen un índice del área que da más peso a aquellas parcelas de la escala de habilidad donde hay un mayor número de sujetos. La expresión de este índice viene dada por:

$$SOS1 = \frac{1}{N} \sum_{j=1}^N [P_R(\theta_j) - P_F(\theta_j)]^2$$

donde,  $N$  es el número total de la muestra en ambos grupos;  $\theta_j$  indica el rango de área considerada para un solo grupo, normalmente para el grupo focal, y la diferencia de probabilidades se obtiene hallando las probabilidades de responder correctamente al ítem en cuestión, pero estimando la probabilidad que le correspondería al otro grupo (grupo de referencia) en el valor  $\theta_j$ . De esta forma, las diferencias entre las CCIs cuantificadas en términos de proporciones son sumadas a través de las diferentes estimaciones obtenidas de  $\theta$ , y, en consecuencia, las regiones de  $\theta$  con mayor número de sujetos obtienen una mayor ponderación. En este caso, la estrategia seguida para obtener la medida del área sin signo no ha

consistido en tomar las diferencias halladas en sentido absoluto, sino elevarlas al cuadrado. Además de éstas, Shepard, Camilli y Williams (1984) también propusieron otras medidas ponderadas del área con y sin signo, en gran parte basadas en las sugeridas por Linn, Levine, Hasting y Wardrop (1980). Sin embargo, algunos estudios realizados muestran que a pesar de ser los índices ponderados más aconsejables desde un punto de vista teórico, en la práctica se encuentran pocas diferencias entre los índices ponderados y no ponderados, obteniéndose en general los mismos resultados (Linn, Levine, Hasting y Wardrop, 1981; Shepard, Camilli y Williams, 1985).

Respecto a la *segunda desventaja*, recientemente se han propuesto algunos índices del área que consideran la naturaleza continua de  $\theta$ , pudiendo ser los límites de la integral presentada en la ecuación (3.4) tanto infinitos como finitos.

Para el caso de los límites infinitos, Raju (1988) en un primer momento propone las *medidas del área exactas con signo* (AECS) y las *medidas del área exactas sin signo* (AESS), y en un artículo posterior (Raju, 1990) determina las distribuciones muestrales asintóticas de la media y varianza de las citadas medidas del área. De esta forma, una vez se conoce el grado de variabilidad esperado a causa del muestreo, se puede determinar con seguridad si las diferencias halladas entre las CCI's se deben a errores aleatorios del muestreo o son estadísticamente significativas. La formulación correspondiente a estas medidas, junto con sus medias y varianzas, así como el estadístico de contraste pertinente ( $Z$ ) para los modelos logísticos de uno, dos y tres parámetros se pueden consultar en las siguientes referencias bibliográficas (Raju, 1988, pp. 495-502; Raju, 1990,

pp.197-207). Por lo que respecta a su *interpretación*, la presencia de valores positivos en los índices con signo mencionados indicarán que el ítem favorece al grupo de referencia, mientras que los valores negativos indicarán lo contrario, es decir, que el ítem favorece al grupo focal. Todos los estadísticos de contraste a utilizar siguen la distribución normal  $N(0,1)$ , por tanto la hipótesis nula basada en que el área observada entre las CCIs de ambas poblaciones es nula se rechazará al nivel de confianza de  $1-\alpha$ , si  $Z \leq z_{(\alpha/2)}$  o  $Z \geq z_{(1-\alpha/2)}$ . En su publicación, Raju (1990, pág. 202) recomienda utilizar valores muy pequeños de alfa (p. ej.  $\alpha=0.0028$ ), ya que la mayoría de los análisis realizados en TRI presentan muestras con tamaños elevados (500 sujetos o más), lo que provoca un aumento de la probabilidad de que diferencias no importantes entre las CCIs consideradas resulten estadísticamente significativas.

Para el caso de los límites finitos, Kim y Cohen (1991) desarrollan la *medida del área de intervalo cerrado con signo* (ACCS) y la *medida del área de intervalo cerrado sin signo* (ACSS). Estos índices de área pueden obtenerse para cada uno de los modelos logísticos propuestos en la TRI: modelos de 1P, 2P y 3P. Dado que estos índices, al igual que los anteriores, no van a ser utilizados en la presente investigación, los lectores interesados en la obtención de su formulación para cada uno de los modelos señalados puede remitirse a la fuente bibliográfica original (Kim y Cohen, 1991, pp. 269-278). En relación a su interpretación, señalar que la obtención de valores positivos serán un indicativo de que el ítem en cuestión es más fácil para el grupo de referencia y los valores negativos que es más difícil para el citado grupo.



Si comparamos las medidas de área propuestas por Kim y Cohen (1991) y Raju (1988, 1990) con el resto, éstas son preferibles a los otros índices de área propuestos, tanto desde el punto de vista teórico, ya que asumen la naturaleza continua de  $\theta$ , como práctico, debido a que sus cálculos son notablemente menores que el resto de los índices de área vistos. Por su parte, las medidas exactas del área (Raju, 1988) frente a las medidas de intervalos cerrados (Kim y Cohen, 1991) presentan como principales ventajas, primero su simplicidad y fácil manejo; segundo, la posibilidad de no tener que elegir un intervalo de valores determinado y en cierta manera arbitrario, y tercero y más importante, la posibilidad de comprobar la significación estadística de las diferencias encontradas entre las CCI's comparadas. Sin embargo, estas medidas de área presentan como principal desventaja la imposibilidad de su cálculo para el modelo logístico de 3P cuando existen diferencias para ambos grupos en los valores del parámetro  $c$ . Debido a ello la integral de la ecuación (3.3) no puede producir valores finitos cuando se integra entre  $-\infty$  e  $\infty$ . En este caso en particular, las medidas del área de intervalos cerrados sin signo son la única alternativa viable para detectar el DIF de los items. También, cuando se poseen intervalos de habilidad con muy pocos sujetos, hecho muy frecuente con tamaños muestrales pequeños, las medidas del área de intervalos cerrados permiten seleccionar únicamente el intervalo con mayor densidad de sujetos para el cálculo de la diferencia entre las CCI's de ambos grupos. Por tanto, las medidas del área de intervalos cerrados se ven menos afectados por el tamaño de la muestra que las medidas exactas del área. Así, pues su uso es aconsejable cuando se tengan tamaños muestrales pequeños.

Por último, antes de finalizar este procedimiento se comentarán los pasos necesarios a seguir para poder llevar a cabo una adecuada evaluación del DIF con aquellas medidas de área que no tienen asociado ningún test de significación estadística para poder interpretar a partir de qué valores existe DIF en los resultados obtenidos. Las soluciones aportadas a este problema en los estudios empíricos realizados ofrecen dos estrategias diferentes (Rogers y Hambleton, 1989). La primera de ellas consiste en obtener una línea base que sirva de criterio para poder juzgar la magnitud de los índices de DIF. Para ello, se dividirá aleatoriamente a los sujetos pertenecientes a un grupo en dos subgrupos, y posteriormente se calculará en dichos subgrupos las medidas de área, siendo el valor más alto encontrado el punto de corte para evaluar el DIF. Sin embargo, esta estrategia presenta dos inconvenientes. Primero, que la línea base se obtiene comparando sujetos pertenecientes al mismo grupo, lo que provoca que esta opción no sea sensible a la influencia que puede ejercer la presencia de impacto entre los grupos que realmente se someterán a comparación (grupo de referencia versus grupo focal). Segundo, que el tamaño muestral empleado para calcular la línea base no es el mismo que el utilizado cuando se calculan las medidas del área dentro de cada grupo. Para evitar estas desventajas, la otra alternativa propuesta consiste en simular las puntuaciones obtenidas por los sujetos en el test de forma que éstas reflejen tan fielmente como sea posible los datos reales, pero sin la presencia de DIF (Rogers y Hambleton, 1989). En este caso, la ausencia de DIF está garantizada ya que las respuestas de cada uno de los grupos fueron generadas utilizando los parámetros de los items obtenidos en el grupo

combinado. Al igual que la estrategia anterior, el valor más alto encontrado se utilizará como punto de corte para evaluar el DIF.

#### **3.4.1.3. Comparación de modelos.**

La siguiente aproximación basada en la comparación de modelos es implementada en la TRI por Thissen, Steinberg y Gerrard (1986). Este procedimiento permite comprobar la igualdad de parámetros de las CCI's de un mismo ítem administrado a distintos grupos de sujetos. Para realizar dicha comprobación se establece una comparación de modelos denominados modelo compacto (modelo C), en el que se establece la restricción de igualdad de parámetros de los ítems en ambos modelos, y modelo aumentado (modelo A), que especifica los parámetros del modelo compacto y además establece una adición de parámetros en los ítems para ambos grupos. Esta nueva estrategia consiste en determinar si los parámetros adicionales del modelo A son significativamente diferentes de 0, es decir, si la inclusión de éstos en el modelo A mejoran significativamente el ajuste de este modelo a los datos. Así pues, una vez obtenidos los resultados si el ajuste a los datos de un modelo que incluye parámetros diferentes en los ítems para el grupo focal y de referencia (modelo A) es significativamente mejor que un modelo en el que los parámetros de los ítems son iguales para ambos grupos (modelo C), entonces se podrá concluir que el ítem presenta DIF (Kelderman, 1989; Thissen, Steinberg y Gerrard, 1986; Thissen, Steinberg y Wainer, 1988).

El estadístico utilizado para comparar ambos modelos suele ser la *razón de verosimilitudes* (LR), expresión que viene dada por:

$$LR = -2 \ln [L(1)/L(2)]$$

$$LR = [-2 \ln L(\text{Modelo compacto})] - [-2 \ln L(\text{Modelo aumentado})]$$

Este último paso puede ser explicado por la equivalencia matemática  $\ln (X/Y) = \ln X - \ln Y$ . El estadístico de contraste empleado sigue la distribución  $\chi^2$  con un número de grados de libertad igual a la diferencia en el número de parámetros entre el modelo aumentado (A) y el compacto (C) bajo la hipótesis nula de que los parámetros del modelo aumentado son iguales a 0. Así pues, si  $LR \geq \chi^2_{1-\alpha, g1}$  entonces se rechazará la hipótesis nula, y por tanto se podrá concluir que el ítem presenta DIF.

Entre sus ventajas, se aprecia que este método no necesita de ninguna equiparación de puntuaciones en los dos grupos sometidos a comparación para poder detectar adecuadamente el DIF de los ítems. Además, como se indica en diversas publicaciones (Thissen, Steinberg y Gerrard, 1986; Wainer, Sireci y Thissen, 1991) esta estrategia puede ser utilizada para comprobar la presencia de DIF en varios ítems simultáneamente. Sin embargo, sus principales limitaciones residen en su incapacidad para detectar diferencias entre parámetros menores de 0.3 y en su dependencia del tamaño muestral. De hecho, algunos autores (Fidalgo, 1996) señalan que al igual que en el procedimiento de Lord (1980) el cálculo de alguna medida del área de intervalos cerrados puede ser una ayuda complementaria cuando la comparación de modelos resulte significativa, contribuyendo a obtener una interpretación adecuada de los resultados obtenidos.

Para finalizar este apartado, destacaremos que la TRI además de permitir la detección del DIF en datos dicotómicos, también ha experimentado algunos avances y progresos en el estudio del sesgo con datos politómicos, aunque debido a su reciente incorporación al estudio del DIF todavía no ofrece resultados concluyentes. En concreto, con ítems politómicos se han utilizado tanto índices de área con y sin signo, como procedimientos basados en la comparación de los parámetros del ítem similares a  $\chi^2$  de Lord (Cohen, Kim y Baker, 1993; Hidalgo, 1995; Hidalgo y López, 1997), además de estrategias de comparación de modelos (Thissen y Steinberg, 1986; Wainer, Sireci y Thissen, 1991). Para una ampliación de esta información se pueden consultar los artículos de Potenza y Dorans (1995) y Mellenbergh (1995).

#### **3.4.2. Procedimientos que no especifican ningún modelo de medida.**

Dentro de esta sección, vamos a exponer aquellos procedimientos que permiten evaluar el DIF sin necesidad de especificar un modelo de medida. Entre ellos, destacan como los más empleados: el procedimiento Mantel-Haenszel (MH), la estandarización (ST) y la técnica de la Regresión Logística (RL). A su vez, dentro de este apartado se expondrá también el método SIBTEST, un procedimiento de detección del DIF que encuentra su fundamento en un modelo de TRI multidimensional no paramétrico.

##### **3.4.2.1. Procedimiento de Mantel-Haenszel (MH).**

Esta medida de asociación fue desarrollada por Mantel-Haenszel (1959), pero no fue aplicada a la investigación sobre la detección del DIF hasta finales de los 70 (Camilli, 1979; Scheuneman, 1975).

La aplicación de esta técnica para la detección del DIF permite clasificar a los sujetos pertenecientes a los grupos focal y de referencia en función del nivel de habilidad alcanzado en el test, de tal manera que la ejecución realizada por el grupo focal, objeto de interés, es comparada con la del grupo de referencia para el mismo nivel de habilidad  $k$  en ambos grupos. La puntuación total en el test es el criterio a utilizar para obtener un estimador de los distintos niveles de la variable habilidad que pretende medir el test. Por tanto, la información obtenida puede clasificarse en tres partes bien diferenciadas: a) las respuestas de los sujetos en el ítem en cuestión, b) la puntuación total de éstos mismos en el test y c) su grupo de pertenencia, puede ser distribuida en  $k$  tablas de contingencia de  $2 \times 2$ , siendo  $k$  el número de intervalos en los que se divide la puntuación total en el test. De esta manera, para cada intervalo o nivel de puntuación  $k$ , tendremos una tabla de contingencia de  $2 \times 2$  con la variable pertenencia al grupo (referencia/focal) en una de las entradas y la respuesta al ítem (acierto/error) en la otra, tal como puede observarse en la Tabla 3.1.

**Tabla 3.1.**

Tabla de contingencia  $2 \times 2$  para el nivel de habilidad  $k$ .

GRUPO	RESPUESTA		TOTAL
	ACIERTO	ERROR	
$G_R$	$A_k$	$B_k$	$N_{Rk}$
$G_F$	$C_k$	$D_k$	$N_{Fk}$
	$N_{1k}$	$N_{0k}$	$N_k$

Los valores  $A_k, B_k, C_k$  y  $D_k$  incluidos en la tabla denotan el número de sujetos en cada categoría. Los valores marginales  $N_{Rk}$  y  $N_{Fk}$  representan el número de sujetos pertenecientes a los grupos de referencia y focal, respectivamente; y los valores  $N_{1k}$  y  $N_{0k}$  el número de sujetos que han contestado de forma correcta e incorrecta el ítem en cuestión, respectivamente. Finalmente,  $N_k$  es el número total de sujetos en el nivel de puntuación  $k$ .

La lógica que corresponde al procedimiento MH es la siguiente. Si el ítem no presenta DIF, entonces la razón entre el número de sujetos que aciertan el ítem y aquéllos que lo fallan debe ser la misma en los dos grupos objeto de comparación a lo largo de todos los niveles de habilidad. De esta manera, las hipótesis sometidas a prueba serán:

$$H_0: (A_k/B_k) = \alpha (C_k/D_k) \text{ siendo } \alpha = 1 \text{ para todo } k$$

$$H_1: (A_k/B_k) = \alpha (C_k/D_k) \text{ siendo } \alpha \neq 1 \text{ en algún } k$$

Precisamente, Holland y Thayer (1986, 1988) son quienes proponen utilizar el estadístico Mantel-Haenszel (Mantel y Haenszel, 1959) para evaluar la hipótesis nula de ausencia de DIF. Este estadístico viene dado por la siguiente expresión:

$$\chi_{MH}^2 = \frac{\left( \left| \sum_{k=1}^m A_k - \sum_{k=1}^m E(A_k) \right| - 0,5 \right)^2}{\sum_{k=1}^m \text{Var}(A_k)}$$

donde  $E(A_k)$  es el valor esperado de  $A_k$  y  $Var(A_k)$  es su varianza, siendo iguales a:

$$E(A_k) = (N_{Rk} N_{1k}) / N_k$$

y

$$Var(A_k) = \frac{N_{Rk} N_{Fk} N_{1k} N_{0k}}{N_k^2 (N_k - 1)}$$

El estadístico  $\chi_{MH}^2$  sigue una distribución  $\chi^2$  con un grado de libertad. Si  $\chi_{MH}^2 \geq \chi_{1-\alpha}^2$  entonces se concluirá que el ítem estudiado presenta DIF al nivel de significación  $\alpha$ .

Además, Mantel y Haenszel (1959) ofrecen un estimador de  $\alpha$  denominado cociente de razones común ( $\hat{\alpha}_{MH}$ , *common odds ratio*), cuya expresión viene dada por:

$$\hat{\alpha}_{MH} = \frac{\sum_{k=1}^m A_k D_k / N_k}{\sum_{k=1}^m B_k C_k / N_k}$$

En concreto, el  $\hat{\alpha}_{MH}$  es un estimador de la magnitud del tamaño del DIF en una métrica que varía entre 0 e  $\infty$ . En cuanto a su interpretación, señalaremos que un valor de 1 representa la hipótesis nula de ausencia de DIF, un valor mayor que 1 sugiere que el ítem estudiado favorece al grupo de referencia y, por el contrario, si  $\hat{\alpha}_{MH}$  es menor que 1 ello indica que el ítem está favoreciendo al grupo focal. Para una mejor comprensión e interpretación del estadístico, Holland y Thayer (1985) proponen una



transformación del estimador  $\hat{\alpha}_{MH}$  a la escala delta. Esta transformación queda reflejada por:

$$MH\ D-DIF = -2,35 \ln[\hat{\alpha}_{MH}]$$

donde la expresión  $\ln[\hat{\alpha}_{MH}]$  es igual al logaritmo neperiano de  $\hat{\alpha}_{MH}$ . En cuanto a la interpretación del DIF, ésta se realiza de la siguiente manera: valores próximos a 0 expresan la ausencia de DIF, valores negativos indican que el ítem favorece al grupo de referencia, y valores positivos que el ítem favorece al grupo focal.

En la actualidad, el método MH es el procedimiento seleccionado por el *Educational Testing Service* (ETS) para la detección de items con DIF en los programas de evaluación del rendimiento de los estudiantes norteamericanos. La adaptación de sus valores a la escala Delta ha permitido al ETS distinguir tres categorías distintas de items, denominadas A, B y C en función de la magnitud de DIF establecido en el ítem (Zieky, 1993). Dichas categorías son:

(1) Items tipo A, aquellos items cuyo valor en MH D-DIF no es significativamente distinto de 0 ( $\alpha=0.05$ ), o dicho valor en valor absoluto  $|MH\ D-DIF|$  es menor que 1. Todos los items que cumplen este requisito son considerados libres de DIF;

(2) Items tipo B, son aquellos con un valor en MH D-DIF significativamente distinto de 0 ( $\alpha=0.05$ ), y poseen o bien un  $|MH\ D-DIF| < 1.5$  o bien un  $|MH\ D-DIF|$  no significativamente diferente de 1.

Este tipo de items pueden ser incluidos en el test pero con algunas reservas, es decir, se aconseja una inspección de los mismos, siendo preferibles los items tipo A, y por último,

(3) Items tipo C, que incluyen a aquellos items significativamente más grandes que 1 ( $\alpha=0.05$ ) y cuyo valor en  $|MH D-DIF| \geq 1.5$ . Estos items son los que presentan problemas de funcionamiento diferencial y por tanto su inclusión en el test sólo se admite si es estrictamente necesario.

Respecto a sus ventajas y desventajas, el procedimiento MH se caracteriza por su simplicidad en cálculo e interpretación, y sus buenos resultados tanto en muestras pequeñas como en la cuantificación del DIF en los items, además de ofrecer un test de significación estadística, siendo así uno de los métodos más utilizados para detectar el DIF. Sin embargo, a pesar de tener varias ventajas su principal deficiencia ha sido su incapacidad para detectar el DIF no uniforme (Rogers y Swaminathan, 1993; Swaminathan y Rogers, 1990), recomendándose hasta hace poco la aplicación de métodos loglineales, logit o de regresión logística. Sin embargo, este déficit pudo ser recientemente resuelto mediante la aplicación de una modificación de cálculo en el método MH propuesta por Mazor, Clauser y Hambleton (1994). Precisamente, dicha modificación consistió en calcular de forma separada los estadísticos MH en el grupo de sujetos con puntuaciones más bajas en el test (en concreto, los sujetos con una puntuación total menor o igual a la media de la distribución de las puntuaciones en el test) y en el grupo de sujetos con mayores puntuaciones (sujetos con una puntuación total mayor que la media de la distribución de las puntuaciones

en el test). De hecho, algunos trabajos de simulación que han utilizado esta variación del método MH la señalan como una alternativa eficaz para detectar el DIF no uniforme aunque a costa de incrementar la tasa de error tipo 1 (Fidalgo, 1996; Fidalgo y Mellenbergh, 1995).

La enorme cantidad de literatura existente en torno a este método habla por sí sola. El método MH es una de las técnicas estadísticas más utilizadas en la detección de DIF. La mayor parte de los trabajos realizados tanto con datos reales como de simulación han sido desarrollados principalmente con el fin de determinar cómo se ve afectado este método ante variables como el tamaño muestral, el porcentaje de items sesgados, la longitud del test, y otras más. Precisamente, en el artículo de Hambleton, Clauser, Mazor y Jones (1993) podemos encontrar un buen resumen de las principales implicaciones de tipo práctico que se desprenden de estos estudios. Estas, junto a algunas otras procedentes de otros estudios (Clauser, Mazor y Hambleton, 1991; Mazor, Clauser y Hambleton, 1992; Clauser, Mazor y Hambleton, 1993; Clauser, Mazor y Hambleton, 1994; Mazor, Clauser y Hambleton, 1994; Miller y Oshima, 1992; Ryan, 1991; Uttaro y Millsap, 1994; Fidalgo, Mellenbergh y Muñiz, 1997) se presentan a continuación:

1) Respecto al *tamaño muestral*, en términos generales los resultados obtenidos muestran claramente que un aumento del tamaño muestral lleva consigo un mayor poder estadístico del método utilizado. Sin embargo, dado el buen comportamiento de este método ante la presencia de tamaños muestrales pequeños, esta cuestión ha sido foco de interés en muchos trabajos y estudios elaborados. Sin embargo, a pesar de la abundante literatura en este tema parece ser que no existe un consenso generalizado

acerca de qué tamaño muestral es necesario para un estudio y detección adecuada del DIF con el método MH. Por ejemplo, Hills (1989) sugiere que este método funciona correctamente con muestras tan pequeñas como 100 sujetos por grupo. Shepard y cols. (1985) señalan que una adecuada detección del DIF en el test se consigue con muestras de 300 sujetos para el grupo focal y 1000 para el grupo de referencia. Por su parte, Kubiak y Colwell (1990) indican que tamaños muestrales con un total de 500 sujetos y un mínimo de 100 para el grupo focal son adecuados. Parshall y Kromrey (1992) y Spray (1989) utilizan tamaños muestrales muy reducidos de 50 y 200 sujetos para el grupo focal y de referencia, respectivamente, y por parte, Mazor y cols (1992) trabajan con muestras de 100 sujetos por grupo. Sin embargo, en la mayoría de estos trabajos el funcionamiento de este estadístico no resultó ser adecuado por la utilización de tamaños muestrales tan reducidos. Precisamente, Mazor y cols (1992) sugieren trabajar con 200 sujetos por grupo, mientras que Spray (1989) encontró que el método MH funcionaba adecuadamente a partir de 250 sujetos por grupo. En base a estos estudios, parece ser que el tamaño muestral mínimo para la utilización con ciertas garantías de la técnica MH en la detección del DIF es de 200 personas por grupo. Además, cuando se aumenta el tamaño del grupo de referencia manteniendo constante el del grupo focal en una razón de 1 a 9, también se produce un incremento del poder estadístico del método MH.

2) En cuanto a las *propiedades del ítem*, los items con DIF que presentan baja discriminación y alta dificultad son los items más difíciles de detectar por el método MH, y en especial cuando la distribución de la habilidad en ambos grupos es desigual (Mazor, Clauser, Hambleton, 1992;

Clauser, Mazor y Hambleton, 1994). De todas formas, aunque el método MH resulta más efectivo con distribuciones iguales para ambos grupos en la habilidad, éste también funciona adecuadamente con distribuciones desiguales. No obstante, lo que si es aconsejable es la utilización de un nivel de significación más pequeño para poder detectar correctamente aquellos items con DIF que son muy difíciles o muy poco discriminativos. E incluso en caso de poseer un porcentaje muy pequeño de items poco discriminativos, proceder a su eliminación. Sin embargo, esta deficiencia no es un problema cuando el test a analizar es un test de aptitud, dado que su construcción técnicamente se basa en la inclusión de items en su mayoría con alta discriminación y de dificultad media.

3) En relación a la *amplitud del criterio*, los estudios realizados intentan examinar qué poder estadístico posee el método MH cuando se reducen el número de grupos formados dentro de la variable criterio. Holland y Thayer (1988) recomiendan realizar los análisis al nivel más fino, es decir, que las puntuaciones de los sujetos en el test sean divididas en  $k + 1$  categorías, siendo  $k$  el número de items en el test. Por su parte, Raju, Bode y Larsen (1989) manifiestan que el número apropiado de niveles está en función del rango de puntuaciones, siendo para su estudio 4 ó 5 grupos los valores óptimos recomendados en un rango de puntuaciones de 0 a 40. Los resultados obtenidos con este método muestran que una reducción del número de grupos a comparar en el criterio va asociada a un mayor poder estadístico del método, siempre y cuando las distribuciones de la habilidad en ambos grupos sean similares. Por contra, cuando existen diferencias entre grupos en las distribuciones de la habilidad aumenta la tasa de error tipo I, es decir, aumenta la identificación incorrecta o falsa de items con

DIF? a medida que disminuye el número de categorías (Clauser, Mazor y Hambleton, 1994).

4) En la medida que el criterio utilizado para comparar sujetos con igual nivel de competencia o habilidad (en este caso la puntuación total en el test) esté contaminada por la presencia de items con DIF, se estará utilizando un criterio distorsionado o sesgado para analizar correctamente el DIF de los items del test, y por tanto, es probable que se produzca un incremento de la tasa de error tipo I. Una solución a este problema es la utilización de métodos de purificación. De hecho, los trabajos realizados en este campo ofrecen apoyo empírico a la utilidad de estos *métodos de purificación* en el criterio, señalando que las ganancias obtenidas son más patentes cuando la magnitud del DIF es mayor (Miller y Oshima, 1992) y la distribución muestral de la habilidad es igual para ambos grupos (Clauser, Mazor y Hambleton, 1993). Holland y Thayer (1988) recomiendan para el método MH una purificación en dos etapas del criterio utilizado. Este procedimiento consiste en eliminar del criterio aquellos items que fueron identificados con DIF en el primer análisis y recalcular convenientemente los estadísticos de DIF para todos los items del test en la segunda etapa.

5) Por lo que respecta a la *inclusión del ítem estudiado en el criterio*, Holland y Thayer (1988) afirman que las definiciones acerca del DIF son equivalentes para las aproximaciones de la TRI y MH, si se cumplen las siguientes condiciones: a) todos los items, incluido el ítem objeto de estudio, se ajustan a un modelo de Rasch; b) la no presencia de DIF en los items del test, a excepción del ítem estudiado; c) la inclusión del ítem bajo estudio en el criterio, y d) la selección aleatoria de los grupos

focal y de referencia. Bajo estas condiciones, la utilización de la puntuación total en el test es un estimador suficiente del nivel de habilidad ( $\theta$ ), dado que es imposible conocer los parámetros estimados para cada nivel de la medida de habilidad. Varios estudios se han dedicado a estudiar el comportamiento de éste y demás métodos bajo condiciones más generales, es decir, bajo el incumplimiento de algunas de las condiciones mencionadas (Lewis, 1993; Meredith y Millsap, 1992; Millsap y Meredith, 1992; Uttaro y Millsap, 1994; Zwick, 1990; Zwick, Donoghue y Grima, 1993). Los resultados aportados para el método MH señalan que cuando el ítem está siendo evaluado, éste debe ser siempre incluido en el cálculo de la puntuación total en el test, incluso si ha presentado DIF en el análisis inicial.

6) En lo referente a la detección de *DIF no uniforme*, se recomienda utilizar la modificación del método MH que hemos descrito con anterioridad (Mazor, Clauser y Hambleton, 1994).

7) El funcionamiento del método MH es excelente cuando los ítems del test sólo varían en dificultad, siendo todos ellos igualmente discriminativos y no existiendo aciertos al azar, es decir, cuando se ajustan al modelo de Rasch. Sin embargo, los datos aportados hasta el momento también señalan que este método sigue proporcionando buenos resultados aún cuando las condiciones mencionadas no se dan.

Por último, señalar que los últimos avances con este método muestran algunas modificaciones del método MH. Precisamente, una de ellas permite su utilización para evaluar el DIF en ítems politómicos (Welch y Hoover, 1993; Welch y Miller, 1995; Zwick y cols, 1993).

### 3.4.2.2. Estandarización.

El siguiente procedimiento de detección de DIF que analizaremos será la Estandarización. El núcleo central de este procedimiento consiste en determinar la diferencia en la proporción de sujetos que aciertan el ítem entre el grupo de referencia y el grupo focal para cada nivel de la variable habilidad. De esta forma, si el ítem evaluado no presenta DIF no existirán diferencias en las proporciones mencionadas entre los dos grupos a lo largo de todos los niveles de habilidad considerados.

Las características que definen a este procedimiento hacen de él un método conveniente y eficaz para la descripción del DIF. Ello se debe, primero, a la posibilidad de obtener una representación gráfica de las diferencias en las proporciones condicionales entre el grupo focal y el grupo de referencia, lo cual permite obtener una medida de la cantidad de DIF que un determinado ítem puede exhibir; y segundo, a que la estandarización proporciona un índice numérico para cuantificar el DIF, conocido como la *diferencia en proporciones estandarizadas* (DPE). Este índice se define como:

$$DPE = \frac{\sum_{k=1}^m [W_k (P_{Fk} - P_{Rk})]}{\sum_{k=1}^m W_k} \quad (3.5)$$



donde,  $W_k$  es el factor de ponderación en el nivel de habilidad  $k$ , utilizado para ponderar las diferencias en la proporción de respuestas correctas entre el grupo de referencia y el focal. Por su parte,  $P_{Fk}$  y  $P_{Rk}$  son las proporciones de sujetos que responden con éxito al ítem en cuestión para el nivel de habilidad  $k$  en el grupo focal y de referencia, respectivamente. Dichas proporciones, teniendo en cuenta la Tabla 1 son iguales a:

$$P_{Fk} = C_k / N_{Fk} \quad \text{y} \quad P_{Rk} = A_k / N_{Rk}$$

Por lo que respecta a su interpretación, el índice DPE puede variar entre -1 y 1, obteniendo valores positivos cuando el ítem favorece al grupo focal y valores negativos cuando favorece al grupo de referencia. Para una mejor interpretación, Dorans y Holland (1993) proponen los siguientes criterios:

- a) Valores entre -0,05 y 0,05 indican ausencia de DIF.
- b) Valores entre -0,10 y -0,05 y entre 0,05 y 0,10 aconsejan una inspección de los ítems.
- c) Valores fuera del rango [-0,10 ; 0,10] son altamente sospechosos, por tanto los ítems que los posean deben ser examinados detenidamente.

Por otra parte, es importante destacar que el factor de ponderación empleado en el método de la estandarización puede variar en función de los propósitos de la investigación. Así pues, algunas posibles opciones de aplicación son las siguientes:

$W_k=N_k$ , número de sujetos en el grupo total en el nivel de habilidad  $k$

$W_k=N_{Rk}$ , número de sujetos en el grupo de referencia en el nivel de habilidad  $k$ .

$W_k=N_{Fk}$ , número de sujetos en el grupo focal en el nivel de habilidad  $k$

$W_k$ =la frecuencia relativa de algún grupo de referencia en el nivel de habilidad  $k$

Habitualmente, en la práctica se utiliza la opción  $W_k=N_{Fk}$  ya que esta opción proporciona la ponderación más alta en cuanto a las diferencias entre  $P_{Fk}$  y  $P_{Rk}$  en los niveles de habilidad más frecuentemente conseguidos por el grupo focal bajo estudio, y también de esta manera se elimina la influencia que puede ejercer la existencia de diferencias en ambos grupos en la habilidad medida.

Al igual que el método MH, los valores obtenidos en el estadístico DPE también pueden transformarse a la escala delta. Esta transformación viene dada por la siguiente expresión:

$$DPE \text{ D-DIF} = -2,35 \ln \left[ \frac{\left( \frac{\sum_{k=1}^m N_{Fk} P_{Rk}}{\sum_{k=1}^m N_{Fk}} \right)}{\left( 1 - \frac{\sum_{k=1}^m N_{Fk} P_{Rk}}{\sum_{k=1}^m N_{Fk}} \right)} \right] / \left[ \frac{\left( \frac{\sum_{k=1}^m N_{Fk} P_{Fk}}{\sum_{k=1}^m N_{Fk}} \right)}{\left( 1 - \frac{\sum_{k=1}^m N_{Fk} P_{Fk}}{\sum_{k=1}^m N_{Fk}} \right)} \right] \quad (9.6)$$

La justificación de esta transformación se localiza en el hecho de que existe una mayor correlación entre los valores de la ecuación DPE D-DIF (9.6) y los obtenidos aplicando la expresión MH D-DIF, que entre los valores en MH D-DIF y DPE (Dorans y Holland, 1993).

Por otra parte, recientemente Dorans y Holland (1993) han deducido el error típico asociado al índice de estandarización, expresión que viene dada en Dorans y Holland (1993, pp. 50), lo cual ha permitido evaluar la hipótesis nula de ausencia de DIF. Por lo tanto, una vez conocido el error típico de este estadístico es posible evaluar no sólo la magnitud de las diferencias consideradas, sino también si éstas son estadísticamente significativas. En este caso, el estadístico de contraste denominado (E) es igual a:

$$E = DPE / \text{VAR}(DPE)$$

que sigue una distribución normal  $N(0,1)$ . Así pues, la hipótesis nula de ausencia de DIF se rechazará si se cumple que  $E \leq Z_{\alpha/2}$  ó  $E \geq Z_{1-\alpha/2}$ .

En lo que respecta a las consideraciones de tipo práctico para el método de estandarización (DPE), destacar que los comentarios señalados anteriormente para el método MH se hacen extensivos a este procedimiento, a excepción del citado para la detección del DIF no uniforme. Precisamente, uno de los inconvenientes de la DPE ha sido su incapacidad para detectar este tipo de DIF, sin embargo esta deficiencia ha quedado cubierta últimamente con la creación de una nueva versión del método SIBTEST para detectar el DIF no uniforme (Crossing SIBTEST; Li y Stout, 1994a, 1994b).

Precisamente, cuando el método SIBTEST posee como objeto de estudio la detección del DIF en un sólo ítem, éste se convierte en una diferencia de proporciones estandarizadas, es decir, en el método de estandarización. También, el procedimiento de estandarización ha sido aplicado en el análisis del funcionamiento diferencial de distractores (FDD) en ítems de elección múltiple (Dorans, Schmitt y Bleistein, 1992) y en el estudio del DIF en ítems omitidos en tests de velocidad (Dorans y Holland, 1993). Además, recientemente se han elaborado algunas nuevas propuestas para este procedimiento. Precisamente, se han elaborado extensiones de este método para evaluar el DIF en datos politómicos (Dorans, Schmitt y Bleinstein, 1992; Dorans y Schmitt, 1993).

#### **3.4.2.3. Modelos loglineales y modelos logit.**

En su momento, la aparición de los modelos loglineales y los modelos logit supusieron un cambio de perspectiva en la detección del funcionamiento diferencial para aquellos métodos que no requerían ningún modelo de medida, debido a que desde sus inicios resultaron igualmente eficaces en la detección de los diferentes tipos de DIF (Mellenbergh, 1982).

La lógica que subyace a ambos métodos es determinar las relaciones existentes entre las variables categóricas presentadas en tablas de contingencia multidimensionales. La tabla de contingencia multidimensional para la detección del DIF incluye las siguientes variables:

H: 1,...i..., p. Nivel de habilidad que posee el sujeto que contesta al ítem

G: 1,...j..., q. Grupos comparados

R: 1,...k..., r. Respuesta al ítem

La aplicación de los mencionados análisis estadísticos permitirán poner a prueba distintas hipótesis sobre las relaciones que mantienen entre sí las diferentes variables consideradas. La formulación de dichas hipótesis se puede concretar como sigue:

Modelo 1 (DIF no uniforme)

$$\ln(F_{ijk}) = \lambda + \lambda_{H(i)} + \lambda_{G(j)} + \lambda_{R(k)} + \lambda_{HG(ij)} + \lambda_{HR(ik)} + \lambda_{GR(jk)} + \lambda_{HGR(ijk)}$$

Modelo 2 (DIF uniforme)

$$\ln(F_{ijk}) = \lambda + \lambda_{H(i)} + \lambda_{G(j)} + \lambda_{R(k)} + \lambda_{HG(ij)} + \lambda_{HR(ik)} + \lambda_{GR(jk)}$$

Modelo 3 (ausencia de DIF)

$$\ln(F_{ijk}) = \lambda + \lambda_{H(i)} + \lambda_{G(j)} + \lambda_{R(k)} + \lambda_{HG(ij)} + \lambda_{HR(ik)}$$

siendo  $F_{ijk}$  la frecuencia esperada de la respuesta k-ésima de los sujetos del grupo j-ésimo en el i-ésimo nivel de habilidad;  $\lambda$  el parámetro que describe los efectos generales y el resto de parámetros el efecto de las distintas variables y sus interacciones. Una vez realizada la formulación del modelo, el siguiente paso es la comprobación del ajuste del modelo que consiste en observar si las frecuencias esperadas (frecuencias

generadas de acuerdo con el modelo en cuestión) coinciden con las frecuencias observadas en la muestra. Además, este tipo de análisis también permite estimar los parámetros que representan las propiedades de las variables categóricas analizadas y sus relaciones mediante la descomposición lineal de los logaritmos naturales de las frecuencias esperadas (Gilbert, 1981; Knoke y Burke, 1980; Pardo y San Martín, 1994; Reynolds, 1977). Los estadísticos de bondad de ajuste más utilizados son el estadístico  $\chi^2$  de Pearson y el estadístico  $G^2$  o razón de verosimilitud (Bishop, Fienberg y Holland, 1975). Ambos estadísticos se distribuyen asintóticamente según una distribución  $\chi^2$  con los grados de libertad asociados al modelo ajustado.

Teniendo en cuenta que los modelos expuestos son modelos anidados o jerárquicos se compara el ajuste de los mismos para determinar la relevancia de los diferentes términos incluidos en ellos. Por tanto, se podrá concluir que el ítem analizado presenta DIF no uniforme si el modelo 1 es el que mejor ajuste ofrece, indicando pues que la respuesta al ítem depende de los términos de interacción  $G \times R$  y  $H \times G \times R$ . Si por el contrario, se da el caso de que el término  $H \times G \times R$  no resulta ser significativo entonces el modelo 2 será el que mejor representará los datos y, por tanto, el ítem presentará DIF uniforme. Y por último, si ninguno de los dos términos de interacción son necesarios entonces el modelo seleccionado será el 3, modelo relativo a la ausencia de DIF.

Por otro lado, si la variable respuesta al ítem se toma como variable dependiente de los efectos inducidos por otras variables y además se realiza una transformación logit del tipo  $\ln[p/1-p]$ , siendo  $p$  la

probabilidad de éxito en el ítem, entonces este tipo de modelos denominados logit pueden ser considerados un caso especial de los modelos loglineales. De esta forma, los modelos loglineales anteriores pueden ser transformados en modelos logit:

Modelo 1 (DIF no uniforme)

$$\ln(p_{ij1} / 1 - p_{ij2}) = \eta + \eta_{H(i)} + \eta_{G(j)} + \eta_{HG(ij)}$$

Modelo 2 (DIF uniforme)

$$\ln(p_{ij1} / 1 - p_{ij2}) = \eta + \eta_{H(i)} + \eta_{G(j)}$$

Modelo 3 (ausencia de DIF)

$$\ln(p_{ij1} / 1 - p_{ij2}) = \eta + \eta_{H(i)}$$

En cuanto a la interpretación de los parámetros logit, indicar que ésta se da de forma similar a los parámetros en un modelo de ANOVA, donde  $\eta$  es el efecto total de la dificultad del ítem,  $\eta_{H(i)}$  es el efecto principal de la variable habilidad de los sujetos (agrupada en intervalos),  $\eta_{G(j)}$  es el efecto principal de la variable grupo y  $\eta_{HG(ij)}$  es el efecto de la interacción habilidad x grupo. Y en relación a la comparación de modelos para probar el ajuste de cada uno de ellos, señalar también que ésta es equivalente a la expuesta anteriormente para los modelos loglineales.

#### 3.4.2.4. Regresión Logística (RL)

Los primeros autores que sugieren la utilización de la regresión logística (RL) como método estadístico para la detección del DIF fueron Spray y Carlson (1986), Bennet, Rock y Kaplan (1987) y Swaminathan y Rogers (1990). Esta técnica aparece como una alternativa eficaz ante los métodos basados en la teoría de respuesta al ítem (TRI) para detectar tanto el DIF uniforme como no uniforme en datos dicotómicos (Swaminathan y Rogers, 1990) y con posterioridad en datos politómicos (Rogers y Swaminathan, 1993; Gómez y Navas, 1996; Navas y Gómez, 1994).

La ecuación general para un modelo de regresión logística (RL) viene dada por la siguiente expresión (Hosmer y Lemeshow, 1989):

$$P_{(u=1|x,g)} = \frac{e^z}{[1 + e^z]}$$

donde

$$z = \beta_0 + \beta_1 \theta + \beta_2 g + \beta_3 (\theta g),$$

de tal modo que  $u$  es la respuesta al ítem;  $\theta$  el nivel de habilidad o competencia del individuo (p.e. puntuación total en el test);  $g$  la pertenencia a uno de los grupos, codificado con valor de 1 para el grupo de referencia y de 0 para el grupo focal, y  $\theta x g$  el producto de las dos variables mencionadas,  $\theta$  y  $g$ . Los  $\beta_0$ ,  $\beta_1$ ,  $\beta_2$  y  $\beta_3$  son los diferentes parámetros a estimar, y cada uno aporta la siguiente información:  $\beta_0$  es el parámetro *intercepto* y representa el efecto total de la dificultad del



ítem;  $\beta_1$  es el parámetro *pendiente* e indica el efecto de la variable habilidad ( $\theta$ ); el parámetro  $\beta_2$  corresponde al efecto de la variable grupo ( $g$ ) y establece la diferencia que existe entre ambos grupos en cuanto a su ejecución en el ítem; y el parámetro  $\beta_3$  representa el efecto de la interacción entre habilidad y grupo ( $\theta \times g$ ), de tal modo que si éste resulta significativo el ítem muestra más diferencias respecto al desempeño de los grupos en unos determinados niveles de habilidad que en otros.

En relación a su interpretación, diremos que un ítem mostrará *DIF uniforme* si  $\beta_2 \neq 0$  y  $\beta_3 = 0$ , y *DIF no uniforme* si  $\beta_3 \neq 0$ , independientemente que  $\beta_2$  sea o no igual a cero (Swaminathan y Rogers, 1990). Cuando ambos parámetros sean igual a cero, entonces diremos que existe ausencia de DIF.

Para la comprobación de estas hipótesis podemos elegir entre dos opciones: el estadístico de Wald o el test de la razón de verosimilitudes. Swaminathan y Rogers (1990) proponen el estadístico de Wald con el fin de poner a prueba la hipótesis nula de la simultaneidad de  $\beta_2 = 0$  y  $\beta_3 = 0$ . En concreto, este estadístico permite probar la significación de los coeficientes  $\beta$  comparando el valor estimado ( $\hat{\beta}_p$ ) para un coeficiente  $p$  con su error típico ( $SE(\hat{\beta}_p)$ ) según la siguiente expresión:

$$W = \left( \frac{\hat{\beta}_p}{SE(\hat{\beta}_p)} \right)^2$$

la cual sigue una distribución  $\chi^2$  con un grado de libertad.

El DIF también puede evaluarse comparando los cambios producidos en el ajuste del modelo a los datos cuando se elimina alguna variable. De esta manera, los modelos a comparar serán:

Modelo 1 (DIF no uniforme):  $z = \beta_0 + \beta_1\theta + \beta_2g + \beta_3(\theta g)$

Modelo 2 (DIF uniforme):  $z = \beta_0 + \beta_1\theta + \beta_2g$

Modelo 3 (ausencia de DIF):  $z = \beta_0 + \beta_1\theta$

La comparación del modelo 1 con el modelo 2 permite comprobar si la inclusión del término interacción es necesario en el modelo. La razón de verosimilitudes (RV) compara el ajuste de ambos modelos. Se calcula dividiendo la verosimilitud asociada al modelo 2 por la verosimilitud asociada al modelo 1.  $-2 \text{ Log RV}$  se distribuye como una distribución  $\chi^2$  con un grado de libertad, y puede ser utilizada para comprobar la hipótesis de que  $\beta_3=0$ . En el caso de que el término de interacción no demuestre su significación, se compararán los modelos 2 versus 3 para comprobar si la pertenencia al grupo tiene un efecto relevante sobre la respuesta al ítem. Así pues, se trata de escoger el modelo que describa los datos con el menor número de términos posible.

Como se ha comentado anteriormente, también se han propuesto algunas extensiones dentro de la técnica de regresión logística para el cálculo del DIF con datos politómicos (French y Miller, 1996; Hidalgo, 1995; Hidalgo y Gómez, 1996; Miller y Spray, 1993; Rogers y Swaminathan, 1993; Wilson, Spray y Miller, 1993). Sin embargo, la principal desventaja de su cálculo con datos politómicos es su complejidad de cálculo e interpretación de

resultados, la cual resulta bastante más difícil. Dada su complejidad, el procedimiento estadístico que parece que podría sustituirle en la detección de DIF para datos politómicos es el *análisis discriminante logístico* (logistic discriminant function analysis) propuesto por Miller y Spray (1993). Precisamente, en su estudio de simulación donde se compararon varios métodos como la regresión logística, el método MH y el análisis discriminante logístico, este último resultó ser tan eficaz como la regresión logística en la detección del DIF uniforme y no uniforme, además de resultar ser más parsimonioso (sólo se requería una ecuación de regresión por ítem) y menos costoso computacionalmente (Miller y Spray, 1993). Por último, señalar que en la actualidad algunos estudios han aportado evidencia empírica acerca una posible aplicación de la regresión logística a la Teoría Multidimensional del sesgo (Mazor, Kanjee y Clauser, 1995).

#### **3.4.2.5. EL SIBTEST**

El SIBTEST es uno de los más recientes y prometedores procedimientos para detectar tanto el funcionamiento diferencial de los ítems como el funcionamiento diferencial del test (Shealy y Stout, 1993a, 1993b). Tal como su propio nombre indica (SIB "*simultaneous item bias*") es un método que permite estudiar la presencia de DIF simultáneamente en varios ítems de un determinado test. La principal ventaja de comprobar la presencia de DIF en varios ítems a la vez reside en la posibilidad de determinar si los ítems con DIF actúan conjuntamente afectando de manera diferencial a las puntuaciones totales del test, según los sujetos pertenezcan a un grupo u otro; dicho de otro modo, si realmente provocan un funcionamiento

diferencial del test (FDT), fenómeno conocido como *amplificación*. En cambio, al fenómeno contrario se le denomina *cancelación*, el cual se produce cuando los efectos de DIF presentes en varios items a la vez se anulan unos con otros no llegando a provocar FDT. Ejemplos de la aparición de la *amplificación* y *cancelación* pueden encontrarse en los trabajos de Drasgow (1987), Roznowski (1987) y Nandakumar (1993b).

En concreto, la base teórica del SIBTEST se encuentra ubicada en un modelo de TRI multidimensional no paramétrico, formulado por Shealy y Stout (1993a, 1993b) para explicar el sesgo del test y cuyos supuestos coinciden en parte con los defendidos por la *Teoría Multidimensional del sesgo (TMS)*.

Para una mejor comprensión del método, seguidamente vamos a exponer brevemente la lógica que subyace a este procedimiento. La teoría multidimensional del sesgo (TMS) señala que los items de un test pueden mostrar DIF cuando, además de la habilidad principal que el test pretende medir ( $\theta$ ), están midiendo otras habilidades adicionales ( $h_1, \dots, h_n$ ). Por lo tanto, los items que potencialmente pueden mostrar DIF serán aquellos que están midiendo otras habilidades complementarias además de la habilidad principal ( $\theta$ ), objeto de medición del test, y los items sin DIF serán aquellos que sólo miden las diferencias de los sujetos examinados en la habilidad objetivo del test. Los items del primer grupo, aquellos con posibilidad de DIF, formarán el *subtest estudiado*, y los items del segundo grupo compuesto por items sin DIF constituirán el denominado *subtest válido*. Para poder detectar el funcionamiento diferencial de los items en la práctica, primero habrá que determinar qué items conforman el subtest

válido, esto es, qué parte del test mide la habilidad principal. Así se podrá igualar a los sujetos de ambos grupos en esa habilidad. Por ejemplo, si tenemos tests suficientemente largos el requisito necesario de formar un subtest que sólo mida la habilidad principal puede sustituirse por el requisito de formar un subtest válido pero esencialmente unidimensional. Este término va ligado a los conceptos de "*dimensionalidad esencial e independencia esencial*" los cuales emergen como una posible solución ante los fuertes supuestos de unidimensional e independencia local defendidos en la TRI (Stout, 1987, 1990; Nandakumar y Stout, 1993; Nandakumar, 1994; Nandakumar y Yu, 1996). Para determinar qué items cumplen el supuesto de la "unidimensional esencial" y por tanto proceder a su inclusión en el subtest válido, podemos recurrir a procedimientos estadísticos que sirvan para identificar a los items que no cumplen el supuesto de unidimensionalidad como por ejemplo el DIMTEST (Stout, 1987; Stout, Nandakumar, Junker, Chang y Steidinger, 1991). Así pues, los items que cumplan el supuesto de "unidimensional esencial" constituirán el subtest válido y el resto de items del test formarán el subtest estudiado. Por otra parte, se asume que en los tests suficientemente largos los sujetos con las mismas puntuaciones en el subtest válido tienen aproximadamente el mismo nivel en la habilidad principal ( $\theta$ ) y por tanto, sus puntuaciones son comparables. De acuerdo con esta lógica, los miembros de ambos grupos, focal y de referencia, son agrupados por su puntuación total en el subtest válido y posteriormente se compara su ejecución en el subtest estudiado.

La detección del DIF con este procedimiento suele adoptar una hipótesis nula de ausencia de funcionamiento diferencial entre grupos del

conjunto de items que componen el subtest a estudiar, quedando representada de la siguiente manera:

$$H_0: \beta_u = 0$$

$$H_1: \beta_u \neq 0$$

siendo  $\beta_u$  un parámetro que permite estimar la *cantidad y dirección* del FDT. Es preciso destacar que por test aquí se entiende al conjunto de items que componen el subtest bajo estudio. El estadístico de contraste para comprobar la hipótesis nula es:

$$B = \frac{\hat{\beta}_u}{\hat{\sigma}(\hat{\beta}_u)}$$

donde

$$\hat{\beta}_u = \sum_{k=0}^n \left[ \hat{P}_{Fk} (\bar{Y}_{Rk} - \bar{Y}_{Fk}) \right] \quad (3.7)$$

siendo  $\hat{P}_{Fk}$  la proporción de sujetos del grupo focal que obtienen una puntuación igual a  $k$  en el subtest válido.  $\bar{Y}_{Rk}$  y  $\bar{Y}_{Fk}$  las medias en el subtest estudiado de los sujetos con una puntuación igual a  $k$  en el subtest válido, para el grupo de referencia y focal respectivamente. En el caso de estudiar un sólo ítem, la ecuación 3.7 se convierte en una diferencia de proporciones estandarizadas, es decir, en el procedimiento de estandarización, y por tanto el estudio del FDT se transforma en el estudio del DIF. Para más información sobre los detalles de su aplicación puede consultarse Shealy y Stout (1993a, pp. 159-194).

La distribución muestral de  $B$  es exacta o aproximadamente normal con  $N(0,1)$ . Por tanto, si se cumple que  $B \leq Z_{\alpha/2}$  ó  $B \geq Z_{1-\alpha/2}$  rechazamos la hipótesis nula de ausencia de un funcionamiento diferencial de los items. Naturalmente, también se puede someter a contrastación la hipótesis de DIF/DFT unidireccional contra un grupo. Tras su aplicación, los valores positivos indicarán que los items son más fáciles para el grupo de referencia, y los valores negativos que los items evaluados son más fáciles para el grupo focal.

En un principio, el SIBTEST fue un procedimiento diseñado para detectar el FDT/DIF uniforme, por tanto su principal limitación era su incapacidad para detectar el FDT/DIF no uniforme. Sin embargo, a pesar de que los estudios de simulación desconocían su comportamiento bajo condiciones de DIF no uniforme, éstos auguraban buenos resultados en base a los estudios realizados con la detección del FDT/DIF uniforme (Ackerman y Evans, 1994; Nandakumar, 1993; Roussos y Stout, 1996b; Shealy y Stout, 1993a). En este sentido, Narayanan y Swaminathan (1994) en un estudio de simulación en el que comparan el SIBTEST frente al procedimiento MH ofrecen las siguientes conclusiones: 1) cuando la distribución de la habilidad es la misma en ambos grupos, el SIBTEST es tan eficaz para detectar el DIF uniforme como el procedimiento MH, pero el SIBTEST es mejor en la detección del DIF cuando existe impacto, es decir, cuando las distribuciones son desiguales; 2) respecto a la tasa de error tipo 1, el método MH se encuentra dentro de los límites nominales establecidos, mientras que en el SIBTEST son ligeramente más altas de lo esperado; 3) en cuanto al tamaño muestral, ambos procedimientos se comportan aceptablemente bien con

muestras tan bajas como 300 examinados en el grupo de referencia y 100 en el grupo focal; 4) en ambos procedimientos, la tasa correcta de detección de DIF se incrementa a medida que el tamaño o magnitud del DIF también aumenta, y por último 5) en relación a las características del ítem, cabe destacar que el poder estadístico de ambos métodos aumenta en los ítems con valores elevados en discriminación y bajos/medio-bajos en dificultad.

En cuanto a sus *limitaciones*, es importante destacar que el método SIBTEST únicamente se puede aplicar a tests con más de 25 ítems y que además no presenten impacto. Cuando no se cumple alguna de estas condiciones, que es la situación más frecuente, se debe utilizar un procedimiento diferente al presentado, pero que por su complejidad hemos omitido. En la actualidad, disponemos de un programa informático que implementa ambos procedimientos (Shealy, Stout y Roussos, 1995). Recientemente se han desarrollado algunas modificaciones del método SIBTEST. Una de ellas permite evaluar el FDT/DIF no uniforme en ítems dicotómicos (Crossing SIBTEST; Li y Stout, 1994a, 1994b). De hecho, algunos trabajos publicados hace poco ofrecen apoyo empírico del SIBTEST para la detección de DIF no uniforme (Narayanan y Swaminathan, 1996). Otra modificación permite su aplicación en la evaluación del FDT/DIF para ítems politómicos (Poly SIBTEST; Chang, Mazzeo y Roussos, 1996). Y la más reciente ofrece la posibilidad de aplicar este modelo en un enfoque multidimensional de la TRI (MULTISIB; Stout, Li, Nandakumar y Bolt, 1997).



### 3.4.3. Desarrollos actuales y futuros

A finales de los 70 y principios de los 80, un número de investigadores empiezan a elaborar una nueva teoría formal en términos de dimensionalidad, derivando todos sus esfuerzos en la formulación de los nuevos *modelos multidimensionales* en TRI (MMTRI). En sus inicios, además del trabajo de Reckase (1972) sobre el modelo de Rasch multidimensional, los estudios realizados por Mulaik (1972), Simpson (1978) y Whitely (1980) son los encargados de aportar nuevos modelos acerca de las interacciones existentes entre personas e items, y donde la mayoría de ellos suele aludir a la multidimensionalidad como la principal causa de la aparición del sesgo de los items.

Sin embargo, en la actualidad la situación que impera es bastante diferente, la *Teoría multidimensional del sesgo* se presenta como una alternativa muy ambiciosa en los recientes estudios basados en la detección del DIF (Ackerman, 1994; Camilli, 1992; Kok, 1988; Roussos y Stout, 1996b; Shealy y Stout, 1993a, 1993b), aún cuando no se posean resultados totalmente satisfactorios debido a que algunos errores en la estimación de los parámetros pueden conducir equivocadamente a la aparición del DIF en determinados items (Reckase, 1979, 1985; Reckase y McKinley, 1991; Roussos y Stout, 1996b).

El elemento central de esta teoría se basa en establecer una distinción entre la *habilidad principal*, aquella que pretende medir el test, y las *habilidades secundarias o espurias*, aquellas variables que no se pretenden medir pero que están siendo medidas por el test y que, por

tanto, pueden influir de forma decisiva en las respuestas a algunos de los items del test. Dentro de esta teoría, la aparición del DIF tiene lugar si se cumplen una serie de condiciones, siendo éstas las siguientes: que dos grupos distintos pero con el mismo nivel de habilidad tengan diferentes distribuciones en las habilidades que no se pretenden medir, que los items del test sean capaces de medir esas múltiples dimensiones, y que la estimación de esas habilidades se concrete en una única medida (p.e. puntuación en el test). Por lo que respecta a la diferenciación entre DIF uniforme y no uniforme, Ackerman (1992) señala que algunas diferencias en los parámetros de las distribuciones condicionales de la variable o habilidad espuria para un nivel concreto de la habilidad principal son causas potenciales del DIF uniforme y no uniforme. En concreto, señala que puede existir DIF uniforme si: (a) los grupos tienen diferentes medias en la habilidad principal (impacto), siempre que haya una correlación significativa entre la habilidad principal y la espuria, o (b) si existen diferencias significativas en las medias en la variable espuria en ambos grupos. Por su parte, el DIF no uniforme puede ocurrir si: (a) la varianza de la habilidad espuria no es la misma entre ambos grupos, ó (b) si la magnitud de la correlación entre la habilidad principal y la espuria no es igual entre ambos grupos.

Para aclarar estos conceptos pongamos un ejemplo. Supongamos que tenemos un test compuesto por 45 items creado para medir la aptitud numérica, que se ha aplicado a una muestra de 1000 personas, de las cuales 500 son sujetos nativos y 500 sujetos no nativos (inmigrantes, lengua materna distinta,...). Si el test administrado midiese la variable objeto

de estudio sin ningún tipo de error y además fuese válido, las personas con el mismo nivel de competencia en aptitud numérica tendrían la misma puntuación total en el test, con independencia de las posibles características sociodemográficas en la que pudiesen variar los individuos mencionados. Del mismo modo, la probabilidad de acertar con éxito cada uno de los items que componen el test sería función exclusivamente del nivel de habilidad alcanzado por los sujetos examinados, y no de cualquier otra variable adicional. En estas condiciones, tanto el test como los items que lo componen estarían completamente insesgados. Ahora bien, supongamos que los sujetos nativos por término medio poseen mayor comprensión lectora que los no nativos, y que alguno de los items del test no sólo mide la aptitud numérica sino que además se requiere una buena comprensión lectora para realizar algunos de ellos correctamente. En este caso entonces la probabilidad de acertar con éxito uno de estos items del test dependería no sólo de la aptitud numérica del sujeto sino también de su comprensión lectora, y debido a que los nativos poseen por término medio mayor nivel de comprensión lectora que los emigrantes, la probabilidad de acertar dichos items, a igual aptitud numérica, será mayor para los nativos. En caso de que esta suposición se confirme, las puntuaciones totales del test también se verán afectadas, y entonces las medidas obtenidas en el test para ambos grupos estarán sesgadas, esto es, que los items que forman parte del test mostrarán un funcionamiento diferencial para los dos grupos sometidos a comparación.

Por otra parte, sería conveniente señalar que éstas no son las únicas causas posibles de la presencia de DIF en los tests, las causas sobre la aparición del DIF en el ámbito aplicado de la medición y evaluación con

tests psicométricos pueden ser varias, sin embargo sí hay que destacar que la Teoría Multidimensional del sesgo es la única teoría formal desarrollada hasta el momento acerca de las causas del DIF, y además siguiendo los últimos datos aportados por algunos estudios muy recientes (Camilli, Wang y Fesq, 1995; Miller y Hirsch, 1992), todo indica que éstos se ajustan fácilmente a la mayoría de las situaciones en las que los items de un test poseen DIF.

Por otro lado, se encuentra la formulación de modelos de TRI semiparamétricos o no paramétricos. Precisamente, uno de los últimos y recientes procedimientos propuestos para detectar el DIF, el SIBTEST es una prueba de ello. Como se ha visto en su anterior presentación, este procedimiento se basa en un modelo multidimensional no paramétrico de TRI (Shealy y Stout, 1993a, 1993b). En definitiva, todo parece indicar que esta nueva visión global del sesgo de los items que tiene como base central *la dimensionalidad* de los tests psicológicos constituye un marco teórico adecuado para afrontar el futuro de los estudios sobre el funcionamiento diferencial de los items.

### **3.5.RELEVANCIA DEL SESGO EN LA INVESTIGACION EMPIRICA**

Una vez presentados los principales métodos estadísticos utilizados en la detección del DIF, se procederá a revisar la parte empírica de este área de estudio. Es importante destacar aquí que la gran cantidad de literatura existente en torno al estudio del DIF se ha centrado principalmente en conocer bajo qué condiciones y con qué método estadístico

se detecta adecuadamente la presencia del DIF en los items del test, y sólo para situaciones reales muy específicas se ha pasado al terreno de la práctica. Dicho esto, queda claro que la utilización de los métodos que detectan el DIF en el ámbito aplicado se encuentra muy reducida limitándose en especial a los contextos educativos y/o aquéllos donde casi siempre existe la predominancia de más de una lengua, esto es, en situaciones de contacto de lenguas. Así pues, dado que el foco de interés de nuestro estudio se sitúa en el ámbito educativo de una comunidad bilingüe, a continuación se ofrecerá una revisión de aquellos estudios empíricos realizados en contextos de naturaleza bilingüe sobre la detección del DIF en los items de un test o instrumento de medida.

Los estudios empíricos que vamos a presentar han sido agrupados según su principal objetivo en dos categorías diferentes: a) la identificación de items con DIF tras la aplicación de un instrumento de medida, en su versión original y adaptada, en contextos de naturaleza monolingüe y bilingüe, y b) la identificación de las causas o factores responsables del DIF en el test.

En la *primera* categoría, encontramos el estudio realizado por Candell y Hulin (1987). Este trabajo presenta los resultados obtenidos tras un análisis de las propiedades psicométricas de una traducción al francés canadiense del test *Job Descriptive Index* (JDI: Smith, Kendall y Hulin, 1969), un instrumento de medida compuesto por 5 subescalas diferentes que permiten medir el grado de satisfacción de los sujetos en el ámbito laboral. Los análisis realizados tuvieron como objetivo el estudio de la equivalencia métrica de la versión traducida respecto a su versión original, es decir, estudiar si realmente los items de cada una de las

versiones, original y traducida, medían lo mismo en muestras con características lingüísticas y/o culturales diferentes. La muestra de este estudio estaba compuesta por sujetos de clase trabajadora procedentes de dos países diferentes: Canadá y Estados Unidos. La administración de ambas pruebas se realizó de la siguiente manera: la versión original del test, la versión inglesa, se aplicó a tres muestras diferentes, a) a una muestra estadounidense monolingüe (N:486), b) a una muestra canadiense monolingüe inglesa (N:435), y c) a una muestra canadiense bilingüe cuya primera lengua es el francés pero que eligió por propia voluntad la versión inglesa del test (N:226); mientras que la versión traducida del test, la versión francesa, se administró únicamente a una muestra canadiense bilingüe, cuya primera lengua era el francés y además la versión seleccionada por el sujeto era la francesa (N:349).

Respecto a la metodología utilizada, se aplicó un método de comparación de parámetros como es el estadístico chi-cuadrado de Lord (1980) basado en la Teoría de Respuesta al Item (TRI). Es convenientes señalar que la longitud del test varía en función de: a) el ajuste de los datos al modelo de TRI utilizado, esto es, aquellos items que no ajustan adecuadamente al citado modelo son eliminados del test, y b) la ampliación de 20 items adicionales a la versión del test aplicado a la muestra canadiense. Los resultados obtenidos se resumen a continuación. Para la primera comparación (aplicación de la versión inglesa a dos muestras canadienses con lengua nativa diferente: inglés (muestra anglófona) y francés (muestra francófona) los datos aportados mostraron que ninguno de los 89 items que formaban parte del test presentaba un funcionamiento

diferencial (DIF), únicamente dos de ellos se encontraban cercanos al nivel de significación estadística utilizado por el investigador ( $\alpha=0.001$ ). La segunda comparación (aplicación de ambas versiones a una muestra canadiense francófona) identificó 6 items con DIF de un total de 86. La tercera comparación (aplicación de las dos versiones de la prueba -inglés y francés- a una muestra canadiense anglófona y francófona respectivamente) dió lugar a la presencia de DIF en 15 de 86 items. Las siguientes comparaciones: cuarta (aplicación de la versión inglesa a una muestra americana y canadiense anglófona) y quinta (aplicación de la versión inglesa a una muestra americana y canadiense francófona) mostraron la existencia de 11 items con DIF de un total de 60 para la muestra americana y canadiense anglófona, y 9 items con DIF de 60 para la muestra americana y canadiense francófona, de los cuales 7 de ellos coincidieron en ambas comparaciones. Y por último, la sexta comparación (aplicación de las dos versiones -inglés y francés- a una muestra americana y canadiense francófona respectivamente) presentó 11 items con DIF de un total de 58 items. En la siguiente tabla 1.1 se sintetiza toda la información relativa a las distintas comparaciones realizadas junto con los resultados obtenidos.

A partir de los resultados obtenidos, el citado estudio concluyó que la presencia de items con funcionamiento diferencial en la versión inglesa del test analizado tras comparar la muestra americana con la muestra canadiense anglófona o francófona (comparaciones 4 y 5), supone un serio obstáculo para llevar a cabo comparaciones interculturales adecuadas con una única versión del test analizado. Por lo que respecta a la comparación de la muestra canadiense francófona en las dos versiones utilizadas del

test (comparación 2), los resultados conseguidos indicaron un descenso en el número de items identificados con DIF, lo que derivó en algunos supuestos acerca de la discrepancia existente en la ejecución realizada por sujetos monolingües y bilingües. Los citados autores aportaron como posible explicación a este hecho la posibilidad de una estructura cognitiva semántica distinta en sujetos monolingües y bilingües, la cual podía ser la causa de una reducción del DIF en la comparaciones realizadas con sujetos bilingües. Esta justificación también es aplicable para la comparación 1 (ausencia de DIF), donde la versión inglesa del test es aplicada a una muestra monolingüe canadiense anglófona y bilingüe canadiense francófona.

**Tabla 3.2.** Resumen de las comparaciones y resultados conseguidos en el estudio.

N° COMPARACION	MUESTRA	VERSIONES DEL TEST APLICADO	RESULTADOS
01	CANADIENSE ANGLOFONA CANADIENSE FRANCOFONA	VERSION INGLESA	89 Items DIF: 0 (0 %)
02	CANADIENSE FRANCOFONA	VERSION INGLESA VERSION FRANCESA	86 Items DIF: 6 (7%)
03	CANADIENSE ANGLOFONA CANADIENSE FRANCOFONA	VERSION INGLESA VERSION FRANCESA	86 Items DIF: 15 (17,5%)
04	AMERICANA CANADIENSE ANGLOFONA	VERSION INGLESA	60 Items DIF: 11 (18,3%)
05	AMERICANA CANADIENSE FRANCOFONA	VERSION INGLESA	60 Items DIF: 9 (15,0%)
06	AMERICANA CANADIENSE FRANCOFONA	VERSION INGLESA VERSION FRANCESA	58 Items DIF: 11 (19%)



Por último, en las comparaciones de ambas versiones en muestras lingüística y/o culturalmente diferentes (comparación 3 y 6), los autores de este estudio concluyeron que la realización de comparaciones interculturales con versiones diferentes de un mismo test debe realizarse con cierta precaución, dada la posible presencia de algunos items con un funcionamiento no equivalente en muestras lingüística y/o culturalmente diferentes.

Los próximos estudios a considerar son los de Hulin y colaboradores (Hulin, Drasgow y Komocar, 1982; Drasgow y Hulin, 1986). Estos se sitúan en la misma línea de investigación que la anterior. Su principal objetivo es la evaluación de la equivalencia métrica mediante un análisis del DIF en una traducción del inglés al castellano del cuestionario *Job Descriptive Index* (JDI: Smith, Kendall y Hulin, 1969). El tipo de muestra utilizada en este estudio estuvo formada por sujetos monolingües y bilingües, de tal modo que cada versión se administró a su correspondiente muestra monolingüe, mientras que ambas (inglés y castellano) se aplicaron a una muestra bilingüe. La muestra bilingüe estuvo compuesta por 178 sujetos, procedentes de Cuba o Puerto Rico pero residentes en Estados Unidos, a los cuales se le administró las dos versiones en un orden contrabalanceado dentro de un intervalo de 30 días (español-inglés y inglés-español). Por su parte, la muestra monolingüe estuvo formada por sujetos americanos que contestaron la versión americana y por sujetos mejicanos que cumplimentaron la versión española. La metodología utilizada también consistió en la aplicación del estadístico chi-cuadrado de Lord (1980). Los resultados obtenidos mostraron que la aplicación de ambas versiones a una muestra bilingüe compuesta por sujetos hispano-americanos, sólo presentó el 4% del

total del test con DIF; mientras que la aplicación de cada versión a su respectiva muestra monolingüe -hispanos y americanos- provocó un aumento considerable de la tasa de items con DIF, ascendiendo a un 30% del total. La obtención de estos resultados llevó a los autores de este estudio a informar de la obligación de tomar ciertas preocupaciones acerca de las posibles generalizaciones de los sujetos bilingües a los monolingües. A su vez, este resultado también permitió ofrecer evidencia empírica de que los sujetos bilingües, en virtud de sus habilidades lingüísticas a la hora de contestar los items de un determinado instrumento de medida, poseen estructuras cognitivas diferentes respecto a sus homólogos grupos monolingües.

En este mismo campo, se encuentra otro trabajo del mismo autor (Hulin y Mayer, 1986) pero utilizando únicamente muestras monolingües. Este estudio es prácticamente una replicación del estudio anterior, excepto que la versión del test analizado (JDI; Smith, Kendall y Hulin, 1969) es una versión traducida del inglés al hebreo. La administración de ambas versiones se efectuó a sus correspondientes muestras monolingües, americana (N:500) e israeli (N:308). Y tras aplicar el estadístico chi-cuadrado de Lord (1980) para detectar aquellos items con posible funcionamiento diferencial (DIF), los resultados mostraron en la comparación intercultural de ambos grupos una presencia de DIF en el 30 % del total del test.

La próxima investigación empírica que analizaremos es la realizada por Budgell, Raju y Quartetti (1995). El objetivo central de este estudio es la evaluación de la equivalencia métrica de dos tests de aptitud

traducidos al francés y aplicados a sujetos procedentes de un mismo país pero residentes en ámbitos lingüísticos diferentes: canadienses anglófonos y canadienses francófonos. Los tests utilizados fueron una prueba de razonamiento cognitivo compuesta por 18 items, y otra prueba de habilidad numérica formada por 15 items. Cada una de ellas en su versión original en inglés y traducida en francés fueron aplicadas a sujetos cuya lengua nativa y dominio lingüístico del entorno eran el inglés o el francés respectivamente. Dado el elevado tamaño muestral conseguido para cada una de las versiones, la muestra se dividió en dos partes, de tal modo que se obtuvieron dos submuestras canadienses anglófonas formadas por 1000 sujetos cada una (muestra 1 y 2) y dos submuestras canadienses francófonas compuestas también por 1000 sujetos (muestras 3 y 4), posibilitando así una validación cruzada de la misma comparación. La metodología empleada para detectar la presencia de items con funcionamiento diferencial consistió en aplicar el estadístico de Lord (LR; Lord, 1980), las medidas del área exactas con signo (AECS; Raju, 1988, 1990) y sin signo (AESS; Raju, 1988, 1990) y el método de MH (Holland y Thayer, 1988). Los resultados obtenidos para la prueba de aptitud numérica mostraron los siguientes porcentajes de DIF en cada uno de los métodos utilizados: 27% (LR), 27% (AECS), 40% (AESS) y 47% (MH) para la primera comparación realizada con las muestras 1 y 3, y el 33%, 33%, 40% y 40% para la segunda comparación formada por las muestras 2 y 4. En cuanto a la prueba de razonamiento cognitivo, los porcentajes de DIF presentados para cada método fueron de 44%, 56%, 44%, 55% en la primera comparación, y 33%, 11%, 39% y 44% en la segunda. Atendiendo a estos resultados, se puede apreciar un elevado grado de consistencia en la detección de DIF para todos los métodos, a excepción de la medida de área sin signo (AESS) en la prueba de razonamiento cognitivo (11-56%). En

relación al grado de correspondencia en la identificación de items con DIF para cada método, el método MH fue el método que más items con DIF identificó, luego le siguió el estadístico de Lord, la medida del área AECS y, finalmente, AECS. Al respecto, los autores del citado estudio señalan que la elevada tasa de detección de DIF en el método MH quizás sea producida por un aumento del error tipo 1, a causa de la escasa longitud total del test analizado (menor de 20 items).

El último trabajo presentado dentro de esta primera categoría se localiza dentro del territorio español, concretamente en el País Vasco (Elosua, 1996). La autora de este estudio analizó la equivalencia métrica de dos pruebas de inteligencia, Habilidad mental verbal (MHV) y Habilidad mental no verbal (HMnV), en su versión original en castellano y la versión adaptada al euskera. Las muestras analizadas estuvieron compuestas por un grupo monolingüe castellano (N:651), el cual cursaba su enseñanza en castellano y por tanto se le aplicó la versión de la prueba en castellano, y un grupo bilingüe vasco-castellano (N:1155), escolarizado en euskera, y al que se le aplicó la versión en euskera. Tras la aplicación de distintos métodos y técnicas estadísticas basados en la detección de items con funcionamiento diferencial (DIF) en los dos grupos comparados, los resultados conseguidos mostraron que la primera prueba (MHV) con alto contenido verbal presentaba 29 items con DIF de un total de 40 items, el 72,5% del test, mientras que la segunda prueba (HMnV) sin contenido verbal sólo presentaba 7 items con DIF de un total de 40 (17,5%). Los resultados obtenidos permitieron concluir a su autora que la evaluación realizada en el proceso de adaptación sobre la calidad lingüística al

euskera de las citadas pruebas psicopedagógicas no garantiza por sí misma la equivalencia métrica de dichas versiones. Para conseguir la equivalencia de las mismas, ésta señala que es necesario un análisis del sesgo de sus items dentro del ámbito de estudio de la validez mediante la aplicación de métodos estadísticos basados en la detección del DIF en los items, así como el empleo de métodos de juicio que puedan dar una explicación acerca del mismo. Tras su aplicación, los resultados conseguidos mostraron que las puntuaciones obtenidas en las dos versiones del test analizadas no podían ser comparables entre sí, dada la elevada presencia de items con funcionamiento diferencial en ambas, tras ser aplicadas a grupos lingüísticamente diferentes tanto en su contexto social más inmediato como escolar.

En cuanto a los estudios referidos a la identificación de los factores responsables de la aparición del funcionamiento diferencial de los items, la revisión que presentaremos aquí se va a limitar a aquellos estudios cuyo objetivo es la búsqueda de las causas responsables del DIF en los items de un determinado test aplicado a diferentes grupos minoritarios. La metodología utilizada por estos estudios se centra básicamente en métodos basados en la TRI, Mantel-Haenszel (MH) y Estandarización (DPE), y los instrumentos de medida analizados han sido: el "American College Testing Program Assessment Test" (ACT), el "Graduate Management Admission Test" (GMAT), el "Graduate Record Examinations" (GRE) General Test, el "NIE Core Battery" (NIE) y el "Scholastic Aptitude Test" (SAT). En ellos, las diferentes escalas o pruebas que los componen se pueden clasificar en dos categorías: tests verbales y tests de aptitud numérica. Para una mejor comprensión, la exposición que ofrecemos seguirá dicha clasificación.

Dentro de la primera categoría referida a los **tests verbales**, la investigación realizada muestra que una de las causas más importantes en la aparición de items con DIF para un determinado test reside en la presencia de ciertos temas objeto de análisis como en la inclusión de tipos de items específicos como son la presencia de sinónimos, antónimos, analogías verbales, etc...

En concreto, los estudios realizados con el test GRE mostraron que los sujetos de raza negra presentaban un peor rendimiento en aquellos items con cierto contenido científico o relativo a las ciencias que los de raza blanca. Sin embargo, estos mismos resultados para los sujetos de raza negra no se consiguieron replicar en los tests SAT y NTE. Otra característica que ha recibido bastante apoyo empírico en la literatura analizada ha sido la superioridad de los individuos negros e hispanos frente a los blancos en items con contenido de interés para los citados grupos minoritarios. Precisamente, la obtención de este resultado ha sido confirmado en los siguientes tests: GRE, GMAT, NTE y SAT (Carlton y Harris, 1989b; McPeck y Wild, 1986; Medley y Quirk, 1974; Wild y McPeck, en preparación). En el test SAT, los negros presentaban mejor rendimiento que los blancos en items con algún referente sobre los americanos negros (Carlton y Harris, 1989a), y los hispanos ofrecían puntuaciones más altas que los blancos en items referidos a temas relacionados con la mujer americana-mexicana o profesores de raza negra (Schmitt, 1986). En definitiva, los resultados aportados denotan que si el ítem administrado tiene como materia de contenido un tema de relevancia especial para el grupo minoritario en cuestión, entonces éste

presentará mejores puntuaciones en el ítem que su correspondiente grupo de referencia.

Sin embargo, la característica más estudiada en esta categoría de tests ha sido el posible funcionamiento diferencial de los ítems ante la presencia de analogías verbales (e.g., Carlton y Harris, 1989b; O'Neill, McPeck y Wild, en preparación; Pearlman, 1987; Schmitt, 1986; Schmitt y Bleistein, 1987; Schmitt y Dorans, 1987a, 1987b; Wendler y Carlton, 1987). Los distintos estudios realizados en este campo han señalado que los peores resultados en este tipo de ítems, las analogías, se pueden apreciar con claridad en los negros e hispanos comparados con los blancos. Además, estos estudios señalan que la presencia de palabras homógrafas en dichos ítems pueden ser la principal causa de la aparición del DIF en las analogías. El término palabras homógrafas se refiere a aquellas palabras que se escriben igual pero con significado diferente. Los análisis realizados en los tests SAT y GRE mostraron que los sujetos negros poseían peor rendimiento que los blancos en aquellas analogías vinculadas a temas afines a las ciencias y relaciones humanas. Sin embargo, Freedle y Kostin (1987, 1988) en su estudio, en contra de los resultados esperados, encontraron que las analogías de contenido científico apenas contribuían como variable independiente en la predicción de DIF en los ítems del test; por el contrario, la dificultad del ítem y aspectos relacionados con la construcción de las analogías fueron los factores identificados como los mejores predictores de la presencia del DIF.

Otra característica que también provocó la aparición de un funcionamiento diferencial en los ítems entre el grupo de referencia

(blancos) y el grupo focal (hispanos) fue la presencia de *true cognates* (palabras con la misma raíz morfológica en inglés y español, y significados similares en ambos idiomas) y *false cognates* (palabras con la misma raíz morfológica en inglés y español, pero con significados diferentes en ambos idiomas) en los items del test. Precisamente, Schmitt (1985, 1988) y Rogers y Kulick (1987) tras realizar una comparación de la ejecución presentada por diferentes grupos minoritarios en tests de rendimiento mostraron que: a) los sujetos hispanos residentes en EEUU presentaban algunas ventajas (DIF positivo) ante items con *true cognates* y cuyo contenido era de interés especial para el grupo evaluado, mientras que los items con presencia de *palabras homógrafas* y *false cognates* mostraban DIF negativo en estos mismos grupos. Ante estos resultados, Freedle y Kostin (1987) realizaron un estudio para investigar si ciertos patrones de DIF operaban dentro de cada tipología de ítem en función del nivel de dificultad del citado ítem. Tras comparar sujetos de raza blanca con sujetos de raza negra, los resultados obtenidos mostraron, que existía una correlación positiva entre la dificultad del ítem y la aparición del DIF en las diferentes categorías de items incluidas en los tests verbales del GRE y del SAT, pero no en el GMAT, siendo las correlaciones más elevadas para las categorías referidas a analogías y antónimos. Además, se observó que, en sentido contrario a lo esperado, los sujetos negros mostraban una cierta superioridad ante los blancos en los items más difíciles y los blancos sobre los negros en los items más fáciles. La obtención de estos resultados permitieron a los autores de este estudio (Freedle y Kostin, 1988) sugerir que el nivel contextual asociados a los items (muy bajo en analogías y antónimos y el más alto en items de comprensión lectora) podía ser un



importante modulador de la magnitud de las relaciones encontradas entre la dificultad del ítem y la presencia de DIF. Así pues, atendiendo a dicho factor es predecible que la magnitud de la correlación entre el DIF y dificultad del ítem sea más grande para los items con un mínimo grado contextual que para aquellos con un mayor contenido contextual??.

En cuanto a los **tests de aptitud numérica**, algunos resultados mostraron que, en términos generales, los sujetos negros e hispanos tendían a obtener mejores puntuaciones en items de álgebra que sus homólogos de raza blanca, a excepción del test SAT (O'Neill, McPeek y Wild, en preparación; O'Neill, Wild y McPeek, en preparación). De esta forma, la aparición de este resultado sugiere que la inclusión de símbolos en los items, propio de los items de álgebra, favorece al desempeño de los sujetos negros e hispanos frente al de los blancos. Otro resultado descubierto señaló que los items con algún contenido geométrico resultaban ser más fáciles para los hispanos que para los blancos, mientras que dichas diferencias no se constataban cuando la comparación se realizaba entre negros y blancos (O'Neill, McPeek y Wild, en preparación; O'Neill, Wild y McPeek, en preparación).

Otra característica relevante a la hora de evaluar la presencia del DIF en los items de contenido matemático es el tipo de presentación utilizada para la formulación de determinados problemas matemáticos. Los resultados aportados en este campo parecen indicar que la elaboración de algunos problemas de matemáticas en términos lingüísticos muestran con claridad la presencia de diferencias significativas entre sujetos blancos y negros. En concreto, Shepard, Camilli y Williams (1985) y Camilli y Smith

(1990) encontraron que los items de contenido matemático que poseían ciertos componentes verbales resultaban ser más fáciles para los sujetos del grupo dominante que para los del grupo minoritario. En esta línea, el estudio de Ryan (1991) demostró que los conocimientos previos en comprensión lectora eran un factor determinante en la aparición de DIF en algunos tests de aptitud matemática. Por otra parte, Dorans, Schmitt y Bleistein (1992) indicaron que la presencia del DIF aumentaba especialmente cuando la parte verbal del test aparecía situada al principio del test y, sobre todo, cuando los grupos minoritarios eran negros e hispanos.

Dentro de este ámbito, estudios adicionales (O'Neill, McPeck y Wild, 1989; Shepard, Camilli y Williams, 1984) encontraron los mismos resultados tras comparar el rendimiento de los americanos (grupo de referencia) frente al de los hispanos (g. focal) y asiático-americanos (g. focal), independientemente de si su primera lengua era o no el inglés (Carlton y Harris, en preparación; Bleistein y Wright, 1986). En concreto, Kulick y Dorans (1983) descubrieron que tras aplicar el método de estandarización a una prueba de aptitud escolar a sujetos americanos y asiático-americanos residentes en Estados Unidos, ésta mostraba 14 items con DIF de un total de 85. Ante este resultado, los citados autores señalaron que la principal causa de la aparición del mencionado DIF podía ser el elevado porcentaje de sujetos de la muestra participante cuya primera lengua no era el inglés. Con posterioridad, este estudio fue replicado por Bleistein y Wright (1987) quienes únicamente utilizaron sujetos cuya primera lengua era el inglés, y los resultados encontrados mostraron que todavía 6 items del total presentaban DIF. Por lo que respecta a la relación existente entre la

dificultad del ítem y la detección de DIF, también en este tipo de ítems parece haber una correlación moderada entre la dificultad del ítem y la presencia de DIF. En concreto, los resultados mostraron que los negros ofrecían un mejor desempeño que los blancos en los ítems de aptitud numérica más difíciles y los blancos en los ítems numéricos más fáciles.

Por su parte, Dorans, Schmitt y Curley (1988) realizaron un estudio para averiguar si la existencia de diferencias individuales en el tiempo de realización de un test afectaba a la magnitud y detección de ítems que presentan DIF en el test. Además, se proponen evaluar la efectividad que se deriva de la eliminación de los sujetos que no llegan a finalizar el test para el cálculo del DIF en los ítems situados al final del test. Los resultados obtenidos mostraron que la eliminación de los mencionados sujetos sí que influía en el cálculo del DIF para dicho ítems, pero el efecto derivado de la velocidad del test no se eliminaba del todo. En la actualidad, éste es un tema crucial que se está estudiando para ver en qué medida afecta a la detección del DIF, especialmente para los ítems localizados al final del test, los cuales no son cumplimentados por falta de tiempo y, por tanto, son los que están más influenciados por el efecto de la velocidad del test, o mejor dicho, los límites de tiempo en la realización del test.

En resumen, los resultados aportados demuestran que la presencia de ciertas características en los ítems de un test están relacionadas con la aparición del DIF. Los estudios realizados por Schmitt y cols (Schmitt y Dorans, 1990; Dorans, Schmitt y Bleistein, 1992) afirman que éstas pueden generalizarse a través de los diferentes grupos minoritarios estudiados.

Así pues, las principales conclusiones aportadas por sus estudios se resumen a continuación:

(1) En relación al *contenido del ítem*, si éste posee algún interés para el grupo minoritario, esto produce ciertas ventajas para dicho grupo. Especialmente, este hecho ocurre con ítems relativos a construcción de frases e ítems de comprensión lectora;

(2) La presencia de ítems con *palabras homógrafas* tiende a desfavorecer a la mayoría de los grupos minoritarios evaluados;

(3) Respecto a las diferencias en la cantidad de *tiempo necesario para realizar un test* (rapidez/lentitud), la aparición de estas diferencias puede ser función de la posición del ítem en el test, por tanto este hecho puede contribuir a la aparición de DIF en ítems situados al final del test. Así pues, la aparición del DIF puede ser una consecuencia directa de la localización de los ítems en el test. Por otra parte, la existencia de diferencias en el tiempo de realización de un test tiene enormes implicaciones en la calidad del criterio, ya que éste puede estar contaminado con la presencia de dichas diferencias y, por consiguiente, contamina también la valoración de los ítems que presentan DIF en el test. Así pues, futuras investigaciones en este campo podrían aportar nuevas estrategias para la eliminación del componente velocidad en la formación del criterio y, por tanto, eliminar el citado DIF;

(4) Cuando el estudio del DIF se limita sólo a aquellos sujetos cuya primera lengua es el inglés, muy pocos ítems presentan DIF a través de las distintas versiones y grupos culturales étnicos.

A partir de los resultados obtenidos en estos estudios, los distintos investigadores de este campo reconocen que el estudio de una población minoritaria con propiedades lingüísticas, culturales y sociales concretas a partir de constructos e instrumentos de evaluación elaborados en otro contexto, difícilmente puede permitir una adecuada medición y evaluación de ésta, resultando en la mayoría de las ocasiones infravalorada. Sin embargo, en ciertos aspectos más prácticos todavía existen algunos desacuerdos en la comunidad científica acerca de qué características deben presentar los tests o instrumentos de medida utilizados en la evaluación del rendimiento académico de los grupos minoritarios. Precisamente, dentro de esa disconformidad aparecen los defensores de que el hecho de poseer una primera lengua distinta a la lengua dominante no es razón suficiente para explicar el bajo rendimiento académico en dichas minorías. Los partidarios de esta postura afirman que muchos de los sujetos con bajos niveles de rendimiento proceden de un ambiente con considerables desventajas socioeconómicas, lo cual puede ejercer una influencia significativa en las dificultades académicas mencionadas (Bempechat y Ginsburg, 1989; Dunn, 1988; Jones, 1989; Lee, 1989; Santa-Cruz y Senour, 1989; Stein, 1990; Steinberg, Blinde y Chan, 1984). No obstante, a pesar de algunas discrepancias como ésta existe un consenso generalizado entre la mayoría de los investigadores que destaca la necesidad urgente de aportar investigaciones con nuevas técnicas y métodos más sofisticados y generalizables, para tomar decisiones más racionales sobre la política educativa, los instrumentos de medida utilizados y las actuaciones realizadas en los programas de enseñanza bilingüe implantados en la gran mayoría de los países y comunidades bilingües.

Precisamente, este cambio de orientación queda claramente reflejado en algunos trabajos y estudios realizados, donde el rendimiento académico de los sujetos se analiza teniendo en cuenta las características más relevantes del contexto lingüístico y/o cultural en el que surgen y para el cual suelen ser adaptativos. Prueba de ello son los dos ejemplos que vamos a comentar a continuación.

El primero de ellos se desarrolló en Holanda. En los últimos tiempos, la sociedad holandesa ha experimentado una alta tasa de sujetos inmigrantes, lo cual ha provocado que algunos investigadores comiencen a cuestionarse la utilidad de los tests de inteligencia habitualmente empleados en algunos ámbitos educativos multiculturales. En este sentido, varios estudios orientados a confirmar esta suposición se proponen comprobar si efectivamente la aplicación de tests de aptitud a sujetos holandeses y emigrantes en el ámbito escolar puede dar lugar a la aparición de ciertas discriminaciones tras someter a comparación a ambos grupos (De Jong, 1987; Resing, Bleichrodt y Drenth, 1986; Helms-Lorens y Van de Vijver, 1995). Así pues, los resultados señalan que, tras comparar la ejecución de los sujetos emigrantes y holandeses nativos en algunos tests de aptitud, los sujetos nativos obtienen puntuaciones más altas en CI, entre 7 y 15 puntos, que los sujetos emigrantes. Además, estas diferencias aumentan cuando los items del test de inteligencia utilizado poseen un elevado contenido verbal. La solución adoptada en este caso fue la construcción y adaptación de un test de inteligencia sin contenido verbal, compuesto de figuras geométricas *'Revised Amsterdamse Kinder Intelligentie*

*Test'* (RAKIT; Resing, Bleichrodt y Drenth, 1986) para escolares inmigrantes con lengua y cultura diferente, y que previamente había sido adaptada a alumnos nativos holandeses. En este sentido, los últimos datos aportados por Willemse (1989) y, más tarde, replicados por Van de Rijt (1990), muestran con claridad que esta nueva tipología de tests elimina las diferencias entre alumnos nativos y emigrantes cuando ambos grupos poseen el mismo nivel de habilidad en cada uno de los items del test.

Un segundo caso se sitúa en Israel. En este país, se ha realizado un esfuerzo considerable en llevar a cabo una adaptación de tests en seis diferentes idiomas para mejorar la imparcialidad de la evaluación, permitiendo a las personas utilizar los tests y las escalas en su idioma preferido. Por ejemplo, los estudiantes de enseñanza secundaria pueden elegir para realizar el examen de selectividad de entrada a la universidad entre seis idiomas diferentes. De este modo, se elimina el sesgo de las puntuaciones del examen asociados a la obligación de que los estudiantes tengan que realizarlo en su segunda o tercera lengua, y por tanto se potencia la validez (Hambleton, 1996).

En definitiva, el objetivo de esta presentación ha consistido en ilustrar la importancia del problema del DIF en los tests o instrumentos de medida habitualmente utilizados por los profesionales e investigadores en el ámbito de la práctica psicológica y educativa. La idea defendida aquí establece que el propósito de los psicómetras expertos en el tema es demostrar que dichos instrumentos de medida empleados en diversos contextos aplicados están libres de la presencia de tipo de sesgo o DIF, de tal forma que su aplicación a grupos de sujetos con diferente composición

sociodemográfica pero el mismo valor en la variable objeto de estudio del test, no resulte en una discriminación o posible infravaloración de las habilidades de un grupo normativo sobre otro.





## ***(CAPITULO 4***

### **Objetivos y Metodología**



#### **4.1. PLANTEAMIENTO DE LA INVESTIGACION**

Nuestro interés se localiza en la Comunidad Valenciana: una comunidad autónoma con lengua propia y competencias exclusivas en la enseñanza. La incorporación de la lengua autóctona al sistema educativo valenciano en los últimos años ha posibilitado la presencia de programas de educación bilingüe en la mayoría de los centros escolares valencianos. De ahí pues que el bilingüismo y la enseñanza bilingüe aparezcan como aspectos relevantes en la educación de los escolares valencianos.

Sin embargo, estas características propias de una situación de contacto de lenguas no parecen haber sido consideradas cuando se establece cualquier intervención psicológica y/o educativa en el ámbito de la Comunidad Valenciana. En la mayoría de los centros escolares valencianos, la aplicación de pruebas colectivas para una evaluación de las aptitudes intelectuales es una práctica educativa bastante frecuente. Sin embargo, esta práctica habitualmente desarrollada en un contexto educativo bilingüe se realiza mediante la administración de tests o pruebas colectivas elaboradas en lengua castellana, con independencia de la lengua familiar y/o modelo educativo cursado por el alumno objeto de estudio. Únicamente en casos muy puntuales, los propios psicólogos del centro escolar proceden a realizar una traducción no validada al valenciano de la prueba que se va a utilizar, y con posterioridad la aplican a aquellos alumnos escolarizados en programas bilingües con un predominio claro del valenciano como lengua de enseñanza y que además poseen el valenciano como lengua habitual de comunicación.

Tomando en consideración esta realidad, el actual trabajo de investigación se plantea ante la necesidad de responder qué tipo de repercusiones educativas y sociales tiene la aplicación de pruebas de aptitud colectivas, elaboradas y baremadas únicamente en la lengua mayoritaria, en una situación de contacto de lenguas, donde coexisten dos lenguas oficiales y, además, los programas de enseñanza bilingüe son una realidad educativa. Atendiendo a esta problemática, el siguiente paso a realizar consistió en concretar dicha cuestión en varios objetivos más específicos y operativos. Éstos se concretan en la contestación a las cuestiones que se detallan a continuación:

*¿los/as niños/as valencianoparlantes resultan perjudicados en su evaluación ante la aplicación de un test de aptitud elaborado y redactado en la lengua mayoritaria, que los/as niños/as castellanoparlantes?*

*¿Los/as niños/as escolarizados de un programa de educación bilingüe son comparativamente discriminados ante un test de aptitud elaborado en castellano que los/as niños/as escolarizados en programas monolingües de castellano?*

*Y los/as niños/as de lengua familiar valenciana que además cursan un modelo educativo bilingüe con predominio del valenciano como lengua de enseñanza, ¿se enfrentan a mayores dificultades que los/as niños/as castellanoparlantes y escolarizados en programas monolingües de castellano cuando ambos responden a un test aptitudinal elaborado en castellano?*

*En caso afirmativo para los enunciados formulados, ¿es posible que la existencia de dichas dificultades esté determinada por los años de escolarización del alumno/a en un modelo educativo determinado?*

Así pues, atendiendo a todas estas cuestiones resulta bastante evidente que el núcleo central de este estudio reside en averiguar si la aplicación colectiva de pruebas de aptitud, elaboradas y baremadas en lengua castellana, a diferentes grupos lingüísticos resulta realmente apropiada para llevar a cabo una adecuada evaluación de las aptitudes intelectuales de los escolares valencianos. Las consecuencias derivadas de este tipo de práctica pueden ser especialmente negativas para la población escolar valenciana, sobre todo si se cumple el supuesto de que ciertos grupos de alumnos con una formación lingüística determinada resultan

perjudicados o discriminados frente a otros en la evaluación de sus propias aptitudes intelectuales.

Si dicha problemática la trasladamos al campo de estudio de la Psicometría, ésta podría ser formulada de la siguiente manera:

*¿Hasta qué punto las propiedades psicométricas de los items de un test de aptitud, elaborado en castellano y aplicado a alumnos bilingües valencianos, varían en función de la lengua que éstos hablan asiduamente en su entorno social y/o el modelo educativo de enseñanza cursado?*

En este sentido, algunos estudios realizados han demostrado que uno de los mayores problemas asociados a los instrumentos de medida o tests es el uso inadecuado de los mismos (Hambleton, 1993, 1994; Van de Vijver y Hambleton, 1996). Por este motivo, los propios profesionales y expertos en este campo insisten cada vez más en la necesidad de estudiar las propiedades psicométricas de los items de aquellos tests o cuestionarios aplicados a poblaciones con características diferentes para la que éstos fueron creados, y prestar especial atención a los items que presenten problemas de funcionamiento diferencial tras ser aplicados a poblaciones de ambientes lingüísticos y/o culturales diferentes (Botempo, 1993; Budgell y cols, 1995; Candell y Hulin, 1987; Ellis, 1989; Ellis y cols., 1992, 1993; Hulin, 1987; Hulin y Mayer, 1986). En tales circunstancias, es muy importante determinar si la aparición de diferencias significativas en los resultados obtenidos en el test son causadas por la existencia de diferencias reales en la variable medida por el test en los distintos grupos sometidos a comparación, o simplemente son producto de un problema del instrumento de medida utilizado, esto es, un test compuesto por items

con problemas de funcionamiento diferencial en los distintos grupos de sujetos analizados.

De la especificación de esta propuesta se deriva que este estudio pretende estudiar las propiedades psicométricas de los items integrantes de varias pruebas de inteligencia, elaboradas y baremadas en castellano y de uso habitual en la evaluación psicológica desarrollada en los centros escolares valencianos y, por consiguiente, en la medida que sea posible garantizar que los grupos escolares con perfil lingüístico no resulten discriminados o perjudicados frente a otros grupos, simplemente por hacer uso de una lengua minoritaria en su entorno social y/o escolar como es el valenciano. En caso de obtener resultados positivos en cuanto a la presencia de items con problemas de funcionamiento diferencial en las pruebas psicológicas analizadas aquí, este estudio sin duda contribuirá a introducir posibles mejoras en la composición de dichas pruebas y, por tanto, en su utilización para la realización de sus diagnósticos, asesoramiento y orientación vocacional en los escolares valencianos.

Como se ha expuesto anteriormente, la gran proliferación de estudios y trabajos en el ámbito del DIF se han desarrollado fundamentalmente bajo situaciones de simulación con el objetivo de averiguar qué técnicas y métodos estadísticos son los más adecuados en la detección del DIF. Este tipo de aportaciones metodológicas ha provocado que la contribución del DIF en la vertiente más aplicada del estudio sea mucho menor comparativamente hablando. Algunas de las razones que argumentan algunos profesionales e investigadores en este campo se sitúan en la compleja sofisticación que requiere el cálculo estadístico e informático de dichos métodos estadísticos, además de necesitar tamaños muestrales grandes, especialmente



para algunos de los métodos basados en la TRI. Así pues, dada la limitada tradición de esta metodología en el marco aplicado de la investigación, el presente trabajo de investigación se presenta como uno de los primeros estudios empíricos realizados sobre el posible sesgo o detección del DIF en instrumentos de medida aplicados en el Estado Español, y la primera experiencia aplicada de esta metodología en la Comunidad Valenciana.

Sin embargo, a pesar de la relevancia empírica de este estudio somos conscientes de que un trabajo como éste no puede aportar una respuesta a todas las repercusiones educativas y sociales que se derivan de la aplicación de instrumentos de medida con problemas de funcionamiento diferencial en sus ítems en contextos de naturaleza bilingüe. Las limitaciones de nuestro estudio empírico vienen dadas por el contexto en el que se desarrolla la Comunidad Valenciana, y una situación de contacto de lenguas con características lingüísticas muy peculiares. No hay que olvidar que la Comunidad Valenciana se caracteriza por ser una comunidad bilingüe donde la relación de desigualdad entre sus dos lenguas oficiales es una nota predominante tanto en el entorno social como académico, lo que repercutirá de forma notable en la realidad lingüística objeto de análisis de este estudio.

Teniendo en cuenta los comentarios realizados, a continuación vamos a introducir las diferentes variables utilizadas, y se presentarán las principales hipótesis de este trabajo de investigación.

#### 4.2. VARIABLES

De todo lo expuesto se deriva que la utilización adecuada de una prueba psicológica en un idioma u otro, en una situación de contacto entre lenguas, requiere tener en consideración el nivel de competencia lingüística adquirida por los sujetos a los cuales ésta va dirigida. La mayoría de los estudios realizados sobre evaluación de programas bilingües implantados en el ámbito educativo valenciano como en el resto del Estado Español señalan que el rendimiento académico o aprovechamiento lingüístico en dichos programas deriva principalmente del nivel de bilingüismo educativo alcanzado por dichos sujetos, o lo que es lo mismo, la competencia lingüística adquirida en ambas lenguas (en Cataluña: Generalitat de Catalunya, 1983; Bel, 1993; en Euskadi: Etxebarria, 1992; Olaziregui, 1993; y en la Comunidad Valenciana: Pascual y Doménech, 1990; Díez, 1995; Bueno, 1995).

Además, el estudio sobre la adquisición de dicha competencia lingüística, objeto de estudio de varios trabajos en programas de educación bilingüe, muestra con cierta transparencia que el aprendizaje y adquisición de las dos lenguas oficiales en una comunidad viene determinada por la presencia de factores diferentes. En términos generales, los resultados aportados en dichos estudios apuntan que mientras la adquisición de la lengua minoritaria en una comunidad depende de factores específicos como por ejemplo la lengua utilizada en el ámbito familiar o el modelo lingüístico cursado por el alumno, el dominio lingüístico adquirido en la lengua mayoritaria es independiente de dichos factores, quizás a causa a su fuerte presión social en ámbitos específicos.

Por lo tanto, las variables que vamos a analizar en esta investigación, y que posiblemente determinen la presencia de rendimientos diferenciales en los tests de inteligencia aplicados a diferentes grupos de escolares valencianos, son las siguientes:

- a) El entorno sociolingüístico de cada sujeto participante en la muestra, medido por el tipo de lengua utilizada habitualmente por el sujeto en su contexto familiar y coloquial (LF).
- b) El grado de utilización de las dos lenguas oficiales de la comunidad como vehículo de enseñanza en cada uno de los centros escolares considerados: Modelo educativo o línea de escolarización (LE).
- c) Los años de escolarización de los individuos en un determinado modelo educativo. Esta variable fue operacionalizada atendiendo al

nivel de aplicabilidad del instrumento de medida utilizado en el estudio (NA).

d) Tipo de prueba administrada. Esta variable se refiere al tipo de items utilizados (verbales o gráficos) en la evaluación de las aptitudes intelectuales de los escolares (TP).

A continuación, pasamos a analizar con más detalle cada una de ellas.

#### LENGUA UTILIZADA HABITUALMENTE POR EL SUJETO (LF).

Dada la importancia del entorno sociolingüístico inmediato del sujeto como factor determinante en la adquisición de su competencia lingüística en dos o más lenguas, ésta será la primera variable a considerar. Esta variable adquiere una relevancia considerable en una comunidad bilingüe donde su situación lingüística, lejos aún de una normalidad o normalización lingüística, está claramente dividida en dos franjas: zona castellanoparlante y zona valencianoparlante de la Comunidad Valenciana. Por tanto, los dos niveles a considerar serán los siguientes:

**1. lengua castellana, y 2. lengua valenciana.**

La operacionalización de dicha variable se realizó mediante la administración de una encuesta sociolingüística a cada sujeto participante de la muestra. El contenido de la misma hace referencia a algunos datos sociodemográficos de la muestra empleada en el estudio, como la formulación de varias cuestiones acerca de qué tipo de lengua es habitualmente utilizada por el sujeto en el ámbito familiar, coloquial, con los amigos,

etc.. Una descripción más detallada de dicha encuesta se localiza en el apartado de instrumentos de medida utilizados en el estudio.

#### MODELO EDUCATIVO O LINEA DE ESCOLARIZACION (LE).

A grandes rasgos, los modelos educativos implantados en la Comunidad Valenciana pueden clasificarse en tres categorías diferentes:

**1. Educación monolingüe en castellano.** Dentro de este modelo se encuentran los centros escolares cuya enseñanza es íntegramente en castellano, a excepción del valenciano como materia de estudio en toda la educación obligatoria, y el idioma extranjero elegido por el centro (francés o inglés). Sin embargo, es importante destacar que en algunas zonas castellanoparlantes de la Comunidad Valenciana la realidad educativa imperante es otra totalmente diferente. Todavía en la actualidad existen centros escolares valencianos que sólo contienen el castellano como lengua de instrucción y la asignatura de valenciano no se imparte en ninguno de los cursos académicos integrantes de la Educación Primaria. Así pues, nuestro estudio incluyó para esta primera categoría: 1a) los modelos educativos cuya línea de escolarización es el castellano y no se imparte la asignatura de valenciano, y 1b) los modelos de enseñanza en castellano, en los que se imparte la asignatura de valenciano en todos o algunos cursos académicos de la escolarización elemental.

**2. Educación bilingüe: Programa de mantenimiento/ Programa de incorporación progresiva del valenciano.** Este modelo educativo incluye a aquellos centros de enseñanza cuya lengua de escolarización es el

castellano durante toda la enseñanza obligatoria, a excepción de algunas variantes introducidas en 3º curso de Primaria y 1º curso de E.S.O., donde una parte de las áreas curriculares se imparten en valenciano. Además, durante toda la Educación General Básica se recibe la asignatura de valenciano, y en el ciclo superior se estudia un idioma extranjero (francés o inglés) según la opción de cada centro. En este segundo modelo educativo, nuestro estudio se realizó en centros escolares cuyas variantes fueron las siguientes: 2a) los modelos educativos con línea de escolarización en castellano para todas las asignaturas del plan de estudios, a excepción de la asignatura de *valenciano* en toda la educación obligatoria, el área de *experiencias del medio* (asignatura referida a la propia herencia étnica y cultural) en el ciclo medio y la asignatura de *Ciencias Sociales y/o Naturales* en el ciclo superior, y 2b) los modelos de enseñanza cursados en lengua castellana en todas las asignaturas, a excepción de la asignatura de *valenciano* en toda la escolarización obligatoria, el área de *experiencias del medio* en el ciclo medio, las *Ciencias Sociales y/o Naturales* en el ciclo superior, y las asignaturas de *Educación Artística y Física* en los ciclos medio y superior.

**3. Educación bilingüe: Programa de enriquecimiento/ Programa de enseñanza en valenciano.** Dentro de este tercer modelo, se incluyen aquellos centros escolares en cuyas líneas de escolarización se incluía el programa de enseñanza en valenciano. La característica principal de estos modelos es la adquisición y aprendizaje de las habilidades lectoescritoras en valenciano. Así pues, en estos modelos el valenciano es la lengua de instrucción utilizada en la mayoría de las áreas curriculares, a excepción de la asignatura referida a la lengua castellana. Es conveniente destacar que este modelo educativo es el que más variantes presenta en los

diferentes centros escolares participantes en el estudio. Dichas variantes considerando el número de asignaturas recibidas en valenciano para cada centro escolar fueron las siguientes: 3a) los modelos educativos de enseñanza en valenciano, a excepción de las asignaturas de lengua castellana, religión/actividades de estudio y matemáticas, ésta última para los ciclos medio y superior; 3b) los modelos de enseñanza bilingüe en valenciano, a excepción de las materias de estudio lengua castellana en toda la escolarización obligatoria y matemáticas en los ciclos medio y superior, y 3c) los modelos de enseñanza en valenciano, excepto la asignatura de lengua castellana para toda la educación general básica.

Diversos estudios señalan la importancia del modelo educativo cursado como una variable determinante en la competencia lingüística adquirida por los sujetos bilingües en una situación de contacto de lenguas (Alsina, 1983; Artigal, 1993; Cummins, 1993; Generalitat de Catalunya, 1983; Fernández, 1981; Mägiste, 1986; Genesse, 1986, 1989 y 1992; Sánchez, 1985; Espí, 1988; Artigal, 1993; Bel, 1993; Olazaregui, 1993; Pascual y Domenech, 1990; Diez, 1995; Bueno, 1995), en especial con respecto a la competencia lingüística adquirida en la lengua minoritaria.

AÑOS DE ESCOLARIZACION (NA).

La tercera variable a considerar se refiere a los años de escolarización recibidos por los escolares participantes en el estudio en uno de los tres modelos educativos comentados previamente. La operacionalización de esta variable se realizó atendiendo al nivel de

aplicabilidad del instrumento de medida utilizado. Así pues, en este estudio podemos distinguir entre:

1. **B.A.D.y.G. Nivel Elemental**, para niños/as de 9 a 12 años de edad. Se aplicó a escolares de 4<sup>º</sup>, 5<sup>º</sup> y 6<sup>º</sup> curso de Primaria, y
2. **B.A.D.y.G. Nivel Medio**, aplicable de 12 a 15 años. Se administró a los cursos académicos de 1<sup>º</sup> E.S.O. y 2<sup>º</sup> E.S.O.

TIPO DE PRUEBA ADMINISTRADA (TP).

Dada la importancia de las características lingüísticas de los diferentes grupos sometidos a comparación para la detección de items con problemas de funcionamiento diferencial en ambos grupos, la última variable a considerar fue el tipo de prueba administrada. Esta variable presentó dos opciones diferentes: 1. una prueba de aptitud intelectual compuesta por items de contenido verbal, denominada escala de **Habilidad Mental Verbal (HMV)**, y 2. una prueba de aptitud intelectual formada por items gráficos, denominada escala de **Habilidad Mental No Verbal (HMNV)**.

Una vez realizada la presentación de dichas variables, se procederá a presentar las diferentes hipótesis de trabajo de nuestro estudio.







#### **4.3. HIPOTESIS**

De acuerdo con lo expuesto en la fundamentación teórica sobre el tema del bilingüismo y la educación bilingüe, como en los objetivos de este trabajo, las hipótesis formuladas a continuación sobre la aparición de items con problemas de funcionamiento diferencial (DIF) se presentarán en cuatro bloques diferentes: 1) Años de escolarización (Hipótesis A), 2) Tipo de prueba administrada (Hipótesis B), 3) Según la lengua familiar del alumno (Hipótesis C), y 4) Según la lengua familiar y el modelo educativo cursado por el alumno (Hipótesis D).

No obstante, antes de pasar a formular dichas hipótesis es importante determinar qué características lingüísticas van a definir a los diferentes

grupos de sujetos sometidos a comparación. En esta investigación, el grupo de referencia (GR) estará compuesto por aquellos sujetos cuya lengua familiar es el castellano y son escolarizados en ese mismo idioma, mientras que los grupos objetivo o focales (GF) estarán formados por dos grupos diferentes: a) los sujetos cuya lengua familiar es el valenciano y el modelo lingüístico que están cursando es un programa de enseñanza bilingüe de mantenimiento, y b) los sujetos cuya lengua familiar y de escolarización es el valenciano.

**A) Años de Escolarización: BADYG Nivel Elemental y Medio**

**Hipótesis A.1:** *Las pruebas referentes al nivel de aplicación BADYG Elemental presentarán mayor porcentaje de items con funcionamiento diferencial (DIF) que las pruebas de aptitud correspondientes al BADYG Medio.*

Básicamente, los estudios realizados cuyo objetivo ha sido evaluar la incidencia de la lengua de escolarización sobre el rendimiento académico global del sujeto han señalado que la mayor presencia de diferencias significativas en dicho rendimiento académico tienen lugar cuando el sujeto se encuentra en el proceso de adquisición-aprendizaje de un idioma a lo largo de su proceso de escolarización y además su nivel de dominio lingüístico en castellano es deficiente comparado con el obtenido por el grupo monolingüe (Bueno, 1995; Díez, 1995). Por ejemplo, en nuestra comunidad los alumnos de 5º curso Primaria escolarizados en modelos de enriquecimiento usaban y dominaban mucho mejor el valenciano en tareas de asociación, comprensión, producción y discriminación de palabras que el resto de los modelos educativos; en cambio, estos mismos alumnos eran

deficitarios en tareas de corrección escrita, asociación, comprensión y velocidad lectora en castellano, comparadas con las obtenidas en los programas monolingüe y de mantenimiento. Por el contrario, dichos estudios en alumnos de los últimos cursos, esto es, al final su educación básica mostraron que éstos presentaban un dominio lingüístico equivalente en ambas lenguas, la lengua castellana y valenciana. Así pues, atendiendo al hecho de que el sujeto con lengua familiar y escolar distinta a la lengua mayoritaria no es capaz de conseguir una competencia lingüística aceptable en castellano hasta finalizar los últimos cursos de la educación obligatoria, y a las aportaciones de algunos estudios (Boada, 1986; Sanchez y Tembleque, 1986) que señalan que la eficacia de un programa bilingüe no aparece hasta los 4 ó 6 años de escolarización bilingüe, la aplicación de pruebas psicológicas en castellano en alumnos a lo largo de su proceso de escolarización (BADYG elemental) puede ser inadecuada. Dado que dichos sujetos todavía no poseen un nivel de bilingüismo equilibrado y aceptable en ambas lenguas, la aplicación de pruebas psicológicas redactadas en castellano puede repercutir en una mayor cantidad de dificultades en la realización de la misma provocando la aparición de resultados diferentes en el test, cuando en realidad éstos no poseen un rendimiento inferior en dicha variable. En cambio, los items de las mismas pruebas aplicadas a los alumnos en los últimos cursos (BADYG medio) no deberán presentar DIF, esto es, que dichos sujetos no deberán resultar ser discriminados frente al otro grupo, dado que la competencia lingüística alcanzada en castellano es equiparable a la obtenida en valenciano. Por tanto, si se cumple esta hipótesis se podrá afirmar que la existencia de funcionamiento diferencial en los items de las pruebas aplicadas en este estudio en detrimento de los grupos con características lingüísticas específicas viene modulada por los años de escolarización en un modelo de enseñanza determinado.

No obstante, es conveniente destacar que esta hipótesis se comprobará únicamente con los grupos de alumnos cuya lengua familiar es castellano o el valenciano, pero su modelo de escolarización es el programa de incorporación progresiva al castellano. Este hecho se debe a que la introducción del modelo de enriquecimiento dentro de la educación bilingüe implantada en la Comunidad Valenciana es relativamente reciente y se ha llevado a cabo de forma progresiva.

***B) Según el tipo de prueba: Habilidad Mental Verbal (HMV) y Habilidad Mental No Verbal (HMNV)***

***Hipótesis B.1:*** *Las pruebas de aptitud formadas por items con contenido verbal presentarán un mayor porcentaje de items con funcionamiento diferencial (DIF) que las pruebas de aptitud sin referentes verbales.*

La formulación de esta hipótesis se basa en la importancia que adquieren las pruebas con contenido verbal y no verbal (gráficas) para la aparición de items con funcionamiento diferencial (DIF). Si realmente la presencia de DIF en los items está directamente relacionada con las características lingüísticas de los sujetos integrantes de los grupos comparados, entonces la prueba compuesta por items de contenido verbal presentará un porcentaje de items con funcionamiento diferencial (DIF) significativamente mayor que aquellas en las que el idioma se limita sólo al protocolo de la explicación. Así pues, se espera que en las pruebas

psicológicas de contenido no verbal no aparezcan items con funcionamiento diferencial, salvo posibles errores aleatorios.

Hay que recordar que en la revisión bibliográfica aportada se presentó un estudio sobre la construcción y adaptación de un prueba de inteligencia sin referentes de tipo verbal en el contexto educativo holandés (Resing, Bleichrodt y Drenth, 1986; Willemse, 1989; Van de Rijt, 1990). Los resultados aportados tras ser aplicada a alumnos nativos holandeses y emigrantes van en el sentido de la hipótesis anterior, es decir, la eliminación de diferencias significativas para ambos grupos en todos los items del test, cuando éstos poseían el mismo nivel de habilidad.

### **C) Según la lengua familiar del alumno**

La comparación establecida en este apartado implicará a los grupos:

**Muestra LF castellano y LE castellano (Grupo de referencia)**

**Muestra LF valenciano y LE incorporación progresiva valenciano  
(Grupo focal)**

**Hipótesis C.1:** *Los alumnos valencianos cuya lengua familiar es el valenciano y línea de escolarización del modelo educativo cursado es el castellano (Modelo de incorporación progresiva al valenciano), y además cursan los niveles correspondientes a la aplicación del BADYG Elemental, resultarán perjudicados durante el proceso de evaluación de sus aptitudes intelectuales con la aplicación de tests verbales elaborados en castellano, y por tanto, obtendrán peores puntuaciones respecto a los sujetos con el*

*mismo nivel de puntuación de su grupo de referencia compuesto por alumnos castellanoparlantes escolarizados en programas monolingües de castellano. Por tanto, las propiedades de los items en la prueba verbal H.M.V. del BADYG Elemental presentarán problemas de funcionamiento diferencial (DIF) en contra del grupo focal, en este caso sujetos valencianoparlantes.*

La justificación de esta hipótesis se localiza en el choque o cambio de lengua del ámbito familiar al escolar al que están expuestos los alumnos valencianoparlantes de programas de enseñanza en castellano. Precisamente, la presencia de dicho corte en el desarrollo evolutivo de su competencia lingüística puede derivar en una serie de deficiencias lingüísticas en dichos sujetos. Y éstas pueden desencadenar ciertas dificultades a la hora de contestar una prueba verbal en castellano, al ser éstos comparados con un grupo monolingüe castellanoparlante el cual no ha experimentado ningún cambio de lengua en su incorporación al sistema educativo.

En la zona valencianoparlante de la Comunidad Valenciana, el valenciano es la lengua más necesaria para la comunicación habitual de los niños pertenecientes al grupo focal de nuestro estudio porque es la lengua generalmente más utilizada en su ambiente social. De esta manera, los alumnos valencianoparlantes escolarizados en programas de enseñanza en valenciano desarrollan en la situación escolar las mismas habilidades de competencia comunicativa que son necesarias y requeridas por su entorno ambiental, mientras que los escolares valencianoparlantes escolarizados en modelos educativos de castellano se encuentran en una situación de constante estructuración de su lenguaje comunicativo para adecuarlo al lenguaje exigido en la institución escolar, y en ocasiones para intervalos temporales muy breves porque la situación del aula, donde se exige el

segundo lenguaje, se combina con la situación de recreo, donde los estudiantes vuelven a hablar la lengua familiar. Por ello, los primeros siguen un desarrollo que da lugar a la formación de bilingües estables, en la medida en que existe una continuidad de uso entre su lengua comunicativa y de instrucción o enseñanza, y pueden acostumbrarse a la lengua castellana en el ambiente donde se presenta (medios de comunicación: TV, prensa, etc) de manera no forzada ni impositiva, mientras que los alumnos valencianoparlantes escolarizados en castellano no encuentran esta continuidad entre la lengua habitualmente usada para la comunicación y la utilizada para la instrucción, con lo cual en estos últimos estudiantes no se puede manifestar tan acusadamente como en los estudiantes de la situación anterior los beneficios que dependen de la instrucción escolar. Al respecto, la consideración de la hipótesis de los umbrales y de la interdependencia lingüística aportan una mayor claridad y coherencia a los resultados expuestos acerca de los beneficios cognitivo-académicos derivados de la enseñanza bilingüe (Cummins 1981, 1984, 1986).

***D) Según la lengua familiar y el modelo educativo cursado***

La siguiente comparación a analizar estará constituida por las muestras:

**Muestra LF castellano y LE castellano (Grupo de referencia)**

**Muestra LF valenciano y LE valenciano (Grupo focal)**

**Hipótesis D.1:** *Los escolares valencianos cuya lengua familiar y de instrucción en el modelo educativo es el valenciano, y además se encuentran*



*cursando la Educación Primaria (BADYG Elemental) resultarán discriminados ante la aplicación de pruebas de aptitud verbales desarrolladas y baremadas en castellano y, en consecuencia, presentarán un peor desempeño en dichas pruebas ante sujetos valencianos cuya lengua familiar y escolar es el castellano, y que poseen el mismo nivel o puntuación en la aptitud medida. Así pues, las pruebas verbales aplicadas del BADYG Elemental presentarán items con funcionamiento diferencial (DIF) en contra de los sujetos valencianoparlantes escolarizados en valenciano.*

Al igual que en el caso anterior, el grupo objetivo o focal estará compuesto por sujetos valencianoparlantes pero que además están escolarizados en programas de enseñanza en valenciano. Las características propias de dichos modelos educativos bilingües son la adquisición y el aprendizaje de las habilidades lectoescritoras en valenciano, por tanto los alumnos que se encuentren cursando cursos académicos intermedios (4º, 5º y 5º de Primaria) estarán menos expuestos a la presencia y aprendizaje de la lengua mayoritaria, el castellano, luego sus habilidades lectoescritoras en castellano no estarán lo suficientemente desarrolladas hasta finalizar la escolarización elemental, como así lo confirman algunos estudios (Bueno, 1995). Por tanto, la aplicación de pruebas verbales en lengua castellana para evaluar las habilidades intelectuales de dichos sujetos será una práctica educativa inadecuada. Dado el menor desarrollo lingüístico en castellano de dichos sujetos a lo largo de su proceso de escolarización en comparación con las habilidades adquiridas por el grupo monolingüe castellano, los primeros, a niveles iguales en la medida de habilidad, se enfrentarán con una mayor cantidad de dificultades o handicaps lingüísticos a la hora de responder a los items de una prueba verbal elaborada en castellano.

En líneas generales y a grandes rasgos, las hipótesis que hemos planteado son paralelas a las hipótesis más generales comentadas anteriormente en el apartado referido a los modelos teóricos del bilingüismo educativo y que han servido de paradigmas principales para la mayoría de las investigaciones realizadas en el mundo del bilingüismo y la educación bilingüe. De los 6 principios presentados en el marco teórico, tres de ellos justifican fundamentalmente las hipótesis del presente trabajo. Éstas son los principios de los "umbrales" y la "interdependencia lingüística" de Cummins, y el principio de bilingüismo aditivo y bilingüismo substractivo de Lambert. Según los principios de los umbrales y de la interdependencia lingüística la presencia curricular del valenciano no perjudica el dominio del castellano sino que lo mejora para ambas lenguas, siempre y cuando los sujetos evaluados hayan cursado un programa bilingüe en valenciano durante toda su escolarización acaadémica. La otra hipótesis presente se basa en el principio de "bilingüismo aditivo y bilingüismo substractivo" de Lambert que en la realidad bilingüe valenciana supone que el bilingüismo producido por el sistema escolar debido a la introducción vehicular del valenciano en la institución escolar al finalizar la escolarización debe resultar más beneficiosa que la práctica monolingüe en castellano. Así pues, en caso de confirmarse la primera hipótesis planteada la aplicación de tests psicológicos y educativos elaborados en castellano en el ámbito educativo valenciano únicamente resultará ser adecuada o correcta a partir de ciertos cursos escolares, esto es, a partir de 1<sup>º</sup> y 2<sup>º</sup> ESO, cuando los individuos han finalizado o están a punto de finalizar su proceso de adquisición-aprendizaje de ambas lenguas. No obstante como se ha expuesto anteriormente, esta última propuesta sólo se someterá a prueba cuando los sujetos evaluados poseen

como lengua familiar el valenciano pero su lengua de escolarización es el castellano.

#### 4.4. INSTRUMENTOS DE MEDIDA.

##### 4.4.1. Prueba de aptitud: 'B.A.D.y.G'.

La prueba de aptitud seleccionada para realizar esta investigación fue la '*Batería de Aptitudes Diferenciales y Generales (B.A.D.y.G)*' (Yuste, 1988). Dicha batería elaborada y baremada en lengua castellana es una de las pruebas más utilizadas en la Comunidad Valenciana para la evaluación de las aptitudes intelectuales de los escolares valencianos. El motivo de la elección de dicha prueba residió, después de una consulta via telefónica con diferentes profesionales de distintos centros escolares y gabinetes psicopedagógicos, en el nivel de utilización de la misma siendo la prueba más utilizada en el ámbito de la medición de la aptitudes intelectuales.

La totalidad de las pruebas, englobadas bajo la denominación B.A.D.Y.G., presenta seis niveles diferentes de aplicabilidad en el ámbito escolar: los niveles A, B y C aplicables a niños/as desde los 4 a los 9 años, el nivel Elemental (de 9 a 12 años), el nivel Medio (de 12 a 15 años) y el nivel Superior (de 15 a 18 años). Atendiendo a los diferentes estadios del desarrollo infantil, los citados niveles se pueden clasificar en dos grandes grupos: por un lado, los niveles gráficos A, B y C que se caracterizan por poseer una presentación más atractiva y menos verbalizada, junto con items más concretos e intuitivos como corresponde a las primeras etapas del pensamiento lógico-concreto del individuo; y por otro, los niveles Elemental (situado entre la lógica concreta y la formal), Medio y Superior. El foco de interés de este trabajo se centra en dos niveles del segundo grupo: B.A.D.y.G. nivel Elemental y Medio, donde el componente verbal juega un papel predominante en la evaluación de las aptitudes intelectuales de los escolares.

Respecto a su contenido, esta batería de tests está compuesta por ocho pruebas diferentes, las cuales en conjunto pretenden ofrecer una medida de la inteligencia general a través de dos escalas generales: 1) la Inteligencia General Verbal (I.G.V.), y 2) la Inteligencia General No-Verbal (I.G.NV.). La *Inteligencia General Verbal (I.G.V)* es un constructo que mide la capacidad intelectual verbal, esto es, la capacidad de relacionar y seleccionar conceptos, razonar verbal y numéricamente, y comprender correctamente tanto vocabulario como órdenes verbales escritas. Precisamente, las pruebas que intentan medir este factor son las que más relación presentan con los procesos de aprendizaje, el nivel socio-cultural del sujeto, la transmisión de la cultura en el ámbito escolar y, por tanto, las que más correlacionan positivamente con el rendimiento escolar. Por su

parte, la *Inteligencia General No-Verbal (I.G.NV)* mide la capacidad de razonamiento lógico y espacial, y de resolver problemas geométricos. Precisamente, es la inteligencia más relacionada con el espacio, ya que los items que la componen se resuelven en un espacio de dos o tres dimensiones.

De las ocho pruebas que componen la citada batería, nuestro interés se sitúa únicamente en dos de ellas: la escala de *Habilidad Mental Verbal (H.M.V.)*, compuesta por 40 items de fuerte contenido verbal y la escala de *Habilidad Mental No-Verbal (H.M.NV.)*, formada por 40 items gráfico-geométricos. Como se comentó previamente, la distinción entre items verbales y no verbales es una cuestión relevante para el objetivo presentado en este estudio, ya que dicha diferenciación será decisiva para la aparición de items con problemas de DIF en función de las características lingüísticas de los diferentes grupos analizados.

Ambas escalas son pruebas de lápiz y papel, de aplicación colectiva y en las que no cuenta mucho el factor TIEMPO, ya que se pretende lograr un equilibrio entre la longitud de las pruebas y el tiempo de aplicación. El tiempo de aplicación de dichas pruebas varía en función del curso académico evaluado, debido a que se pretende conseguir una dificultad similar y una mejor discriminación de las puntuaciones a lo largo de todas las edades. En la siguiente tabla se exponen las condiciones de aplicación de los niveles B.A.D.Y.G. Elemental y Medio para cada curso académico o edad de aplicación correspondiente.

Por otra parte, también es importante destacar que el ACIERTO POR AZAR en los items que componen ambas subescalas es prácticamente escaso, debido a que éstos poseen 6 alternativas diferentes de respuesta.

**Tabla 4.4.1.1.**  
Tiempos de aplicación de las escalas H.M.V. y H.M.NV.  
pertenecientes al BADIYG nivel Elemental y Medio.

	H.M.V.	H.M.NV.
4º Primaria	20 minutos	20 minutos
5º Primaria	16 minutos	16 minutos
6º Primaria	12 minutos	12 minutos
1º E.S.O	20 minutos	22 minutos
2º E.S.O	18 minutos	20 minutos

A continuación, vamos a presentar las características y propiedades psicométricas de cada una de ellas.

#### **4.4.1.a. Habilidad Mental Verbal (H.M.V.)**

La escala de Habilidad Mental Verbal (H.M.V.) es una prueba aptitudinal integrada en la Batería de Aptitudes Diferenciales y Generales (B.A.D.Y.G.) niveles elemental y medio, y cuyo objetivo es la medición de la inteligencia verbal. Ésta consta de 40 ítems verbales, de dificultad creciente y una escala de respuesta con seis alternativas diferentes donde únicamente una opción es la correcta. Los distintos ítems que forman parte de dicha prueba pueden clasificarse en función del tipo de tarea a realizar. Éstas se detallan a continuación para cada uno de los niveles de aplicación. En ellas, la puntuación total se suele interpretar como la capacidad que posee el individuo en tareas de clasificación, deducción, análisis e integración de conceptos tanto verbales como numéricos, pero siempre formulados de forma verbal.

**Tabla 4.1.1.a1.**  
Distribución de los items de la escala H.M.V en función de la  
tarea a realizar en el *BADYG Elemental*

HABILIDAD MENTAL VERBAL	
*Constancia de una característica	9 items
*Secuencias lógicas con números	7 items
*Ordenar palabras sueltas formando una frase correcta	6 items
*Hallar el género o característica clasificatoria de una serie de palabras	8 items
*Problemas numéricos de comprensión lógico-numérica	7 items
*Problemas de resolución espacio-temporal	3 items
<b>TOTAL</b>	<b>40 items</b>

**Tabla 4.1.1.a2.**  
Distribución de los items de la escala H.M.V en función de la  
tarea a realizar en el *BADYG Medio*

HABILIDAD MENTAL VERBAL	
*Constancia de una característica	4 items
*Secuencias lógicas con números	10 items
*Ordenar palabras sueltas formando una frase correcta	5 items
*Hallar el género o característica clasificatoria de una serie de palabras	7 items
*Problemas numéricos de comprensión lógico-numérica	11 items
*Problemas de resolución espacio-temporal	3 items
<b>TOTAL</b>	<b>40 items</b>

Respecto a la baremación original de la prueba, la correspondiente a la escala elemental se realizó con una muestra total de 439 sujetos procedentes de 5 colegios diferentes de Madrid y Pontevedra, abarcando desde niveles socioeconómicos bajos hasta altos. Ésta escala presentó una media aritmética y una desviación estándar de 21.26 y 6.90, respectivamente. El coeficiente de fiabilidad con el método de las dos mitades ofreció un valor de 0.88, y el de test-retest de 0.92. Por lo que respecta a la escala nivel medio, cabe destacar que la baremación de ésta al igual que en el caso anterior se estableció a partir de una muestra de



413 sujetos de 4 colegios distintos procedentes de Madrid y Pontevedra, y de nivel socioeconómico entre medio y medio-alto. Los valores obtenidos en la media aritmética y la desviación típica fueron 22.89 y 6.02, respectivamente. Y la fiabilidad obtenida mediante los métodos dos mitades y test-retest fue de 0.77 y 0.79.

#### **4.4.1.b. Habilidad Mental No Verbal (H.M.NV.)**

La escala denominada *Habilidad Mental no Verbal (H.M.NV.)* consta de 40 items de tipo gráfico-geométrico, con 6 opciones diferentes de respuesta de las cuales una es la correcta, y ordenados progresivamente por orden de creciente dificultad. Los items componentes de dicha escala pueden clasificarse en:

- a) ITEMS ESTATICOS, los cuales representan a figuras incompletas que deben ser acabadas formando parte de un todo ordenado lógica o simétricamente. Su correcta resolución requiere cierta capacidad de observación, buen sentido espacial y una adecuada captación de la forma total o completa. Todos ellos forman un total de 13 ítems estilo MOSAICO.
  
- b) ITEMS DINÁMICOS, formados por grupos de figuras lógicamente interconectadas, que requieren para su solución el ejercicio directo de las funciones lógicas y cierto sentido espacial, o al menos una actitud positiva hacia el trabajo con figuras geométricas. El movimiento cambiante de las figuras puede hacerse de forma horizontal o vertical, o en ambas direcciones alternadas. Y los cambios lógico-

dinámicos presentes en ellas pueden combinar las relaciones básicas de tamaño, cantidad (número), color, posición relativa de las figuras, complementación de figuras y síntesis deductiva.

Respecto a su baremación, ésta se realizó con la misma muestra que la escala de H.M.V. La media aritmética y la desviación típica obtenidas en dicha escala fueron de 20.21 y 7.10 para el BADYG elemental, y de 18.40 y 6.73 para el BADYG Medio. En relación a los análisis de fiabilidad con los métodos de dos mitades y test-retest, éstos presentaron valores de 0.87 y 0.75 para el nivel BADYG elemental, y de 0.87 y 0.79 para el nivel BADYG medio, respectivamente.

Por último lugar, en lo relativo a la administración de ambas pruebas señalar que las instrucciones dadas para su realización y el orden de ejecución de las mismas se realizó siguiendo las normas dictadas por el autor de éstas. Antes de comenzar la aplicación de las pruebas, los psicólogos aplicadores de las mismas procedieron a explicar la mecánica de éstas realizando cuatro ejemplos diferentes con el fin de asegurar que todos los sujetos participantes en el estudio entendían perfectamente el funcionamiento de ambas pruebas. Tras realizar algunas aclaraciones y responder a varias dudas acerca de su correcta ejecución, se procedió primero a la administración de la escala de Habilidad Mental Verbal , y con posterioridad la escala de Habilidad Mental No verbal. Para su consulta, remitirse al apartado de anexos situado al final de la tesis, donde se adjunta una copia de las mismas.

#### **4.4.2. ENCUESTA SOCIOLINGÜÍSTICA.**

Además, a cada sujeto participante del estudio se le administró una encuesta sociolingüística. La composición de dicha encuesta puede clasificarse en dos partes bien diferenciadas. La primera parte está compuesta por 4 cuestiones referentes a ciertos datos sociodemográficos de cada sujeto, por ejemplo sexo, edad, lugar de nacimiento y nivel socioeconómico de los padres; y la segunda parte que consta también de 4 preguntas que permiten conocer el uso real que cada sujeto realiza de las dos lenguas oficiales del territorio en su contexto social más inmediato, por ejemplo a nivel de relación familiar (en casa), escolar (en el patio del colegio), personal (con sus amigos fuera del colegio), etc.. Las distintas alternativas de respuesta utilizadas, seis opciones para la primera pregunta y cinco para el resto, hacen referencia a qué lengua o lenguas son utilizadas habitualmente en las situaciones mencionadas. Por regla general, la administración de dicha encuesta se realizó al principio de cada sesión.

La utilidad de dicha encuesta se desarrolla con más detalle en el siguiente apartado, en el se procede a la descripción del proceso de obtención de la muestra. La encuesta utilizada se presenta en un anexo al final de la tesis.

#### **4.5. MUESTRA**

##### **4.5.1. Proceso de Recogida de Datos**

La selección de las escuelas participantes en el presente estudio se realizó partiendo de una serie de listados facilitados por la Conselleria de Cultura, Educació i Ciència de la Comunitat Valenciana, y atendiendo a las diferencias cualitativas y cuantitativas existentes en el uso y dominio lingüístico de la lengua propia en la Comunidad Valenciana. Dicho de otro modo, que un criterio designado para la selección de dichos centros fue el territorio histórico en el cual se hallaba enclavado cada centro escolar, esto es, su localización bien en la zona castellanoparlante o bien en la zona valencianoparlante de la propia comunidad. Al menos de esta manera, la selección de los tres modelos lingüísticos implantados en la Comunidad Valenciana quedaba totalmente garantizada.

Una vez realizada la selección aleatoria de los diferentes centros escolares implicados, se procedió a informar a los mismos sobre el objetivo de nuestro trabajo de investigación y, en consecuencia, pedir su apoyo y colaboración. Sin embargo, dada la enorme complejidad hallada en cuanto a la organización y distribución de los distintos centros escolares seleccionados, la estrategia adoptada para su localización y posterior contacto fue doble: a) mediante una entrevista personal y/o telefónica con el psicólogo y director de cada centro escolar, o bien b) en algunas zonas donde un mismo psicólogo ejercía en 3 ó 4 centros escolares distintos, el punto de conexión para informar y solicitar colaboración fueron los Gabinetes Psicopedagógicos Municipales o los Servicios Psicopedagógicos Escolares existentes, dependiendo de los servicios existentes en la zona.

Durante el proceso de recogida de datos, es importante destacar que descubrimos dos acontecimientos inesperados y comentados previamente: a) el primero de ellos fue que en la gran mayoría de los centros educativos en la zona castellanoparlante de la Comunidad Valenciana (en los que se sigue el modelo de enseñanza monolingüe en castellano) no se impartía la asignatura de valenciano como materia de estudio, y b) el segundo fue que el modelo bilingüe de enriquecimiento de valenciano no era un modelo extendido en la mayoría de los centros escolares, aunque sólo fuera por obligación legal. La realidad descubierta fue que la aplicación de la normativa lingüística sobre la extensión curricular del valenciano al sistema educativo se había producido a un ritmo muy diferente, según las zonas lingüísticas consideradas, las características de los alumnos y la disponibilidad y formación del profesorado, entre otros motivos. De hecho, esta circunstancia limitó muchísimo los posibles centros escolares participantes en este estudio, y por tanto la selección aleatoria del modelo bilingüe de

enriquecimiento de valenciano se realizó atendiendo a nuestras posibilidades y la disponibilidad de los colegios existentes.

A continuación, se ofrece una distribución de los distintos centros escolares participantes en el presente estudio (ver tabla 4.5.1.1.), donde se indica el número asignado a cada escuela participante, el nombre del centro y población donde se halla ubicado, la zona lingüística considerada y el/los modelos lingüísticos impartidos en dichos centros escolares.

**Tabla 4.5.1.1.** Distribución de los diferentes centros escolares participantes en esta investigación.

Nº	CENTRO ESCOLAR (POBLACION)	ZONA LINGÜÍSTICA	MODELO LINGÜÍSTICO
01	C.P. "San Jaime Apostol" (ALFARP)	> 80 %	Mod. 2: Mantenimiento
02	C.P. "Tomás Albert" (ALBUIXECH)	40-80 %	Mod. 2: Mantenimiento
03	C.P. "Miguel de Cervantes" (CATADAU)	> 80 %	Mod. 2: Mantenimiento
04	C.P. "San Francisco de Borja" (LIOMBAI)	> 80 %	Mod. 2: Mantenimiento
05	C.P. "Sant Francés" (GUADASUAR)	> 80 %	Mod. 2 y 3: Manten. y Enriquecimiento
06	C.P. "Almazaf" (ALMUSSAFES)	> 80 %	Mod. 2 y 3: Manten. y Enriquecimiento
07	C.P. "" (ALBAL)	< 40 %	Mod. 2: Mantenimiento
08	C.P. "Misericordia" (VINAROS)	> 80 %	Mod. 2 y 3: Manten. y Enriquecimiento
09	C.P. "Sant Sebastià" (VINAROS)	> 80 %	Mod. 2: Mantenimiento
10	C.P. "Fray Pere" (ALCALA de XIVERT)	> 80 %	Mod. 3: Enriquecimiento
11	C.P. "Marqués de Benicarló" (BENICARLO)	> 80 %	Mod. 2 y 3: Manten. y Enriquecimiento
12	C.P. "E.M. Ródenas" (BENICARLO)	> 80 %	Mod. 3: Enriquecimiento

(cont.)

Nº	CENTRO ESCOLAR	ZONA	MODELO
	(POBLACION)	LINGÜÍSTICA	LINGÜÍSTICO
13	C.P. "Verge Font de la Salut" (TRAIGUERA)	> 80 %	Mod. 3: Enriquecimiento
14	C.P. "Jaume Sanz" (PENYISCOLA)	> 80 %	Mod. 2: Mantenimiento
15	C.P. "Blasco Ibañez" (UTIEL)	< 40 %	Mod. 1: Monolingüe Cast.
16	C.P. "M. Garcia Izquierdo" (SAN ANTONIO)	< 40 %	Mod. 1: Monolingüe Cast.
17	C.P. "Higuerillas" (REQUENA)	< 40 %	Mod. 1: Monolingüe Cast.
18	C.P. "Sagrada Familia" (MANISES)	< 40 %	Mod. 2: Mantenimiento
19	C.P. "Vicente Nicolau" (MANISES)	< 40 %	Mod. 2: Mantenimiento
20	C.P. "Lucio Gil Fagoaga" (REQUENA)	< 40 %	Mod. 1: Monolingüe Cast.
21	C.P. "Canónigo Muñoz" (UTIEL)	< 40 %	Mod. 1: Monolingüe Cast.
22	C.P. "N.Sra de la Asunción" (VINAROS)	> 80 %	Mod. 2 y 3: Manten. y Enriquecimiento
15	C.P. "Blasco Ibañez" (UTIEL)	< 40 %	Mod. 1: Monolingüe Cast.
16	C.P. "M. Garcia Izquierdo" (SAN ANTONIO)	< 40 %	Mod. 1: Monolingüe Cast.
17	C.P. "Higuerillas" (REQUENA)	< 40 %	Mod. 1: Monolingüe Cast.
18	C.P. "Sagrada Familia" (MANISES)	< 40 %	Mod. 2: Mantenimiento
19	C.P. "Vicente Nicolau" (MANISES)	< 40 %	Mod. 2: Mantenimiento
20	C.P. "Lucio Gil Fagoaga" (REQUENA)	< 40 %	Mod. 1: Monolingüe Cast.
21	C.P. "Canónigo Muñoz" (UTIEL)	< 40 %	Mod. 1: Monolingüe Cast.
22	C.P. "N.Sra de la Asunción" (VINAROS)	> 80 %	Mod. 2 y 3: Manten. y Enriquecimiento
23	C.P. "Benjamin Benlloch Verdejo" (MANISES)	< 40 %	Mod. 2: Mantenimiento
24	C.P. "Ausias March" (MANISES)	< 40 %	Mod. 2: Mantenimiento
25	C.P. "Felix Rodriguez de la Fuente" (MANISES)	< 40 %	Mod. 2: Mantenimiento
26	C.P. "Joan Fuster" (MANISES)	< 40 %	Mod. 2: Mantenimiento
27	C.P. "El Carmen" (MANISES)	< 40 %	Mod. 2: Mantenimiento

\* Zona lingüística: se refiere al porcentaje de valencianoparlantes en la zona considerada. Tasa alta de valencianización: >80 %; Tasa de valencianización media: 40-80%, tasa baja de valencianización:<80 %.

Todos los colegios seleccionados pertenecen a centros escolares públicos distribuidos por toda la Comunidad Valenciana. La selección de una muestra compuesta por centros escolares públicos en lugar de centros privados fue motivada por los siguientes argumentos:

a) El carácter obligatorio y gratuito de los distintos centros escolares seleccionados posiblemente neutralice los posibles efectos derivados de las diferencias socioeconómicas presentes en los diferentes alumnos participantes en la muestra, y

b) por otra parte, si tomamos en consideración que todos los centros públicos siguen las mismas normas de organización y funcionamiento, y además están controlados todos por la misma inspección educativa, entonces es posible que la existencia de diferencias en los contenidos impartidos sea mínima.

#### **4.5.2. Descripción de la muestra**

La totalidad de la muestra obtenida en la presente investigación, tras eliminar algunas pruebas por haber sido contestadas en gran medida al azar, fue de 2128 alumnos con edades comprendidas entre los 9 y los 15 años, y escolarizados en los siguientes niveles de la Educación Primaria: 4ª Primaria, 5ª Primaria, 6ª Primaria, 1ª E.S.O. y 2ª E.S.O. durante el curso académico 1997-1998.



Recordemos que nuestro objetivo se sitúa en una situación de contacto de lenguas, donde los diferentes grupos sometidos a comparación están formados por personas que conviven en una comunidad bilingüe y donde coexisten dos lenguas oficiales distintas, aunque a niveles muy diferentes dependiendo de la zona lingüística considerada. En este sentido, los estudios empíricos realizados en la Comunidad Valenciana (Conselleria de Cultura, Educació i Ciència, CCEC, 1990, 1995; Bueno, 1995; Díez, 1995) señalan que el grado de bilingüismo adquirido por los valencianos viene determinado principalmente por el modelo lingüístico cursado y el contexto o ámbito sociolingüístico del individuo.

En este sentido, es conveniente resaltar que en los últimos años los sectores más industrializados de la zona valencianoparlante de la Comunidad Valenciana han experimentado un elevado incremento de población inmigrante procedente del resto del Estado Español, así como de fuera de éste, como es la inmigración de la población magrebí a nuestra comunidad, caso en el que además del inconveniente de aprender un idioma nuevo y la existencia de deficiencias lingüísticas se añaden las características propias de todo proceso de aculturación. La forma de controlar esta variable para la selección de la muestra se realizó de la siguiente forma: (a) averiguando la condición lingüística familiar de cada alumno de la muestra mediante la aplicación de una encuesta sociolingüística donde se preguntaba de forma individualizada cuál es lengua que habitualmente se utiliza en el entorno familiar, o bien (b) solicitando el lugar de nacimiento del alumno, que como bien indican los datos aportados por algunos estudios valencianos (CCEiC, 1990, 1995) ésta es una variable que permite definir con mayor claridad la conducta lingüística adoptada por los valencianos. Una vez obtenida dicha información, se procedió a eliminar del total de la muestra a aquellos

alumnos cuya lengua familiar era diferente a las dos lenguas oficiales de dicha comunidad, así como a los nacidos fuera del Estado Español que en su gran mayoría también poseían como lengua familiar un idioma distinto al castellano y el valenciano.

El siguiente paso fue determinar qué grupos de sujetos iban a ser sometidos a comparación en función de las tres primeras variables consideradas en el presente estudio, ya que la cuarta variable referida al tipo de prueba administrada es una variable común a todos los grupos comparados. Recordemos que dichas variables son: a) la *lengua familiar* (LF) del alumno, que posee dos niveles: lengua castellana (valor 1) y lengua valenciana (valor 2); b) la *lengua de escolarización* (LE) que presenta tres niveles: programa monolingüe en castellano (valor 1), programa bilingüe (modelo de mantenimiento o incorporación progresiva al valenciano) (valor 2) y programa bilingüe (modelo de enseñanza en valenciano) (valor 3), y c) los años de escolarización (NA) de los sujetos, variable que fue operacionalizada en el nivel de aplicabilidad de las dos escalas del BADYG: nivel elemental (valor 1) y nivel medio (valor 2). De todas ellas, la primera variable relativa a la lengua familiar de los alumnos se operativizó de dos formas distintas, para ver en qué sentido ello afectaba a los diferentes tamaños muestrales (Ns) obtenidos en cada caso.

Por lo que respecta a la primera opción (OPCION 1), la variable lengua familiar tomó el valor 1 (lengua castellana) cuando el sujeto contestaba la opción "siempre en castellano" en la pregunta 6 de la encuesta sociolingüística<sup>1</sup> administrada a cada sujeto y tomó el valor 2 (lengua valenciana) cuando el sujeto contestaba las opciones "siempre en valenciano" ó "más en valenciano que en castellano" de las preguntas 6 y 8 de la misma

encuesta sociolingüística<sup>1</sup>. Los resultados obtenidos para esta primera elección se ofrecen en la tabla 1.

**Tabla 4.5.2.1.**  
Distribución de los tamaños muestrales en los grupos  
definidos para la OPCION 1.

	BADYG 1		BADYG 2	
	LF C	LF V	LF C	LF V
LE C	324	0	311	0
LE CV	162	173	88	209
LE V	16	199	9	22

En cuanto a la segunda opción (OPCION 2), la variable lengua familiar adquirió el valor 1 (lengua castellana) cuando el sujeto únicamente selecciona la opción "siempre en castellano" en la pregunta 6 de la citada encuesta, y tomó el valor 2 (lengua valenciana) cuando el sujeto contestó las opciones "siempre en valenciano" ó "más en valenciano que en castellano" de la misma pregunta. La resultante de esta segunda opción se ofrece en la tabla 4.5.2.2.

**Tabla 4.5.2.2.**  
Distribución de los tamaños muestrales en los grupos  
definidos para la OPCION 2.

	BADYG 1		BADYG 2	
	LF C	LF V	LF C	LF V
LE C	333	2	316	2
LE CV	195	193	149	285
LE V	43	249	12	37

Respecto a las razones que motivaron la selección de las dos opciones presentadas en la variable lengua familiar, conviene destacar que en este estudio la población objeto de análisis está compuesta por alumnos de 9-13

años en edad escolar, lo que implica que las relaciones informales establecidas por éstos se desarrollan mayoritariamente en el ámbito familiar y/o escolar. Por este motivo, decidimos que los dos criterios presentados (Opción 1 y 2) fueran igualmente válidos para definir la lengua que habitualmente utilizada el sujeto evaluado. Teniendo en cuenta los diferentes Ns conseguidos, se puede apreciar que la segunda alternativa parece ser la más recomendable ya que el tamaño muestral conseguido es superior a 200 sujetos en una buena parte de los diferentes grupos. Además, no hay que olvidar que uno de los requisitos en cuanto a tamaño muestral se refiere para la detección adecuada del DIF con la técnica Mantel-Haenzel es un tamaño mínimo de 200 personas por grupo. Por lo tanto, en este estudio la opción 2 será la alternativa seleccionada.

En definitiva, el grupo de referencia en esta investigación está compuesto por aquellos sujetos cuya lengua familiar es el castellano (LF C) y son educados en ese mismo idioma (LE C), y los grupos objetivo o focales serían dos: a) los sujetos cuya lengua familiar es el valenciano (LF CV) y que cursan sus estudios siguiendo un programa de enseñanza bilingüe de mantenimiento (LE CV), y b) los sujetos cuya lengua familiar y de escolarización es el valenciano (LF V y LE 3 V). Si nos fijamos en los Ns obtenidos en esta segunda opción, podemos ver que éstos permiten realizar todas las comparaciones mencionadas con el BADYG elemental (BADYG 1), y sólo una comparación para el BADYG medio (BADYG 2). El pequeño N obtenido para el grupo focal b en el caso del BADYG medio tiene su justificación en que la incorporación de la educación bilingüe en el sistema educativo valenciano, en concreto el modelo de enriquecimiento de la lengua valenciana es relativamente reciente y, además, dicha introducción se ha realizado de forma progresiva.

A continuación, vamos a describir la composición muestral de cada uno de los diferentes grupos objeto de estudio.

- (1) MODELO EDUCATIVO CON LENGUA DE ESCOLARIZACION EN CASTELLANO (LE C)  
Y LENGUA FAMILIAR CASTELLANO (LF C).

Dentro de este grupo se incluyen únicamente aquellos centros escolares localizados en la zona castellanoparlante de la Comunidad Valenciana. Así pues, la muestra que integra este primer grupo estará formada por sujetos monolingües castellanos, cuya lengua familiar y de escolarización es el castellano. En él, podemos distinguir entre:

*1a) Muestra con lengua de escolarización en castellano (LE C) y lengua familiar castellano (LF C)- BADYG 1: Nivel Elemental*

La muestra utilizada en este grupo está compuesta por 333 alumnos escolarizados en 5ª Primaria y 6ª Primaria. Sus edades están comprendidas entre los 10 y 12 años, con una edad media de 11.06 años y una desviación típica de 0.78. Respecto a la distribución por sexos, un 49.2 % de la muestra son varones y un 50.8 % mujeres. En cuanto al nivel socioeconómico de dichos alumnos, éste fue evaluado a través de la profesión laboral de los padres. Esta variable se presentó con formato abierto en la encuesta sociolingüística y posteriormente su operacionalización se realizó utilizando las siguientes categorías de respuesta: 1) Miembro de Dirección (directivos, altos directivos, etc.); 2) Profesional con Titulación Superior (licenciado, médico, abogado..); 3) Profesional con Titulación Media (Profesor EGB, graduado social, ATS-DUE..); 4) Profesional cualificado con

formación profesional 1<sup>º</sup>-2<sup>º</sup> grado (oficiales, secretaria, auxiliares, mecánico, fontanero,..); 5) Profesional no cualificado (peones, obreros..); 6) Autónomo; 7) Jubilado/pensionista; 8) Ama de casa; 9) Parado. En caso de poseer un sólo miembro familiar, el valor obtenido fue el asignado a dicho sujeto. En caso de obtener dos valores diferentes, uno para cada miembro familiar, el valor asignado fue un promedio de ambos valores. Una tabulación de la variable nivel profesional del padre permite conocer que el 3.9 % del total son profesionales con titulación superior, el 5.4 % profesionales con titulación media, el 18 % profesionales cualificados con formación profesional, el 48 % profesionales no cualificados, el 18.6 % autónomos, en su gran mayoría de carácter agrícola, el 0.3 % jubilados y el 0.9 % parados, siendo el 4.8 % datos faltantes en dicha variable. En cuanto a la ocupación profesional desempeñada por la madre, señalar que el 0.6 % son profesionales con título superior, el 6 % profesionales con titulación media, el 8.1 % con cualificación de formación profesional, el 17.7 % profesionales no cualificados, el 5.4 % trabajadoras autónomas, en su mayoría de carácter agrícola, el 60.7 % amas de casa y el 0.3 % en paro, constituyendo el 1.2 % los datos faltantes.

*1b) Muestra con lengua de escolarización en castellano (LE C) y lengua familiar castellano (LF C)- BADYG 2: Nivel Medio*

La muestra siguiente está compuesta por 316 alumnos escolarizados en 1<sup>º</sup> E.S.O. y 2<sup>º</sup> E.S.O, de los cuales 47.8 % son varones y el 52.22% mujeres. Su edad oscila entre los 12 y los 15 años de edad, siendo su edad media de 12.92 años y la desviación típica de 0.832. Respecto al nivel socioeconómico de los padres, éste presentó para la variable referida al nivel profesional del padre la siguiente distribución: el 5.1 % profesionales con titulación

susuperior, el 5.1 % profesionales con titulación media, el 13 % profesionales cualificados con formación profesional, el 43.7 % profesionales no cualificados, el 27.8 % autónomos, el 0.6 % jubilados y el 0.6 restante en paro, siendo el 4.1 % datos faltantes en dicha variable. Por su parte, la distribución obtenida para la madre estuvo formada por: el 2.2 % profesionales con titulación superior, el 3.5 % profesionales con titulación media, el 3.8 % cualificados con formación profesional, el 15.2 % profesionales no cualificados, el 5.1% autónomos, el 0.3 % en paro y el 67.7 % amas de casa, siendo el 2.2 % datos faltantes en dicha variable.

(2) MODELO EDUCATIVO DE INCORPORACION PROGRESIVA AL VALENCIANO (LE CV)  
Y LENGUA FAMILIAR VALENCIANO (LF V) .

Al igual que en el modelo anterior, podemos distinguir dos muestras diferentes en función de los niveles de aplicación de las pruebas utilizadas.

*2a) Muestra escolarizada en un programa de incorporación progresiva al valenciano (LE V) y lengua familiar valenciano (LF V) - BADYG 1: Nivel Elemental.*

Esta primera muestra está compuesta por 193 sujetos escolarizados en 5º y 6º curso de Primaria, con edades comprendidas entre los 10 y 12 años, con un promedio de 11.00 años de edad y una desviación típica de 0.74. La distribución por sexos muestra que el 58.5 % son varones y el 41.5 % mujeres. En relación al nivel socioeconómico de los alumnos seleccionados, éste ofreció éste presentó para el nivel socioprofesional del padre la siguiente distribución: el 0.5 % altos directivos, el 4.7 % profesionales con titulación susuperior, el 5.7 % profesionales con titulación media, el

23.3 % profesionales cualificados con formación profesional, el 45.1 % profesionales no cualificados, el 11.9 % autónomos y el 1 % parados, siendo el 7.8 % datos faltantes en dicha variable. Por su parte, la distribución obtenida para la madre fue la siguiente: el 2.1 % del total por profesionales con titulación superior, el 5.7 % profesionales con titulación media, el 6.2 % cualificados con formación profesional, el 32.1 % profesionales no cualificados, el 5.7 % autónomos, el 0.5 % en paro y el 44 % amas de casa, siendo el 3.6 % datos faltantes en dicha variable.

*2b) Muestra escolarizada en un programa de incorporación progresiva al valenciano (LE V) y lengua familiar valenciano (LF V) - BADYG 2: Nivel Medio.*

De los 285 individuos que constituyen esta muestra, el 54.4 % son varones, el 45.6 % mujeres. Todos ellos están escolarizados en 1º E.S.O. y 2º E.S.O, y la distribución de sus edades se sitúa entre los 12 y los 15 años de edad, siendo su media y desviación típica en años de 12.93 y 0.73, respectivamente. En cuanto su nivel socioeconómico de los padres, señalar que éste presentó para caso del miembro paterno la siguiente distribución: el 0.4 % miembros de dirección o altos directivos, el 4.2 % profesionales con titulación superior, el 7.4 % profesionales con titulación media, el 21.4 % profesionales cualificados con formación profesional, el 43.5 % profesionales no cualificados, el 18.9 % autónomos, el 0.7 % jubilados y otro 0.7 en paro, siendo el 2.8 % datos faltantes. Por su parte, la distribución obtenida para el miembro materno se detalla a continuación: el 0.7 % profesionales con titulación superior, el 5.6 % profesionales con titulación media, el 8.1 % profesionales cualificados con formación profesional, el 24.9 % profesionales no cualificados, el 7 % autónomos y el



53.3 % restante amas de casa, siendo el 1.4 % datos faltantes en dicha variable.

(3) MODELO EDUCATIVO CON LENGUA DE ESCOLARIZACION EN VALENCIANO (LE V)  
Y LENGUA FAMILIAR VALENCIANO (LF V).

Dentro de este modelo educativo en particular, es conveniente resaltar que la única muestra cuyos datos son analizados en el presente estudio está compuesta por los sujetos a los que se les aplicó el BADYG elemental. Como se ha comentado anteriormente, la reciente y progresiva implantación de este tipo de programas ha impedido la realización de un estudio completo de este modelo educativo.

*3a) Muestra cuya lengua de enseñanza es el valenciano (LE V) y lengua familiar valenciano (LF V) - BADYG 1: Nivel Elemental.*

Esta muestra está compuesta por 249 alumnos escolarizados en 4º, 5º y 6º curso de Primaria, de los cuales 48.2 % son varones y el 51 % mujeres. La distribución de la edad se extiende desde los 9 y los 12 años de edad, siendo su edad media de 10.80 años y su desviación típica de 0.89. Respecto al nivel socioeconómico de los padres, éste presentó para el nivel socioprofesional del padre la siguiente distribución: el 4.0 % profesionales con titulación superior, el 6 % profesionales con titulación media, el 15.3 % profesionales cualificados con formación profesional, el 51.4 % profesionales no cualificados, el 18.1 % autónomos, el 0.8 % jubilados y el 0.8 % en paro, siendo el 3.6 % datos faltantes. En cambio, la distribución obtenida para el nivel socioprofesional de la madre estuvo formada por: el 2 % profesionales con titulación superior, el 6.4 % profesionales con

titulación media, el 10.8 % profesionales cualificados con formación profesional, el 32.1 % profesionales no cualificados, el 12 % autónomos, y el 34.1 % restante amas de casa, siendo el 2.4 % datos faltantes en dicha variable.

Para una mejor comprensión de todo lo expuesto hasta ahora, la siguiente tabla ofrece un resumen de los datos sociodemográficos aportados para cada uno de los diferentes grupos objeto de estudio.

**Tabla 4.5.3**

Tabla resumen de los datos sociodemográficos en las diferentes muestras utilizadas.

	N	SEXO		EDAD		
		% Hb	% Muj.	$\bar{X}$	$S_x$	
BADYG Nivel Elemental						
Muestra LF C LE C	333	49.2	50.8	11.06	0.78	
Muestra LF V LE CV	193	58.5	41.5	11.01	0.75	
Muestra LF V LE V	249	48.2	51	10.80	0.89	
BADYG Nivel Medio						
Muestra LF C LE C	316	47.8	52.2	12.92	0.83	
Muestra LF V LE CV	285	4.4	45.6	12.93	0.73	

En definitiva, podemos apreciar que en conjunto las muestras utilizadas en nuestro estudio no presentan grandes diferencias respecto a las variables de sexo, edad y nivel socioeconómico. No obstante, se llevaron a cabo diferentes pruebas estadísticas para confirmar la igualdad en las dos muestras utilizadas en cada comparación realizada en cuanto a la proporción

de varones y mujeres, y de la media y varianza de edad de los sujetos. Respecto a la variable sexo, se realizó una prueba estadística para comprobar la hipótesis acerca de la igualdad de las proporciones de varones y mujeres en las dos muestras utilizadas para cada una de las diferentes comparaciones establecidas. Los resultados obtenidos permiten confirmar que no había diferencias significativas entre los diferentes grupos.

En cuanto a la variable edad se realizó también una prueba estadística de comprobación de hipótesis acerca de la igualdad de las medias y varianzas de la edad de los sujetos. Tras aplicar la prueba t, se observó que dicha variable presentaba diferencias estadísticamente significativas en los grupos LF C LE C y LF V LE V ( $t=3.66$ ,  $p<0.001$ ). La existencia de dichas diferencias en la variable edad se debe a la incorporación de 4º de Primaria en el grupo LF V LE V, sin embargo, si observamos tal diferencia ésta no resulta ser muy elevada (LF C LE C  $\bar{X}=11.06$  y LF V LE V  $\bar{X}=10.80$ ).

#### 4.6. ANALISIS

En primer lugar se realizó un análisis de las propiedades psicométricas de las dos escalas utilizadas en el presente estudio. En segundo lugar, se analizó la estructura factorial de las mismas, con el fin último de averiguar en qué medida las citadas escalas cumplen el supuesto de unidimensionalidad. Y por último, se procedió a la detección del DIF en dichas escalas tras ser aplicadas a los grupos con características lingüísticas diferentes.

#### 4.6.1. Fiabilidad.

La fiabilidad en ambas escalas se estimó mediante la aplicación de la fórmula KR-20 (Kuder y Richardson, 1937), implementada en el programa estadístico SPSS (Norusis, 1990).

#### 4.6.2. Dimensionalidad.

El estudio de la unidimensionalidad del constructo medido por ambas escalas será el siguiente paso a realizar. Como se comentó en el capítulo anterior, hasta hace poco la mayoría de los métodos estadísticos establecidos para la detección del funcionamiento diferencial (DIF) en los ítems de un determinado test asumían la naturaleza unidimensional de la variable objeto de estudio, esto es, que una única aptitud o constructo era suficiente para explicar los resultados obtenidos por los sujetos en relación a los ítems componentes del test. Bajo estas condiciones se asumía que el rendimiento ante un determinado ítem puede ser explicado adecuadamente por un solo rasgo o constructo, aunque los propios investigadores eran conscientes de que este supuesto era demasiado exigente debido a los múltiples factores que podían afectar en un momento dado a las respuestas dadas a los ítems de un test, como por ejemplo la atención, la motivación de los sujetos, la ansiedad, etc... La aceptación de estos modelos generalmente denominados unidimensionales requieren que el rendimiento en el test sea explicado por un factor principal o dominante medido por el propio test. Por el contrario, si se asume la necesidad de más de un factor para explicar el rendimiento en dicho test, entonces éstos son denominados *multidimensionales*. Sin embargo, en la actualidad estos

últimos modelos todavía se encuentran escasamente desarrollados y en consecuencia no serán tratados aquí.

Respecto a la comprobación de la unidimensionalidad, la literatura especializada en el tema propone varios procedimientos de análisis (Hambleton y Swaminathan, 1985). Para una revisión de los distintos métodos disponibles puede consultarse los trabajos de Muñiz y Cuesta (1993), Hambleton y Rovinelli (1986) o Hattie (1984). De todos modos, el análisis factorial sigue siendo el método por excelencia, el más utilizado en los estudios empíricos realizados (Botempo, 1993; Budgell y cols, 1995; Cohen y Kim, 1992; Ellis, 1989; Ellis y cols., 1993; Ellis y Kimmel, 1992; Hulin y Mayer, 1986). Dentro de este procedimiento, el cálculo de la dimensionalidad se convierte en un problema de grado: *cuánta más varianza sea capaz de explicar el primer factor más apoyo existirá en favor de la unidimensionalidad del test analizado*. Al respecto, Lumsten (1961) propone como índice de unidimensionalidad el cálculo del cociente entre la varianza explicada por el primer factor y la explicada por el segundo. Por su parte, Lord (1980) señala que el estudio de la unidimensionalidad en un instrumento de medida requiere el empleo de métodos estadísticos más rigurosos. Precisamente, uno de los problemas conocidos tras someter datos dicotómicos a análisis factorial son los denominados "factores de dificultad", esto es, la obtención de factores que dependen en buena parte del grado de dificultad de los items analizados.

En la presente investigación, la unidimensionalidad de los datos se analizó mediante una aplicación del análisis de componentes principales sobre la matriz de correlaciones tetracórica de los items componentes del test. Dichas correlaciones se obtuvieron con el programa PRELIS (Jöreskog y

Sörbom, 1993) y sobre su matriz se realizó el análisis de componentes principales con el programa SPSS-FACTOR (Norusis, 1990). El procedimiento de estimación empleado estuvo basado en el método de mínimos cuadrados no ponderados (ULS) el cual minimiza las diferencias al cuadrado entre las proporciones de aciertos muestrales y las estimadas según el modelo ajustado (Ferrando, 1994; Martínez Arias, 1994). De hecho, Hambleton y Rovinelli (1986) lo consideran un procedimiento adecuado para evaluar la dimensionalidad de conjuntos de ítems dicotómicos. Por lo que respecta a la interpretación de la unidimensionalidad, una práctica muy extendida para obtener un estimador de la unidimensionalidad consiste en elaborar algún índice basado en la magnitud del primer valor propio obtenido al factorizar la matriz de correlación o, lo que es lo mismo, evaluar la proporción de varianza total explicada por el primer componente obtenido. La idea básica que subyace a la elaboración de dichos índices consiste en que, si el primer componente explica una buena proporción de la varianza total, en tanto que las aportaciones de los sucesivos componentes son mucho menores, entonces se puede considerar al conjunto de ítems analizado como aproximadamente unidimensional. La literatura consultada destaca dos opciones diferentes. La primera de ellas (Lord, 1980, p.21), afirma que existe unidimensionalidad en el constructo evaluado cuando el patrón de valores propios obtenido muestra que (a) el valor propio para el primer componente es más grande comparado con el del segundo, y que (b) el segundo valor propio no es mayor que los restantes. Y la segunda opción propuesta por Reckase (1979), señala que se podrá asumir la unidimensionalidad del rasgo evaluado si la contribución del primer componente en una solución no rotada, tras someter a análisis de componentes principales a una matriz de correlaciones tetracórica, es al menos del 20 % de la varianza total explicada.

#### 4.6.3. Análisis estadístico del funcionamiento diferencial (DIF)

El presente estudio decidió optar por los análisis estadísticos de detección de DIF con datos dicotómicos. Es importante destacar que la información que se pierde con la utilización de datos dicotómicos es a bajo precio, debido a que el conocimiento sobre el comportamiento de los parámetros a estimar en este tipo de modelos es bastante amplio y exhaustivo. Al respecto, Hulin (1987) señala que el empleo de puntuaciones adecuadamente dicotomizadas puede proporcionar tanto información del rasgo o constructo medido por un test que la mayoría de escalas con respuesta politómica no óptimas. Además, recientes estudios realizados (Hidalgo y Gómez, 1997) sobre la idoneidad de los datos dicotómicos frente a los politómicos así lo demuestran. Por lo tanto, las respuestas dadas a los items de cada prueba se dicotomizaron. La respuesta correcta a cada ítem se puntuó con un valor de 1, y el resto de respuestas incorrectas fueron computadas con valores de 0. Una vez realizada la dicotomización de las puntuaciones, se procedió al análisis del DIF en ambas pruebas.

El método utilizado fue el estadístico Mantel-Haenszel (MH) (Holland y Thayer, 1988). La selección de este procedimiento se realizó teniendo en cuenta ciertas normas de uso, efectividad y disponibilidad de recursos informáticos correspondientes a dicho método. Precisamente, la aplicación dicho método se realizó mediante el programa MHDIF (Fidalgo, 1994). El MHDIF implementa la técnica Mantel-Haenszel (MH) propuesta por Holland y Thayer (1988) para detectar la presencia de DIF uniforme y no uniforme en datos dicotómicos de un test. Sus principales limitaciones se encuentran en un máximo de 75 items y 5000 sujetos. Su aplicación ofrece el estadístico



$\chi^2_{MH}$  que permite evaluar la hipótesis nula de ausencia de DIF, y el índice MH D-DIF una medida de la magnitud del tamaño del DIF para cada ítem. Para el cálculo del DIF uniforme, el programa computa el estadístico MH de forma estándar. Mientras que para el cálculo del DIF no uniforme, el MHDIF utiliza una modificación de cálculo, propuesta por Mazor, Clauser y Hambleton (1994) en el estadístico MH. Para más información, una descripción del mismo se ofrece en el capítulo 3 dentro del apartado referido al procedimiento Mantel-Haenszel. Algunos trabajos de simulación afirman que la citada variación en el método MH es una alternativa útil para detectar el DIF no uniforme, no obstante por contrapartida existe un incremento de la tasa de error tipo 1 (Fidalgo, 1996; Fidalgo y Mellenbergh, 1995). Además, el citado programa implementa una versión bietápica de purificación del criterio (Holland y Thayer, 1988), esto es, una vez calculado el estadístico  $\chi^2_{MH}$  para cada ítem, son eliminados aquellos ítems con valores significativos en dicho estadístico y se vuelve a recalcular los estadísticos  $\chi^2_{MH}$ , utilizando como criterio la puntuación total del sujeto en los ítems restantes. Un aspecto importante a considerar es que cuando un ítem está siendo analizado, éste se incluye en el criterio aunque previamente haya presentado problemas de funcionamiento diferencial.

En el presente estudio, se declaró que los ítems analizados presentaban DIF uniforme y no uniforme cuando el estadístico  $\chi^2_{MH}$  para cada tipo de DIF en cada ítem resultó ser estadísticamente significativo, asociado a una probabilidad de  $p \leq 0.05/n^2$  de ítems de la escala analizada (H.M.V.:  $p \leq 0.05/40=0.0012$ ; H.M.NV.:  $p \leq 0.05/40=0.0012$ ). La justificación de esta forma de proceder representa un compromiso entre el nivel de significación adoptado por comparación y el extenso número de comparaciones hechas en el estudio. Además, en cuanto a la detección del DIF no uniforme,

es importante destacar que algunos trabajos de simulación realizados (Fidalgo, 1996; Fidalgo y Mellenbergh, 1995) muestran un incremento de la tasa de error tipo 1 en el cálculo del estadístico MH para el DIF no uniforme. Así pues, tal vez resulte más apropiado considerar los resultados en el DIF no uniforme al nivel de significación de 0.005, como un intento de prevenir la aparición de dicho error y, en consecuencia, asegurar que los items poseen DIF cuando éste realmente está presente. Sin embargo, dado que el programa informático empleado trabaja con los niveles 0.005 y 0.001, en este trabajo usaremos el nivel de significación de 0.005.



## **CAPITULO 5**

### **Resultados**



### **5.1. ANALISIS DESCRIPTIVOS**

Los análisis descriptivos correspondientes a las dos escalas H.M.V. y H.M.NV. utilizadas en este estudio de investigación son presentados en las tablas 5.1.1. y 5.1.2, en función del nivel de aplicación BADYG Elemental y Medio. La contribución de estos resultados puede aportar un primer indicio acerca de ciertas desigualdades o diferencias entre las muestras seleccionadas y las escalas aplicadas.

**Tabla 5.1.1.**  
Análisis descriptivos para las escalas H.M.V. y H.M.NV. en el *BADYG Elemental*.

	Nº Items	$\bar{X}$	$S_x^2$	KR-20
<i>Habilidad Mental Verbal (HMOV)</i>				
Muestra LF C LE C	40	21,71	37,32	0,842
Muestra LF V LE CV	40	21,49	42,92	0,860
Muestra LF V LE V	39	21,25	42,02	0,860
<i>Habilidad Mental No Verbal (HMNV)</i>				
Muestra LF C LE C	40	22,10	52,62	0,866
Muestra LF V LE CV	40	23,92	51,93	0,872
Muestra LF V LE V	40	23,53	59,44	0,888

El cálculo de la media aritmética y la varianza de ambas escalas en el *BADYG Elemental* permite observar que la escala H.M.NV., compuesta por ítems no verbales, ofreció un mejor rendimiento en los dos grupos focales que en el grupo de referencia, presentando en este caso diferencias estadísticamente significativas (*Muestra LF C LE C vs. Muestra LF V LE CV*  $t=2.78$ ,  $p=0.006$ ; *Muestra LF C LE C vs. Muestra LF V LE CV*  $t=-2.28$ ,  $p=0.023$ ).

Respecto al modelo *BADYG Medio*, las desigualdades descubiertas en la escala H.M.NV. permiten observar que los escolares pertenecientes al grupo focal (*muestra LF V LE CV*) presentan mejores puntuaciones en dicha escala que los componentes del grupo de referencia (*muestra LF C LE C*) ( $t=-2.55$ ,  $p=0.011$ ).

**Tabla 5.1.2.**  
Análisis descriptivos para las escalas H.M.V. y H.M.NV. en el *BADYG Medio*.

	Nº Items	$\bar{X}$	$S_x^2$	KR-20
<i>Habilidad Mental Verbal (HMV)</i>				
Muestra LFCLE C	40	20,62	39,35	0,830
Muestra LFVLE CV	40	20,00	43,36	0,838
<i>Habilidad Mental No Verbal (HMNV)</i>				
Muestra LFCLE C	40	16,12	46,86	0,843
Muestra LFVLE CV	40	17,63	52,62	0,858

En relación a la fiabilidad de ambas escalas, ésta se estimó mediante el coeficiente KR-20 para datos dicotómicos. Tal y como se puede apreciar en las tablas 5.1.1. y 5.1.2., el coeficiente de fiabilidad en la escala H.M.V. oscila entre valores de 0.83 a 0.86 en las diferentes muestras consideradas. Por su parte, la escala H.M.NV. incluye valores que van de 0.84 a 0.88. En definitiva, en conjunto podemos afirmar que ambas escalas ofrecen una buena consistencia interna en los diferentes grupos sometidos a comparación.

## 5.2. UNIDIMENSIONALIDAD

Con el fin de comprobar la unidimensionalidad de las dos escalas empleadas, primero se calculó la matriz de correlaciones tetracórica de items para cada condición -escala y muestra aplicada- mediante el programa PRELIS (Jöreskog y Sörbom, 1993), y segundo, ésta se sometió a un análisis de componentes principales con el paquete estadístico SPSS-FACTOR (Norusis,



1990). Los resultados conseguidos tras la realización de dichos análisis se presentan a continuación:

**Tabla 5.2.1.**

Valores propios y proporción de varianza explicada por los 4 primeros componentes para las escalas *H.M.V.* y *H.M.NV.* en el *BADYG Nivel Elemental*.

VALORES PROPIOS								
	F1	% Varianza explicada	F2	% Varianza explicada	F3	% Varianza explicada	F4	% Varianza explicada
<i>Escala HMV</i>								
Muestra LFC LEC	6,15	15,4	2,49	6,2	1,59	4,0	1,54	3,9
Muestra LFV LECV	8,04	20,1	3,31	8,3	1,74	4,4	1,61	4,0
Muestra LFV LEV*	6,60	16,9	2,63	6,8	1,72	4,4	1,49	3,8
<i>Escala HMNV</i>								
Muestra LFC LEC	7,11	17,8	2,17	5,4	1,59	4,0	1,45	3,6
Muestra LFV LECV	8,41	21,0	1,98	5,0	1,66	4,2	1,58	4,0
Muestra LFV LEV	8,25	20,6	1,93	4,8	1,72	4,3	1,50	3,8

\*El ACP de la escala *H.M.V.* para la muestra *LFV LEV* se realizó únicamente con 39 ítems, dado que el ítem 11 fue eliminado de dicha muestra por falta de variabilidad.

Un estudio de los valores propios y de la proporción de varianza explicada por los componentes extraídos tras factorizar la matriz de correlaciones tetracórica de ítems en las dos pruebas y en cada muestra estudiada permite a simple vista la detección de un componente principal el cual explica una buena proporción de la varianza total del test y un punto de inflexión que se sitúa en el segundo componente, siendo las aportaciones de los valores propios en el resto de componentes mucho menores.

Tabla 5.2.2.

Valores propios y proporción de varianza explicada por los 4 primeros componentes para las escalas H.M.V. y H.M.NV. en el BADYG Nivel Medio.

VALORES PROPIOS								
	F1	% Varianza explicada	F2	% Varianza explicada	F3	% Varianza explicada	F4	% Varianza explicada
<i>Escala HMV</i>								
Muestra LFCLEC	5,53	13,8	2,10	5,3	1,58	4,0	1,54	3,9
Muestra LFVLECV	6,67	16,7	2,17	5,4	1,69	4,2	1,52	3,8
<i>Escala HMNV</i>								
Muestra LFCLEC	6,64	16,6	2,00	5,0	1,54	3,9	1,50	3,8
Muestra LFVLECV	8,83	22,1	1,73	4,3	1,58	4,0	1,49	3,7

En la muestra monolingüe castellana (LF C LE C), las escalas de H.M.V. y H.M.N.V. del BADYG Elemental presentaron 14 y 12 componentes respectivamente, con valores propios superiores a la unidad y que explicaban el 59.1 % y 54.7 % de la varianza total en cada escala. Por su parte, las citadas escalas en el BADYG Medio ofrecieron un total de 14 y 13 componentes para cada una, siendo la aportación de varianza explicada en ellos del 57.1% y 55.9 % respectivamente. En la muestra bilingüe con lengua familiar valenciana (LF 2 LE 2), los resultados conseguidos en el BADYG Elemental mostraron una extracción de 12 componentes para la escala de H.M.V. con un valor propio superior a la unidad y un porcentaje de varianza explicada del 61.4 %, y de 13 componentes para la escala H.M.NV., y una varianza total del 61.1 %. Respecto al comportamiento de dichas escalas en el BADYG Medio, señalar que éstas ofrecen un total de 13 y 11 componentes respectivamente, siendo la varianza explicada por ellas de 56.9 % para la H.M.V. y de 55.3 % para la H.M.NV. Por último, en relación a la muestra bilingüe con lengua familiar y escolar valenciana (LF 2 LE 3) es

importante destacar que la unidimensionalidad de la escala H.M.V. se realizó únicamente con 39 de los 40 ítems que formaban la escala, debido a que el ítem 11 de la citada escala fue eliminado de la misma por falta de variabilidad, siendo un ítem contestado incorrectamente por la totalidad de los sujetos participantes en la presente muestra. El número de componentes extraídos para la escala H.M.V. fue de 12 componentes, siendo la varianza total explicada del 57.4 %. Por su parte, la escala H.M.NV. también presentó 12 componentes con valores propios mayores que la unidad, los cuales a su vez explicaban el 57.4 % de la varianza total.

Ante los resultados obtenidos, la existencia de la unidimensionalidad para ambas escalas en cada una de las muestras analizadas puede resultar no del todo concluyente. Teniendo en cuenta los diversos índices que existen para la evaluación de la unidimensionalidad de los ítems mediante análisis de componentes principales (ACP) (Hambleton y Rovinelli, 1986; Hattie, 1984, 1985; Muñoz y Cuesta, 1993), a continuación se aplicó un método propuesto por Lord (1980) basado en el cálculo de la razón entre la diferencia de los dos primeros valores propios y la diferencia entre el segundo y el tercero. Los valores obtenidos en este índice por el BADYG elemental fueron de 4.06, 3.01 y 4.36 para la escala H.M.V. en las muestras LF 1 LE 1, LF 2 LE 2 y LF 2 LE 3, respectivamente, y de 8.51, 20.09 y 30.09 para la escala H.M.NV. Por su parte, los valores presentados para el BADYG medio fueron de 6.59 y 9.37 para la escala de H.M.V., y de 10.08 y 47.33 para la de H.M.NV. en las muestras LF 1 LE 1 y LF 2 LE 2, respectivamente. En la obtención de dichos resultados resulta evidente el mayor grado de unidimensionalidad de la escala H.M.NV. respecto a la de H.M.V. Por otra parte, si tomamos en consideración el criterio propuesto por Reckase (1979)

para evaluar la unidimensionalidad de dichas escalas, se puede observar que en sólo una ocasión para la escala H.M.V. y tres para la escala H.M.NV. de las cinco aportadas la contribución del primer componente de los análisis factoriales realizados es mayor del 20 % de la varianza explicada total.

En definitiva, a partir de los resultados obtenidos podemos decir que si parece que exista una tendencia de la unidimensionalidad del constructo o rasgo evaluado en las dos pruebas o escalas utilizadas.

### 5.3. EVALUACION DEL FUNCIONAMIENTO DIFERENCIAL (DIF)

La evaluación estadística de la presencia de DIF en las escalas H.M.V. y H.M.NV. del presente estudio se realizó con la aplicación del estadístico Mantel-Haenszel (HM) mediante el programa MHDIF (Fidalgo, 1994). Este programa permite la detección del funcionamiento diferencial DIF uniforme así como del DIF no uniforme tras incluir las aportaciones realizadas por Mazor, Clauser y Hambleton (1994). También incluye un proceso de purificación de la puntuación total del test utilizada como criterio en la formación de los diferentes niveles de aptitud comparables entre sí (Holland y Thayer, 1988). Además, el ítem analizado es incluido en el cálculo de la puntuación total, aunque éste previamente haya sido detectado con problemas de funcionamiento diferencial en el test.

A continuación, vamos a presentar los resultados obtenidos en los análisis correspondientes a la detección del funcionamiento diferencial uniforme y no uniforme en las escalas H.M.V. y H.M.NV. tras ser aplicadas a diferentes grupos de sujetos. Bajo la denominación de DIF uniforme, cada

tabla ofrece los valores obtenidos en el parámetro  $\hat{\alpha}_{MH}$ , un estimador de la magnitud del DIF localizado, y su transformación a la escala delta MH D-DIF, sin y con purificación de la variable criterio a un nivel de significación de 0.005. Bajo el enunciado de DIF no uniforme, se presentan los valores obtenidos en los parámetros  $\hat{\alpha}_{MH}$  y MH D-DIF de los dos subgrupos resultantes tras dividir a la muestra total en dos partes, siendo el grupo inferior (subgrupo 1:  $\hat{\alpha}_{MH1}$ ) el formado por los sujetos con una puntuación total menor o igual a la media de las puntuaciones en el test, y el grupo superior (subgrupo 2:  $\hat{\alpha}_{MH2}$ ) el compuesto por los sujetos con puntuaciones totales situadas por encima de la media muestral. En ciertas ocasiones, algunos items presentan el símbolo "---", lo que indica que los índices de DIF de tales items no pudieron ser calculados debido a problemas de estimación en sus parámetros. Un ejemplo claro se localiza en el ítem 11 de la escala H.M.V. aplicada al grupo LF V LE V. Recordemos que este ítem fue eliminado en el cálculo de la unidimensionalidad debido a que éste no presentó variabilidad en su respuesta, esto es, que ningún sujeto de la mencionada muestra fue capaz de responder correctamente a dicho ítem.

La primera comparación a abordar será la establecida entre la muestra LF C LE C vs. muestra LF V LE CV. Los resultados obtenidos de dicha comparación en la detección del DIF uniforme y no uniforme en las escalas H.M.V. y H.M.NV. son presentados en las tablas 5.3.1. y 5.3.2., respectivamente.

En la detección del DIF uniforme, la puntuación total del test se dividió en 33 categorías correspondientes al rango de puntuaciones obtenidas, siendo su puntuación mínima de 4 y su máxima de 37, y bajo la

condición de que el número de sujetos en cada categoría considerada fuera al menos de uno. Tras una primera etapa de detección del DIF, la presente escala no presentó problemas de funcionamiento diferencial en sus ítems.

En la detección del DIF no uniforme, los grupos de referencia y focal fueron divididos de dos subgrupos en función de la media aritmética obtenida en la muestra total. De esta forma, el estadístico MH en sus dos etapas se aplicó a cada uno de los grupos considerados, el grupo inferior (subgrupo 1) y el grupo superior (subgrupo 2) situados por debajo y por encima de la media respectivamente. En este caso, la media aritmética del grupo total fue igual a 21.63. Así pues en la primera etapa en el grupo inferior la puntuación total se estratificó en 18 intervalos diferentes de puntuación con una puntuación mínima de 4 y una máxima de 22, y el grupo superior se clasificó en 15 niveles distintos, los correspondientes a cada una de las puntuaciones dentro del intervalo rango 23-37. La tasa de detección de DIF no uniforme para la citada escala fue calculada a un nivel de significación de 0.005, y los análisis realizados no encontraron ítems con problemas de funcionamiento diferencial en los dos grupos considerados. También, destacar que la estimación del DIF no uniforme en el ítem 7 para el grupo superior ( $\hat{\alpha}_{MH2}$ ) no fue posible, debido a su nula variabilidad en las respuestas dadas por los sujetos integrantes en uno de los dos grupos sometidos a comparación.

En relación a la escala H.M.NV., a la vista de los resultados obtenidos en la tabla 4.3.2 se puede apreciar que la hipótesis formulada acerca de la no existencia de funcionamiento diferencial en la escala H.M.NV. se cumple al nivel de significación considerado.

La segunda comparación realizada fue entre la muestra *LF C LE C* y la muestra *LF V LE V*. La evaluación del DIF uniforme y no uniforme correspondiente a las escalas H.M.V. y H.M.NV. se encuentra presente en las tablas 5.3.3. y 5.3.4., respectivamente.

Para la primera escala H.M.V., sin purificación del criterio los análisis estadísticos realizados detectaron la presencia de 3 items (items 10, 22 y 26) con DIF uniforme. Tras una purificación del criterio, los items detectados con problemas de DIF fueron los mismos.

Respecto a la interpretación de resultados, el índice de DIF mostrado por el estadístico  $\hat{\alpha}_{MH}$  con valores mayores que uno indicará que el ítem estudiado favorece al grupo de referencia formado por la muestra *LF C LE C*, mientras que valores por debajo de la unidad indicarán que el ítem está favoreciendo al grupo focal, en este caso la muestra *LF V LE V*. En cuanto al índice MH D-DIF, valores próximos a 0 serán un indicativo de ausencia de DIF, valores negativos que el ítem favorece al grupo de referencia, y valores positivos que éste favorece al grupo focal. Atendiendo a los resultados obtenidos en esta segunda comparación, podemos observar que en la escala H.M.V. todos los items presentaron un funcionamiento diferencial en contra del grupo focal, formado por sujetos cuya lengua familiar y de escolarización es el valenciano, a excepción del ítem 10 que ofreció un funcionamiento diferencial en sentido contrario a lo esperado favoreciendo a los sujetos pertenecientes al grupo focal. Otro comentario a destacar para esta escala es la aparición de varios items con problemas de estimación en sus índices de DIF, debido a los motivos expuestos

anteriormente. Tales items son el ítem 7 para el cálculo del DIF no uniforme en el subgrupo superior ( $\hat{\alpha}_{MH2}$ ), al igual que en la anterior comparación, y el ítem 11 para ambos tipos de DIF, que si recordemos fue eliminado de dicha escala para calcular la unidimensionalidad de la misma a causa de su invariabilidad en las respuestas dadas por los sujetos integrantes en el grupo focal.

Por lo que respecta a la escala H.M.NV., ésta no presentó problemas de DIF uniforme, pero sí de DIF no uniforme. Independientemente de la purificación del criterio, esta escala detectó al ítem 38 con un funcionamiento diferencial no uniforme. En cuanto a la dirección del DIF, destacar que dicho ítem presentó un Índice con signo positivo a favor del grupo focal, compuesto por alumnos cuya lengua familiar y escolar es el valenciano.

La tercera y última comparación se realizó entre la muestra *LF C LE C* vs. muestra *LF V LE CV* para las dos pruebas H.M.V y H.M.NV. del BADYG medio. Los análisis correspondientes a la detección del DIF uniforme y no uniforme en ambas escalas se exponen en las tablas 5.3.5. y 5.3.6., respectivamente.

En la escala H.M.V., los resultados obtenidos mostraron después de purificar la medida de habilidad que 4 de los 40 items pertenecientes a dicha escala presentaban problemas de funcionamiento diferencial. En concreto, los items 5, 10, 19 y 38 mostraron DIF uniforme, mientras que los items 5 y 38 presentaron DIF no uniforme. En cuanto a la dirección del DIF localizado, señalar que todos los items detectados en esta escala a



excepción del ítem 10 exhibieron un funcionamiento diferencial a favor del grupo de referencia y en contra del grupo focal, formado por alumnos cuya lengua familiar es el valenciano pero cuyo programa de enseñanza es la incorporación progresiva al valenciano.

En último lugar, se realizó la detección del DIF uniforme y no uniforme para la escala H.M.NV. del BADYG medio. Los resultados aportados en la tabla 5.3.6. muestran claramente que dicha escala no presentó ningún problema de funcionamiento diferencial en sus ítems, tras ser aplicada a los dos grupos sometidos a comparación.

**Tabla 5.3.1.**

Funcionamiento diferencial de los items en la comparación  
 Muestra LLF 1 LE 1 vs. Muestra LF 2 LE 2 para la escala H.M.V. del BADYG  
 Nivel Elemental.

Items	DIF UNIFORME				DIF NO UNIFORME					
	ETAPA1		ETAPA2		ETAPA1		ETAPA2			
	$\hat{\alpha}_{MH}$	MMH D-DIF	$\hat{\alpha}_{MH}$	MMH D-DIF	$\hat{\alpha}_{MH1}$	$\hat{\alpha}_{MH2}$	MMH D-DIF	$\hat{\alpha}_{MH1}$	$\hat{\alpha}_{MH2}$	MMH D-DIF
1	0.89	0.27			0.92	0.84	0.19	0.40		
2	1.08	-0.19			1.25	0.63	-0.53	1.10		
3	0.67	0.94			0.74	0.50	0.71	1.62		
4	1.48	-0.92			2.08	0.49	-1.72	1.69		
5	1.04	-0.08			0.95	1.76	0.12	-1.33		
6	1.41	-0.80			1.01	2.84	-0.03	-2.46		
7	4.75	-3.66			4.75	--	-3.66	--		
8	1.04	-0.10			1.04	1.03	-0.10	-0.08		
9	0.96	0.10			0.92	1.15	0.19	-0.32		
10	0.76	0.65			0.67	0.94	0.95	0.15		
11	0.51	1.56			0.63	0.32	1.09	2.64		
12	1.14	-0.31			1.06	1.33	-0.15	-0.66		
13	0.83	0.44			0.58	1.97	1.28	-1.60		
14	0.95	0.13			0.88	1.52	0.30	-0.98		
15	0.91	0.23			0.76	1.22	0.66	-0.46		
16	0.90	0.26			2.19	0.47	-1.84	1.76		
17	1.51	-0.97			1.60	1.43	-1.10	-0.84		
18	0.61	1.16			0.70	0.37	0.84	2.36		
19	1.43	-0.84			1.39	1.51	-0.77	-0.96		
20	0.77	0.62			0.86	0.68	0.37	0.91		
21	0.91	0.23			1.00	0.73	-0.01	0.74		
22	1.11	-0.24			1.09	1.13	-0.21	-0.28		
23	0.84	0.41			0.80	0.91	0.51	0.23		
24	0.57	1.34			0.59	0.53	1.24	1.48		
25	1.45	-0.87			1.29	1.75	-0.60	-1.32		
26	1.14	-0.31			1.11	1.19	-0.24	-0.41		
27	0.71	0.80			0.60	0.90	1.20	0.25		
28	0.93	0.18			0.76	1.16	0.66	-0.35		
29	1.23	-0.49			1.27	1.18	-0.57	-0.39		
30	0.92	0.19			1.67	0.61	-1.20	1.16		
31	0.90	0.25			0.68	1.04	0.90	-0.09		
32	1.53	-1.00			3.34	1.27	-2.83	-0.56		
33	1.41	-0.81			1.17	1.62	-0.37	-1.13		
34	1.16	-0.36			1.96	0.96	-1.59	0.10		
35	1.33	-0.68			2.67	1.13	-2.31	-0.30		
36	1.28	-0.58			4.12	1.09	-3.33	-0.21		
37	1.01	-0.01			1.25	0.96	-0.52	0.10		
38	0.52	1.53			0.62	0.49	1.13	1.70		
39	1.18	-0.39			1.44	1.09	-0.86	-0.20		
40	0.76	0.65			1.71	0.60	-1.25	1.19		
% DIF	(0.00				0.00	0.00				
*p<0.005										

Tabla 5.3.2.

Funcionamiento diferencial de los items en la comparación  
Muestra LF 1 LE 1 vs. Muestra LF 2 LE 2 para la escala H.M.NV. del BADYG  
Nivel Elemental.

Items	DIF UNIFORME				DIF NO UNIFORME					
	ETAPA 1		ETAPA 2		ETAPA 1			ETAPA 2		
	$\hat{\alpha}_{MH}$	MH D-DIF	$\hat{\alpha}_{MH}$	MH D-DIF	$\hat{\alpha}_{MH1}$	$\hat{\alpha}_{MH2}$	MH D-DIF	$\hat{\alpha}_{MH1}$	$\hat{\alpha}_{MH2}$	MH D-DIF
1	0.86	0.36			0.60	1.79	1.20	-1.37		
2	0.78	0.59			1.07	0.44	-0.15	1.94		
3	1.12	-0.27			1.06	1.21	-0.13	-0.46		
4	0.74	0.70			0.69	0.80	0.87	0.54		
5	0.67	0.95			0.60	0.81	1.21	0.49		
6	0.62	1.12			0.50	0.81	1.64	0.50		
7	1.01	-0.02			1.06	0.94	-0.14	0.14		
8	0.74	0.69			0.42	2.01	2.01	-1.64		
9	0.85	0.38			0.88	0.81	0.31	0.49		
10	0.99	0.01			0.80	1.21	0.52	-0.46		
11	0.86	0.34			0.71	1.05	0.79	-0.13		
12	1.09	-0.21			1.38	0.88	-0.76	0.31		
13	1.13	-0.28			1.06	1.30	-0.13	-0.61		
14	0.80	0.52			0.75	0.86	0.67	0.34		
15	1.09	-0.21			1.01	1.22	-0.03	-0.46		
16	1.39	-0.77			1.45	1.23	-0.87	-0.49		
17	1.04	-0.08			0.91	1.27	0.22	-0.56		
18	1.37	-0.74			1.43	1.26	-0.84	-0.54		
19	0.79	0.55			0.86	0.71	0.36	0.80		
20	1.11	-0.25			1.10	1.14	-0.23	-0.31		
21	0.78	0.58			0.52	1.14	1.55	-0.31		
22	0.97	0.07			1.24	0.58	-0.50	1.30		
23	1.67	-1.21			1.83	1.48	-1.42	-0.93		
24	0.80	0.54			1.12	0.37	-0.26	2.34		
25	1.40	-0.79			1.10	1.76	-0.23	-1.33		
26	1.42	-0.83			1.75	1.14	-1.32	-0.31		
27	1.06	-0.13			1.08	1.03	-0.17	-0.07		
28	1.03	-0.06			0.97	1.09	0.07	-0.21		
29	0.93	0.18			0.96	0.90	0.09	0.26		
30	1.10	-0.23			2.42	0.50	-2.08	1.61		
31	1.40	-0.78			1.29	1.46	-0.60	-0.89		
32	1.16	-0.35			1.76	0.49	-1.34	1.68		
33	1.11	-0.25			0.95	1.22	0.13	-0.47		
34	1.03	-0.07			0.67	1.25	0.93	-0.52		
35	1.05	-0.11			1.55	0.89	-1.03	0.28		
36	1.06	-0.13			1.23	1.00	-0.49	-0.01		
37	1.31	-0.63			1.07	1.54	-0.17	-1.01		
38	0.82	0.46			1.07	0.68	-0.16	0.90		
39	0.79	0.55			0.81	0.78	0.50	0.58		
40	1.85	-1.45			0.65	3.26	1.01	-2.78		
% DIF	0.00				0.00	0.00				
*p<0.005										

**Tabla 4.3.3.**  
 Funcionamiento diferencial de los items en la comparación  
 Muestra LF 1 LE 1 vs. Muestra LF 2 LE 3 para la escala H.M.V. del BADYG  
 Nivel Elemental.

Items	DIF UNIFORME				DIF NO UNIFORME					
	ETAPA 1		ETAPA 2		ETAPA 1		ETAPA 2			
	$\hat{\alpha}_{MH}$	MH D-DIF	$\hat{\alpha}_{MH}$	MH D-DIF	$\hat{\alpha}_{MH1}$	$\hat{\alpha}_{MH2}$	MH D-DIF	$\hat{\alpha}_{MH1}$	$\hat{\alpha}_{MH2}$	MH D-DIF
1	1.14	-0.30	1.08	-0.18	1.18	1.07	-0.39	-0.16		
2	1.31	-0.63	1.28	-0.58	1.20	1.65	-0.43	-1.17		
3	0.45	1.88	0.41	2.09	0.41	0.55	2.08	1.40		
4	0.85	0.37	0.82	0.48	1.21	0.26	-0.45	3.14		
5	0.82	0.45	0.83	0.43	0.95	0.00	0.12	999		
6	1.05	-0.11	0.99	0.02	0.92	1.45	0.19	-0.88		
7	1.71	-1.26	1.25	-0.53	1.32	--	-0.64	--		
8	1.32	-0.65	1.51	-0.96	1.21	1.61	-0.44	-1.12		
9	1.06	-0.13	1.06	-0.13	1.03	1.20	-0.08	-0.43		
10	0.50*	1.61	0.55*	1.41	0.53	0.46	1.51	1.81		
11	--	--	--	--	--	--	--	--		
12	1.37	-0.75	1.24	-0.50	1.56	0.98	-1.05	0.05		
13	0.83	0.43	0.90	0.26	0.89	0.65	0.27	1.01		
14	1.00	0.00	0.99	0.01	0.71	5.63	0.81	-4.06		
15	1.01	-0.02	1.01	-0.02	1.15	0.79	-0.33	0.56		
16	0.52	1.53	0.57	1.32	0.95	0.30	0.11	2.80		
17	1.17	-0.36	1.21	-0.45	1.34	1.02	-0.69	-0.04		
18	0.57	1.33	0.63	1.08	0.68	0.31	0.91	2.78		
19	1.04	-0.09	1.09	-0.19	1.17	0.82	-0.37	0.46		
20	1.42	-0.83	1.49	-0.94	1.36	1.48	-0.72	-0.92		
21	0.92	0.20	0.98	0.05	0.94	0.87	0.15	0.31		
22	1.93*	-1.54	1.83*	-1.42	2.03	1.80	-1.67	-1.38		
23	1.09	-0.19	1.23	-0.49	1.11	1.04	-0.25	-0.09		
24	0.77	0.61	0.78	0.59	0.83	0.71	0.45	0.81		
25	1.49	-0.93	1.47	-0.90	1.40	1.64	-0.79	-1.16		
26	1.82*	-1.41	1.82*	-1.41	1.72	1.96	-1.27	-1.58		
27	0.71	0.82	0.72	0.78	0.56	0.96	1.37	0.09		
28	0.68	0.90	0.67	0.93	0.58	0.82	1.28	0.48		
29	1.10	-0.22	1.10	-0.22	1.14	1.04	-0.31	-0.09		
30	0.83	0.45	0.90	0.25	1.05	0.67	-0.12	0.94		
31	0.75	0.67	0.77	0.60	0.45	1.05	1.87	-0.12		
32	1.11	-0.25	1.18	-0.39	1.37	1.04	-0.73	-0.10		
33	0.90	0.26	0.94	0.14	0.83	0.95	0.43	0.11		
34	0.78	0.59	0.84	0.41	1.11	0.67	-0.25	0.92		
35	1.13	-0.30	1.29	-0.59	0.97	1.19	0.06	-0.42		
36	1.46	-0.89	1.46	-0.88	1.57	1.44	-1.06	-0.85		
37	0.80	0.54	0.88	0.30	1.71	0.71	-1.27	0.80		
38	0.87	0.32	0.97	0.08	0.77	0.92	0.62	0.20		
39	0.88	0.29	0.92	0.20	0.46	1.30	1.82	-0.61		
40	1.36	-0.72	1.37	-0.73	1.45	1.34	-0.87	-0.68		
% DIF	7.50		7.50		0.00	0.00				
*p<0.005										

**Tabla 5.3.4.**  
 Funcionamiento diferencial de los items en la comparación  
 Muestra LF 1 LE 1 vs. Muestra LF 2 LE 3 para la escala H.M.NV. del BADYG  
 Nivel Elemental.

Items	DIF UNIFORME				DIF NO UNIFORME								
	ETAPA 1		ETAPA 2		ETAPA 1				ETAPA 2				
	$\hat{\alpha}_{MH}$	MH D-DIF	$\hat{\alpha}_{MH}$	MH D-DIF	$\hat{\alpha}_{MH1}$	$\hat{\alpha}_{MH2}$	MH D-DIF	$\hat{\alpha}_{MH1}$	$\hat{\alpha}_{MH2}$	MH D-DIF	$\hat{\alpha}_{MH1}$	$\hat{\alpha}_{MH2}$	MH D-DIF
1	1.04	-0.09			0.91	1.47	0.22 -0.90	1.12	0.94	-0.27	0.14		
2	0.77	0.63			1.08	0.39	-0.19 2.24	0.96	0.57	0.08	1.31		
3	1.62	-1.13			1.86	1.32	-1.46 -0.65	1.85	1.34	-1.44	-0.68		
4	1.07	-0.16			1.07	1.07	-0.16 -0.16	1.05	1.02	-0.12	-0.04		
5	1.05	-0.11			0.88	1.39	0.30 -0.78	0.82	1.41	0.48	-0.81		
6	0.85	0.39			0.84	0.87	0.42 0.34	0.99	0.78	0.01	0.58		
7	0.64	1.04			0.54	0.81	1.45 0.48	0.59	0.75	1.26	0.68		
8	0.95	0.12			1.01	0.72	-0.02 0.77	1.32	0.46	-0.66	1.81		
9	0.86	0.34			0.78	1.02	0.59 -0.04	0.64	1.01	1.04	-0.01		
10	1.27	-0.56			1.23	1.30	-0.48 -0.62	1.43	1.25	-0.85	-0.52		
11	0.91	0.22			1.02	0.78	-0.05 0.57	1.11	0.80	-0.25	0.53		
12	0.96	0.10			1.07	0.85	-0.15 0.38	1.01	0.93	-0.01	0.18		
13	0.99	0.03			1.04	0.81	-0.09 0.49	0.92	1.06	0.20	-0.14		
14	0.91	0.23			1.22	0.61	-0.47 1.14	1.21	0.63	-0.45	1.09		
15	1.00	0.01			0.85	1.32	0.38 -0.66	1.00	1.02	0.01	-0.06		
16	1.25	-0.53			1.37	0.98	-0.74 0.05	1.36	0.90	-0.73	0.26		
17	0.75	0.69			0.69	0.87	0.86 0.32	0.60	1.16	1.19	-0.36		
18	0.96	0.10			1.02	0.82	-0.05 0.46	1.11	0.67	-0.25	0.94		
19	0.98	0.06			0.84	1.19	0.41 -0.41	0.85	1.17	0.38	-0.37		
20	1.13	-0.29			0.97	1.64	0.07 -1.16	1.04	1.17	-0.08	-0.37		
21	0.78	0.60			0.74	0.81	0.69 0.50	0.82	0.74	0.47	0.72		
22	0.74	0.70			0.78	0.66	0.59 0.98	0.74	0.58	0.71	1.27		
23	1.05	-0.11			0.90	1.29	0.24 -0.60	0.97	1.10	0.08	-0.23		
24	0.87	0.32			0.90	0.83	0.26 0.42	0.90	0.85	0.24	0.37		
25	0.96	0.09			0.88	1.08	0.30 -0.19	0.85	1.19	0.37	-0.41		
26	1.48	-0.92			1.68	1.26	-1.22 -0.55	1.53	1.45	-1.00	-0.87		
27	1.35	-0.71			1.34	1.36	-0.69 -0.73	1.21	1.38	-0.45	-0.76		
28	1.01	-0.03			0.86	1.23	0.36 -0.49	0.77	1.20	0.61	-0.44		
29	1.02	-0.06			1.12	0.93	-0.27 0.18	1.00	0.95	-0.00	0.12		
30	0.93	0.17			1.47	0.59	-0.90 1.23	1.19	0.71	-0.40	0.81		
31	1.66	-1.19			2.33	1.38	-1.99 -0.76	2.32	1.34	-1.98	-0.68		
32	1.14	-0.32			1.02	1.39	-0.04 -0.78	1.11	1.17	-0.25	-0.37		
33	1.35	-0.70			1.42	1.31	-0.83 -0.63	1.78	1.09	-1.36	-0.21		
34	1.04	-0.09			0.73	1.21	0.74 -0.44	0.65	1.20	1.03	-0.44		
35	1.21	-0.45			1.02	1.33	-0.05 -0.67	1.24	1.15	-0.50	-0.33		
36	0.90	0.24			0.88	0.91	0.29 0.22	1.01	0.83	-0.01	0.44		
37	1.06	-0.14			1.13	1.00	-0.30 -0.01	1.18	1.03	-0.38	-0.07		
38	0.65	1.02			1.38	<b>0.38*</b>	-0.75 2.26	1.65	<b>0.38*</b>	-1.17	2.26		
39	0.75	0.68			0.64	0.84	1.06 0.40	0.70	0.81	0.85	0.50		
40	1.51	-0.97			0.53	2.61	1.49 -2.25	0.45	2.42	1.88	-2.08		
% DIF	0.00				0.00	2.50			0.00	2.50			
*p<0.005													

**Tabla 5.3.5.**

Funcionamiento diferencial de los items en la comparación  
 Muestra LLF 1 LE 1 vs. Muestra LF 2 LE 2 para la escala H.M.V. del BADYG  
 Nivel Medio.

Items	DIF UNIFORME				DIF NO UNIFORME							
	ETAPA 1		ETAPA 2		ETAPA 1				ETAPA 2			
	$\hat{\alpha}_{MH}$	MH D-DIF	$\hat{\alpha}_{MH}$	MH D-DIF	$\hat{\alpha}_{MH1}$	$\hat{\alpha}_{MH2}$	MH D-DIF	$\hat{\alpha}_{MH1}$	$\hat{\alpha}_{MH2}$	MH D-DIF	$\hat{\alpha}_{MH1}$	$\hat{\alpha}_{MH2}$
1	1.25	-0.53	1.18	-0.39	1.32	1.08	-0.65	-0.18	1.15	1.24	-0.33	-0.50
2	0.64	1.06	0.66	0.97	0.72	0.50	0.77	1.63	0.63	0.61	1.08	1.16
3	1.01	-0.02	1.04	-0.09	1.27	0.48	-0.55	1.71	1.21	0.78	-0.45	0.57
4	0.83	0.44	0.92	0.20	0.80	0.94	0.53	0.16	1.00	0.78	0.01	0.60
5	1.99*	-1.61	1.97*	-1.59	1.55	3.09*	-1.03	-2.65	1.82	2.89*	-1.41	-2.50
6	1.26	-0.54	1.35	-0.70	1.36	1.06	-0.73	-0.13	1.46	1.11	-0.88	-0.25
7	1.00	-0.00	1.03	-0.07	1.07	0.91	-0.16	0.22	1.03	0.90	-0.07	0.24
8	0.86	0.34	0.82	0.45	0.73	1.06	0.75	-0.14	0.75	1.14	0.69	-0.32
9	0.85	0.38	0.88	0.29	0.95	0.71	0.12	0.80	0.96	0.71	0.09	0.80
10	0.49*	1.67	0.53*	1.49	0.47	0.54	1.79	1.45	0.59	0.74	1.24	0.70
11	1.02	-0.04	1.13	-0.29	0.97	1.27	0.06	-0.57	1.01	1.37	-0.03	-0.73
12	0.68	0.90	0.78	0.59	0.67	0.72	0.94	0.79	0.73	0.69	0.73	0.86
13	0.60	1.20	0.50	1.61	0.59	0.63	1.26	1.09	0.48	0.62	1.71	1.11
14	1.18	-0.39	1.19	-0.41	1.40	0.73	-0.79	0.74	1.22	1.17	-0.46	-0.36
15	0.80	0.52	0.86	0.35	0.70	1.06	0.83	-0.13	0.83	0.96	0.45	0.09
16	1.17	-0.36	1.18	-0.39	1.53	0.80	-1.00	0.53	1.48	0.94	-0.92	0.13
17	1.04	-0.10	1.06	-0.15	1.31	0.85	-0.64	0.38	1.19	0.95	-0.40	0.13
18	1.40	-0.79	1.48	-0.93	1.35	1.49	-0.71	-0.93	1.26	1.98	-0.55	-1.60
19	1.67	-1.21	1.85*	-1.44	2.02	1.38	-1.65	-0.76	2.54	1.56	-2.19	-1.05
20	0.64	1.04	0.72	0.76	0.61	0.67	1.16	0.94	0.68	0.67	0.91	0.95
21	0.67	0.94	0.72	0.76	0.70	0.65	0.83	1.02	0.72	0.71	0.79	0.81
22	0.97	0.07	1.04	-0.10	0.94	1.01	0.14	-0.01	1.00	1.00	0.01	-0.01
23	0.74	0.70	0.82	0.47	0.73	0.77	0.75	0.61	0.90	0.67	0.26	0.93
24	1.22	-0.47	1.30	-0.61	1.40	1.14	-0.79	-0.32	1.37	1.31	-0.73	-0.64
25	0.81	0.50	0.84	0.40	0.85	0.76	0.37	0.64	1.00	0.70	-0.01	0.83
26	0.68	0.89	0.71	0.80	0.75	0.65	0.68	1.02	0.61	0.78	1.17	0.59
27	0.87	0.31	0.92	0.18	0.89	0.86	0.28	0.35	0.82	1.05	0.47	-0.11
28	1.72	-1.27	1.84	-1.44	1.81	1.62	-1.39	-1.13	2.40	1.59	-2.05	-1.10
29	1.29	-0.60	1.37	-0.73	1.55	1.04	-1.02	-0.10	1.56	1.23	-1.05	-0.48
30	1.18	-0.38	1.16	-0.35	1.09	1.25	-0.20	-0.53	1.14	1.23	-0.31	-0.50
31	0.88	0.31	1.03	-0.06	0.69	1.03	0.88	-0.07	0.71	1.23	0.81	-0.48
32	1.18	-0.38	1.21	-0.45	1.97	1.02	-1.59	-0.04	2.76	0.94	-2.38	0.14
33	0.89	0.28	1.05	-0.11	1.23	0.71	-0.48	0.81	1.06	1.07	-0.14	-0.15
34	1.27	-0.56	1.31	-0.63	1.12	1.38	-0.26	-0.75	1.30	1.42	-0.61	-0.82
35	0.84	0.42	0.88	0.31	0.85	0.82	0.39	0.47	0.78	1.17	0.57	-0.36
36	0.91	0.22	0.97	0.08	0.38	1.56	2.29	-1.04	0.55	1.67	1.41	-1.20
37	0.96	0.10	1.02	-0.06	0.73	1.14	0.75	-0.31	0.92	1.12	0.19	-0.26
38	2.19**	-1.84	2.41*	-2.07	1.52	2.71*	-0.99	-2.35	1.83	2.48*	-1.42	-2.13
39	1.24	-0.51	1.37	-0.74	1.35	1.20	-0.71	-0.43	1.31	1.39	-0.63	-0.77
40	1.21	-0.45	1.33	-0.67	1.63	0.96	-1.15	0.09	1.66	1.16	-1.20	-0.34
% DIF	77.50		10.00		0.00	5.00			0.00	5.00		

\*p<0.005

Tabla 5.3.6.

Funcionamiento diferencial de los items en la comparación  
Muestra LF 1 LE 1 vs. Muestra LF 2 LE 2 para la escala H.M.NV. del BADYG  
Nivel Medio.

Items	DIF UNIFORME				DIF NO UNIFORME					
	ETAPA 1		ETAPA 2		ETAPA 1			ETAPA 2		
	$\hat{\alpha}_{MH}$	MH D-DIF	$\hat{\alpha}_{MH}$	MH D-DIF	$\hat{\alpha}_{MH1}$	$\hat{\alpha}_{MH2}$	MH D-DIF	$\hat{\alpha}_{MH1}$	$\hat{\alpha}_{MH2}$	MH D-DIF
1	1.26	-0.55			1.26	1.27	-0.54	-0.56		
2	1.11	-0.25			0.98	1.58	0.05	-1.08		
3	0.92	0.19			0.86	1.06	0.35	-0.13		
4	1.08	-0.18			1.16	0.90	-0.35	0.25		
5	1.59	-1.08			1.43	1.75	-0.83	-1.31		
6	0.96	0.09			0.83	1.06	0.43	-0.15		
7	0.99	0.01			0.97	1.02	0.06	-0.06		
8	0.81	0.48			0.73	1.02	0.73	-0.04		
9	1.35	-0.70			1.42	1.29	-0.82	-0.60		
10	0.89	0.28			1.07	0.56	-0.17	1.36		
11	0.86	0.34			0.69	1.18	0.89	-0.40		
12	1.20	-0.43			1.34	1.03	-0.68	-0.08		
13	0.91	0.22			1.14	0.69	-0.31	0.86		
14	0.92	0.19			0.71	1.24	0.80	-0.50		
15	0.81	0.49			0.57	1.34	1.32	-0.70		
16	1.14	-0.30			1.03	1.26	-0.06	-0.54		
17	0.74	0.70			0.67	0.82	0.93	0.48		
18	0.74	0.71			0.57	0.99	1.33	0.02		
19	0.88	0.30			0.98	0.69	0.04	0.87		
20	1.02	-0.04			1.16	0.95	-0.36	0.13		
21	0.96	0.11			0.89	1.07	0.26	-0.16		
22	1.00	0.01			1.77	0.68	-1.34	0.91		
23	1.56	-1.04			2.25	1.25	-1.90	-0.52		
24	0.87	0.33			0.81	0.93	0.51	0.16		
25	0.87	0.32			0.78	0.97	0.57	0.07		
26	1.27	-0.56			1.07	1.47	-0.16	-0.90		
27	1.01	-0.03			0.92	1.15	0.20	-0.33		
28	0.98	0.05			1.43	0.81	-0.84	0.49		
29	0.63	1.07			0.65	0.63	1.02	1.10		
30	0.98	0.04			0.62	1.48	1.11	-0.93		
31	0.87	0.32			0.81	0.93	0.49	0.18		
32	1.23	-0.49			1.30	1.18	-0.62	-0.38		
33	1.18	-0.40			0.84	1.81	0.40	-1.40		
34	0.65	1.00			1.23	0.46	-0.48	1.84		
35	1.12	-0.26			1.62	0.75	-1.13	0.69		
36	0.80	0.53			0.80	0.79	0.51	0.55		
37	0.93	0.18			1.52	0.51	-0.99	1.58		
38	1.62	-1.14			1.83	1.47	-1.43	-0.91		
39	0.85	0.37			1.15	0.68	-0.33	0.90		
40	1.02	-0.05			1.20	0.92	-0.43	0.20		
% DIF	0.00				0.00	0.00				
*p<0.005										

Para una mejor descripción de qué items son detectados con DIF uniforme y DIF no uniforme, tanto sin (etapa 1) como con (etapa 2) purificación del constructo evaluado en ambos niveles de significación se construyeron las tablas 5.3.7. y 5.3.8. En ellas, se puede apreciar con más claridad qué los items son los que presentan problemas de funcionamiento diferencial.

**Tabla 5.3.7.**

Resumen de los análisis de DIF en las escalas H.M.V. y H.M.NV. en la comparación Muestra LF C LE C vs. Muestra LF V LE CV

	BADYG ELEMENTAL		BADYG MEDIO	
	HMV	HMNV	HMV	HMNV
ALFA=0.005				
DIF UNIFORME				
POSITIVO			I10	0
NEGATIVO			I5, I19, I38	0
% ITEMS DIF	0 %	0 %	10 %	0 %
DIF NO UNIFORME				
POSITIVO				0
NEGATIVO			I5, I38	0
% ITEMS DIF	0 %	0 %	12.5 %	0 %

**Tabla 5.3.8.**

Resumen de los análisis de DIF en las escalas H.M.V. y H.M.NV. en la comparación Muestra LF C LE C vs. Muestra LF V LE CV

	BADYG ELEMENTAL	
	HMV	HMNV
ALFA=0.005		
DIF UNIFORME		
POSITIVO	I10	0
NEGATIVO	I22, I26	0
% ITEMS DIF	7.5 %	0 %
DIF NO UNIFORME		
POSITIVO	0	I38
NEGATIVO		
% ITEMS DIF	0 %	2.5 %



En términos generales, podemos apreciar que la aplicación de la prueba H.M.V. formada por items de contenido verbal presenta problemas de funcionamiento diferencial en sus items, tal y como se había hipotetizado. Por lo que se refiere a la escala H.M.NV. compuesta por items de tipo gráfico, los resultados obtenidos sugieren que estos items no suponen ningún esfuerzo adicional para los diferentes grupos de sujetos lingüísticos comparados. En cuanto al tipo de DIF detectado, se puede observar una cierta predominancia del DIF uniforme. Por último, respecto a la utilización de un método de purificación del constructo analizado, es conveniente señalar que éste no ha repercutido en un descenso significativo de los items detectados con DIF. Quizás la justificación a este hecho se encuentre en el pequeño número de items con problemas de funcionamiento diferencial.

***CAPITULO 6***

**Conclusiones**



En la actualidad, los estudios sobre DIF se constituyen como una iniciativa metodológica para evaluar la objetividad en la evaluación de un constructo o rasgo psicológico con un determinado instrumento de medida. La importancia de este tema encuentra su apoyo en los últimos avances alcanzados por una gran cantidad de trabajos y estudios metodológicos sobre los métodos de detección de DIF, así como en las implicaciones sociales derivadas de su práctica aplicada a diversos contenidos, aunque esta última con menor frecuencia. Nuestro interés se sitúa en el segundo ámbito.

En sus comienzos, la investigación empírica muestra que las variables de DIF más estudiadas y analizadas fueron las referidas a la raza y el sexo. Sin embargo, en los últimos años se ha constatado un creciente interés por el estudio de variables de tipo educativo, lingüístico, cultural, etc.. constituyendo un nuevo mundo de posibilidades sin explorar hasta entonces. De hecho, las sociedades multilingües o situaciones de naturaleza bilingüe se presentan como uno de los escenarios más habituales de estudio del DIF. Precisamente, dentro en estos ámbitos se desarrolla la línea de investigación iniciada por Hulin y cols (Hulin, Drasgow y Komocar, 1982; Drasgow y Hulin, 1986; Hulin y Mayer, 1986; Candell y Hulin, 1987) y, posteriormente, adoptada por Ellis (1989, 1995; Ellis y Kimmel, 1992; Ellis, Becker y Kimmel, 1993). Su principal objetivo reside en el estudio del DIF en tests o instrumentos de medida aplicados a muestras monolingües y/o bilingües con el fin de descubrir en qué medida las características lingüísticas de los sujetos tienen un efecto no deseado en la medición y evaluación de sus capacidades cognitivas y psicológicas. Además, ésta deriva en la aparición de nuevos trabajos de investigación sobre el estudio de la equivalencia de la medición de los tests traducidos con respecto a la versión original, y el grado en que los items con DIF entre las distintas versiones de un mismo test constituyen una fuente de información sobre posibles diferencias culturales.

Una realidad empírica con las características propias de una situación de contacto de lenguas podemos encontrarla en la Comunidad Valenciana, una comunidad bilingüe donde el uso y aprendizaje de sus dos lenguas oficiales es un hecho bastante común. Partiendo de esta

circunstancia, podemos distinguir tres situaciones diferentes en el sistema educativo valenciano:

1) aquellas zonas donde el valenciano está presente con fuerza y los sujetos residentes en ellas no sólo lo entienden y lo hablan cotidianamente, sino que también cursan programas bilingües de enseñanza de la lengua minoritaria. En concreto, estos sujetos serán los destinatarios de un bilingüismo aditivo al finalizar su educación obligatoria, aunque a lo largo de su proceso de escolarización éstos puedan presentar algunas deficiencias lingüísticas en la competencia lingüística de la lengua mayoritaria;

2) aquellas zonas con un predominio claro de la lengua valenciana y donde los sujetos utilizan el valenciano como lengua habitual de comunicación en sus relaciones informales, pero en cambio son escolarizados en programas de inmersión progresiva al valenciano. En estos casos, la formación académica de dichos sujetos sufre una especie de choque cognitivo en los primeros años de escolarización, debido a la exigencia de un cambio de lengua en su incorporación al sistema educativo, y por contra

3) los alumnos pertenecientes a las zonas lingüísticas donde el valenciano tiene una presencia muy débil, y el modelo educativo implantado son programas de enseñanza monolingüe en castellano, siendo la adquisición de la lengua minoritaria muy limitada.

En todas estas situaciones existe una generalización en la evaluación de las aptitudes intelectuales de los alumnos valencianos, basada

fundamentalmente en la aplicación colectiva de pruebas redactadas en lengua castellana, las más predominantes en el mercado profesional.

Ante esta situación, el presente trabajo de investigación pretende averiguar en qué medida las características lingüísticas de los grupos definidos previamente afectan a los resultados obtenidos en rendimiento tras la administración de tests o instrumentos de medida en lengua castellana a lo largo de parte del proceso de escolarización. Para dar una respuesta a dicha problemática, en el presente estudio se planteó realizar un análisis de detección de DIF en dos pruebas aptitudinales elaboradas en lengua castellana y de uso habitual en el ámbito educativo valenciano, con el objetivo principal de averiguar si los items de las mencionadas pruebas favorecen sistemáticamente a los miembros de un grupo sobre otro, igualmente capaces, pero con características lingüísticas diferentes, tales como poseer una lengua familiar y/o lengua de escolarización diferente, como hemos visto anteriormente. La novedad de este procedimiento frente a los métodos estadísticos habitualmente utilizados en los estudios sobre bilingüismo consiste en la igualación de los diferentes grupos a comparar en el rasgo o constructo medido por el test, lo que sin duda aporta una garantía de no confundir los términos de IMPACTO con DIF.

Para conseguir nuestro propósito, las siguientes dos pruebas de la batería de inteligencia BADYG Elemental y Medio fueron objeto de estudio: la escala H.M.V. compuesta por 40 items con fuerte contenido verbal, y la escala H.M.NV. formada por otros 40 items pero siendo su contenido meramente gráfico. Estas pruebas se aplicaron a los tres grupos lingüísticos definidos previamente. Las citadas escalas fueron sometidas a

un análisis de componentes principales para evaluar su dimensionalidad, ya que uno de los requisitos a cumplir para obtener un análisis adecuado del DIF es la unidimensionalidad del rasgo o constructo evaluado. En términos generales, los resultados obtenidos indicaron la predominancia de un primer componente principal en dichas escalas, aunque atendiendo a los niveles de unidimensionalidad para cada escala y grupo, éstos denotaron con bastante claridad una mayor unidimensionalidad de la escala H.M.NV. que de la escala H.M.V.

Respecto a los análisis correspondientes a la detección del DIF en los ítems de ambas escalas tras purificar la medida de la habilidad, los resultados aportados mostraron, tal como se esperaba, un mayor predominio del funcionamiento diferencial en los ítems verbales de la escala H.M.V. que en los ítems integrantes de la escala H.M.NV. También es conveniente destacar que el método de purificación utilizado, una versión bietápica del estadístico MH, no repercutió de forma substancial en la cantidad y número de ítems detectados con DIF. Al respecto, algunos estudios realizados muestran que dicho método de purificación no resulta muy útil cuando la presencia de ítems con problemas de DIF en magnitud y número es mínima (Linn y cols, 1981), aspecto característico de los estudios sobre DIF en las variables sexo, raza, etnia, etc.

Seguidamente, se comentarán los resultados derivados de la comprobación de las hipótesis planteadas en este estudio. Conviene recordar que la primera de ellas suponía que la presencia de ítems con problemas de funcionamiento diferencial en una prueba estaba relacionada con las características lingüísticas de los sujetos sometidos a evaluación. Por



tanto ésta proponía que la prueba de H.M.V. compuesta por items verbales presentaría un mayor porcentaje de items con funcionamiento diferencial que la prueba de H.M.NV. sin referentes verbales en su composición. Los resultados obtenidos ofrecen apoyo empírico a esta primera hipótesis. Al respecto, es interesante destacar que este hallazgo se encuentra en sintonía con un estudio realizado en el contexto educativo holandés sobre la construcción y adaptación de una prueba de inteligencia sin referentes de tipo verbal (Resing, Bleichrodt y Drenth, 1986; Willemse, 1989; Van de Rijt, 1990). Los resultados aportados por este estudio tras aplicar una escala de inteligencia sin referentes verbales a alumnos nativos holandeses y emigrantes van en el sentido de la hipótesis formulada.

No obstante, a pesar de una clara tendencia de ausencia de DIF en la escala de H.M.NV., ésta presentó un ítem con problemas de funcionamiento diferencial no uniforme, en concreto el ítem 38, en la comparación *muestra LF C LE C vs. muestra LF C LE V*. Hay que recordar que una limitación en el cálculo del DIF no uniforme con el estadístico MH es un aumento de la tasa de error tipo 1, por tanto pensamos que es razonable creer que la presencia de DIF en dicho ítem es debida a ciertos problemas de sobreestimación en la detección del DIF no uniforme. Otra posible justificación a este resultado podría residir en la posición del ítem en la escala. De hecho, algunos estudios realizados (Dorans, Schmitt y Bleistein, 1992) señalan que un factor muy importante en la aparición de items con problemas de funcionamiento diferencial puede ser su localización al final de la escala, y en concreto, éste es el caso que nos ocupa. El único ítem detectado con DIF en la escala H.M.NV. es el ítem 38, lo que sugiere que la aparición de DIF en dicho ítem puede ser producto de su localización en el test. Dado

que la prueba administrada es un test de inteligencia con tiempo límite en su realización, el hecho de que algunos sujetos alcancen el ítem y otros no puede ser la causa de la aparición del DIF.

Las siguientes hipótesis planteadas en este estudio se formularon con el fin de averiguar si las características lingüísticas de los diferentes grupos de sujetos sometidos a comparación tenían algún efecto no deseado en la medición de sus aptitudes con instrumentos de medida elaborados y baremados en lengua castellana, y aplicados a lo largo de su proceso de escolarización. Así pues, atendiendo a este objetivo se formularon las siguientes dos hipótesis. La *primera* de ellas proponía que los alumnos pertenecientes al grupo focal, compuesto por sujetos cuya lengua habitual de comunicación es el valenciano (un idioma distinto a la lengua del test administrado) y que están escolarizados en programas bilingües de incorporación progresiva al valenciano, probablemente resultarían perjudicados en la prueba verbal de inteligencia (H.M.V.) del BADYG elemental, respecto a la ejecución presentada por los sujetos del grupo de referencia, cuya lengua familiar y escolar es el castellano. Los resultados obtenidos al respecto no confirmaron la hipótesis. Los análisis de detección de DIF realizados en la escala H.M.V. del BADYG elemental en la comparación *muestra LF C LE C vs. muestra LF V LE CV* no identificaron ningún ítem con funcionamiento diferencial.

Por lo que respecta a la *segunda* hipótesis, ésta señalaba que los escolares con una lengua familiar y de escolarización distinta a la versión de la prueba administrada, como son los escolares valencianoparlantes escolarizados en programas de enseñanza en valenciano, presentarían una

mayor cantidad de dificultades o handicaps lingüísticos al ser evaluados durante su escolarización con pruebas de tipo verbal en dicho idioma, que los sujetos cuya lengua familiar y de escolarización es idéntica a la de la prueba administrada, como es el caso de los alumnos monolingües castellanos. Es decir, en la comparación entre las muestras LF C LE C y LF V LE V se esperaba que aparecieran items con DIF que perjudicaran el rendimiento de los sujetos del grupo focal (LF V y LE V). Los resultados procedentes de los análisis de DIF (ver comparación *muestra LF C LE C vs. muestra LF V LE V*) corroboraron dicha hipótesis. Por tanto, los resultados aportados confirmaron que la aplicación de una prueba de inteligencia verbal en lengua castellana a escolares de Educación Primaria con lengua familiar y de escolarización distinta a ésta conlleva una cierta discriminación en los resultados de rendimiento obtenidos en el test por dicho grupo.

En concreto, la escala H.M.V. del BADYG elemental en versión castellana aplicada a sujetos con características lingüísticas diferentes en lengua familiar y de escolarización presentó los siguientes problemas:

(a) En relación a los items detectados con problemas de funcionamiento diferencial (DIF), un total de 3 items fueron detectados con DIF uniforme para la escala H.M.V. En cuanto al signo del índice de DIF calculado, conviene destacar que dos de los items mostraron un funcionamiento diferencial en contra del grupo focal, y el ítem 10 a favor. Los enunciados correspondientes a los citados items son respectivamente:

*Item 10:* La MESA siempre debe tener: a) Madera; b) Algo que sirva para apoyar en el suelo; c) Cajones; d) Hierros; e) Libros encima, y f) Adornos.

*Item 22:* ¿Qué palabra sobra?: FANTA- VINO- ANIS- LICOR- BRANDY- GINEBRA a) Vino; b) Ginebra; c) Anís; d) Licor; e) Brandy, y f) Fanta.

*Item 26:* ¿Qué número sigue? 12-14-28-30-60.. a) 8; b) 5; c) 2; d) 4; e) 15, y f) 6.

Con referencia al ítem 22, puede resultar interesante señalar que algunos de los sujetos encuestados del grupo focal manifestaron algunas dudas acerca del significado de la palabra BRANDY presente en la formulación dicho ítem.

(b) la falta de variabilidad en las respuestas dadas por los sujetos del grupo focal al ítem 11 de la escala H.M.V., siendo incorrectamente respondido por la totalidad de sus componentes. El enunciado correspondiente a dicho ítem es el siguiente: *Item 11:* ¿Qué palabra sobra?: CEREZO- GUINDO- PERAL- MATORRAL- NARANJO- MANZANO. Al respecto, quizás resulte interesante destacar que durante las aplicaciones de esta prueba a dicho grupo, una de las preguntas más comunes fue solicitar a los aplicadores del test una explicación acerca del significado de la palabra "guindo", tal vez el desconocimiento de este término provocó una falsa identificación en la respuesta correcta del ítem en cuestión y, por tanto, ningún sujeto de la presente muestra fue capaz de resolver con éxito la pregunta en cuestión;

(c) Otro ítem que presentó problemas fue el ítem 7, aunque sólo para el cálculo del DIF no uniforme en el grupo superior debido de su falta de variabilidad en dicho grupo.

En conjunto, podemos afirmar que la aplicación de la escala H.M.V. en versión castellana del BADYG elemental a grupos de alumnos diferentes en lengua familiar y escolar (grupo de referencia: muestra LF C LE C, y grupo focal: LF V LE V) presentó problemas tanto de invariabilidad en sus respuestas como de funcionamiento diferencial en 3 ítems de un total de 40, lo que equivale al 7.5 % del test analizado.

La última hipótesis formulada se basó en la posibilidad de que la presencia de ítems con funcionamiento diferencial en las escalas aplicadas estuviera determinada por los años de escolarización cursados por el alumno. Para someter a prueba dicha hipótesis, ésta se planteó en base al siguiente razonamiento: puesto que los alumnos que se encuentran en proceso de escolarización en programas de enseñanza bilingüe están menos expuestos a la presencia y aprendizaje de la lengua mayoritaria (el castellano) es posible que sus habilidades lectoescritoras en dicha lengua no estén suficientemente desarrolladas y consolidadas hasta llegar a los últimos cursos de la educación obligatoria. Luego, los alumnos escolarizados en programas de enseñanza en valenciano y cuya lengua es el valenciano, y que cursan la Educación Primaria se encontrarán con mayores dificultades o handicaps lingüísticos al ser evaluados mediante pruebas verbales en castellano, que los sujetos del mismo grupo que cursan la E.S.O. Es decir, se esperaba que los análisis de detección de DIF ofrecieran un mayor porcentaje de ítems con funcionamiento diferencial para la prueba BADYG

Elemental que para la prueba BADYG Medio cuando se comparan los grupos LF C más LE C (grupo de referencia) vs. LF V más LE V ( grupo focal). Antes de pasar a comentar los resultados obtenidos, cabe recordar que esta comparación sólo se realizó para alumnos castellanoparlantes escolarizados en castellano (muestra LF C y LE C) y escolares valencianoparlantes escolarizados en programas de incorporación progresiva al valenciano (muestra LF V y LE CV) debido a las razones señaladas anteriormente (ver apartado referente al tipo de muestra empleada).

Los resultados obtenidos en relación a esta hipótesis no sólo no la confirmaron sino que la tasa de identificación de items con DIF para la escala H.M.V. del BADYG medio fue superior a la obtenida por la escala H.M.V. del BADYG elemental. La comparación de las escalas H.M.V. en ambos niveles de aplicación mostraron la no existencia de funcionamiento diferencial para el nivel BADYG elemental y un total de 4 items con problemas de funcionamiento diferencial para el BADYG medio, donde los items 10 y 19 presentaron DIF uniforme y los items 5 y 38 DIF no uniforme. Respecto al signo del índice calculado, conviene destacar que el ítem 19 mostró un funcionamiento diferencial en contra del grupo focal y el ítem restante, el 10 a favor de éste. El enunciado de dichos items se detalla a continuación:

*Item 5: ¿Qué palabra sobra en la serie siguiente?: AVISPA -ARAÑA - MOSCA - LAGARTO -GUSANO -ABEJA a) Gusano; b) Araña; c) Avispa; d) Mosca; e) Lagarto, y f) Abeja*

*Item 10:* ¿Qué letra o grupo de letras continúa en la serie siguiente?: ab- b- cd- d- ef... a) g; b) h; c) e; d) f; e) fg, y f) Otra no indicada

*Item 19:* ¿ Qué número dividido entre 5 nos dará lo mismo que la mitad de 8 por 2? a) 20; b) 40; c) 30; d) 5; e) 4; y f) 10

*Items 28:* ¿Cuánto es un tercio del cuádruple de 30?: a) 60; b) 30; c) 40; d) 9; e) 80, y f) 100

*Item 38:* Ordena estas palabras sueltas en una frase correcta y contéstala: VIVIENDA- LOS- LLAMAN- FIJA- SE- PUEBLOS- QUE- TIENEN: a) Extranjeros; b) Nómadas, c) Indígenas; d) Habitantes; e) Fijos, y f) Sedentarios

Una posible justificación a la obtención de este resultado tal vez podría encontrarse en el grado de unidimensionalidad presentado por dicha escala. Un examen de los análisis de unidimensionalidad obtenidos permite observar que la escala H.M.V. del BADYG medio es la escala que peor resultados ofrece en cuanto a unidimensionalidad del constructo se refiere, explicando el primer componente extraído únicamente el 13.8 % de la varianza total junto con un valor propio de 5.53. Así pues, estos resultados quizás sean la causa principal de la aparición de tantos ítems con problemas de funcionamiento diferencial. Además, no hay que olvidar la perspectiva del DIF a consecuencia de la violación de la unidimensionalidad, según la cual la presencia de DIF en el test tiene lugar cuando, bajo ciertas condiciones, se incumple el supuesto de unidimensionalidad en el test. De hecho, en la actualidad la multidimensionalidad aparece como una de las causas más aceptadas de la

aparición del funcionamiento diferencial de los items en un determinado test.

Si comparamos nuestros resultados con los obtenidos en otros estudios sobre detección de DIF en muestras bilingües, se puede apreciar que el porcentaje de items detectados con DIF en estas situaciones no es del todo concluyente, siempre bajo la condición de que la misma versión del test sea aplicada a una muestra monolingüe y bilingüe. Precisamente, los estudios de Candell y Hulin (1987) mostraron que la aplicación de una escala verbal en versión inglesa a dos grupos lingüísticos diferentes pero procedentes del mismo país (muestra canadiense francófona: m. bilingüe, y muestra canadiense anglófona: m. monolingüe) no derivó en la aparición de DIF en la prueba aplicada. Sin embargo, cuando ésta misma se administró a grupos lingüísticos similares pero culturalmente diferentes (muestra canadiense francófona: m. bilingüe, y muestra americana: m. monolingüe) se detectó DIF en el 15% de los items de la escala. Otro estudio en la línea de investigación propuesta por Hulin y cols (Hulin, Drasgow y Komocar, 1982; Drasgow y Hulin, 1986) detectó 3 items con DIF de un total de 70 items, tras ser aplicada la versión utilizada a una muestra monolingüe (hispana ó americana) y bilingüe (hispano-americanos).

En definitiva, los resultados obtenidos en el presente estudio contribuyen a demostrar que la aplicación de la prueba H.M.V. de inteligencia verbal a alumnos con características lingüísticas diferentes para la que ésta originariamente fue creada repercute directamente en la aparición de DIF en los items del test. Por su parte, la escala gráfica analizada, la escala H.M.NV. no presentó problemas de funcionamiento



diferencial en sus items, a excepción de un caso ya comentado. Estos hallazgos permiten concluir que la utilización de la citada escala de inteligencia verbal para la evaluación de las aptitudes intelectuales en los diferentes grupos de escolares valencianos analizados aquí no es del todo adecuada. La ocurrencia de problemas de DIF en dicha prueba suponen un primer obstáculo para llevar a cabo estudios comparativos adecuados sobre rendimiento o inteligencia en base a una única versión del test. Es cierto que la tasa de items detectados con DIF en la escala verbal no ha resultado ser muy elevada, sin embargo este hecho no invalida en absoluto los comentarios señalados. No hay que olvidar que todas las comparaciones realizadas en este estudio siempre se han establecido entre alumnos monolingües y bilingües lo cual deriva, tal y como señalan algunos estudios realizados (Hulin, Drasgow y Komocar, 1982; Drasgow y Hulin, 1986), en una menor presencia de items con DIF. La principal causa aportada para explicar estos resultados es la naturaleza cognitiva de los sujetos bilingües. Como se ha expuesto en los capítulos anteriores, los sujetos bilingües que manejan dos códigos lingüísticos diferentes poseen un desarrollo diferente en las habilidades cognitivas no lingüísticas frente a los monolingües, lo cual provoca que las diferencias encontradas en las comparaciones entre monolingües y bilingües en la detección del DIF sean menores que las diferencias halladas entre monolingües con lenguas diferentes, aspecto que se reflejará en un descenso de la cantidad de items identificados con problemas de DIF (Hulin, Drasgow y Komocar, 1982; Drasgow y Hulin, 1986).

Resumiendo, en base a los resultados obtenidos, los estudios sobre DIF son necesarios en el ámbito de la medición de aptitudes en contextos bilingües ya que la presencia de diferencias lingüísticas entre los

diferentes sujetos pueden determinar la aparición de diferentes probabilidades de responder correctamente a un ítem determinado cuando en realidad éstos poseen el mismo nivel de habilidad en el constructo medido. La realización de este tipo de estudios contribuirá a garantizar que los ítems funcionen de forma equivalente, y por tanto que las comparaciones entre diferentes grupos que han contestado a una misma versión de la prueba tengan sentido. Atendiendo a los resultados conseguidos en este trabajo, la solución al problema del DIF en el contexto bilingüe analizado podría residir en las siguientes pautas de actuación:

1) la eliminación de aquellos ítems que presentan problemas de funcionamiento diferencial de las pruebas en castellano.

2) el empleo de pruebas de inteligencia sin contenido verbal dado que su funcionamiento adecuado para la evaluación de las aptitudes intelectuales en los alumnos bilingües analizados en este estudio queda totalmente demostrado,

3) la traducción y posterior validación de tests o pruebas verbales de inteligencia en la lengua minoritaria, siempre y cuando se demuestre que las versiones construidas son métricamente equivalentes a la versión original en castellano,

4) el desarrollo de nuevas pruebas en la lengua minoritaria

Por otra parte, también cabe destacar que el presente trabajo de investigación presenta algunas limitaciones. Estas son las siguientes:

La primera de ellas reside en el hecho de que se trata de un estudio empírico y, por tanto, es prácticamente imposible conocer la relación exacta existente entre el número de detecciones correctas y los errores Tipo I (falsos positivos (FP)) y tipo II (falsos negativos (FN)) implicados en los análisis realizados. En este sentido, nosotros sugerimos que en futuras investigaciones empíricas se debería examinar la consistencia del método estadístico utilizado en la detección del DIF mediante estudios de replicación con diferentes muestras (Budgell, Raju y Quartetti, 1995).

La segunda limitación se refiere al tamaño muestral utilizado. Es importante recordar que una de las limitaciones de los estudios empíricos de DIF se refiere al tamaño muestral empleado, dado que en la práctica este tipo de estudios disponen de pocos sujetos en el grupo focal y, en ocasiones, el grupo de referencia tampoco es suficientemente numeroso. En nuestro estudio, el tamaño muestral utilizado ha sido el requerido para poder aplicar el método estadístico MH. Sin embargo, en próximos estudios resultaría conveniente ampliar el tamaño muestral de los diferentes grupos comparados para poder aplicar diferentes métodos estadísticos de detección de DIF y, por tanto, examinar la congruencia entre los diferentes métodos aplicados. En la práctica, pocos estudios empíricos (Budgell y cols, 1995) han analizado la convergencia entre dichas técnicas. Parece evidente que si bien el desarrollo metodológico del tema ha sido extenso y riguroso, las aplicaciones prácticas de este tipo de estudios están prácticamente en sus comienzos. Ante nosotros se abre un futuro prometedor sobre el estudio del sesgo o DIF en los instrumentos de medida aplicados en contextos bilingües, aspecto que como se ha comentado varias veces anteriormente es patente en el sistema educativo desarrollado en el Estado Español.

Una vez hemos llegado aquí, somos conscientes de que la línea de investigación iniciada en esta tesis no acaba aquí. Una vez se ha detectado la presencia de funcionamiento diferencial en los ítems de un test mediante la aplicación de cualquier método estadístico el siguiente paso consistirá en averiguar las causas de la presencia de ese DIF y encontrar una explicación teórica de la ocurrencia del mismo. En los estudios empíricos sobre DIF no sólo es de gran interés detectar y eliminar del test aquellos ítems que muestran DIF con respecto a ciertos grupos de la población, sino conocer el porqué del funcionamiento diferencial de los ítems. Así pues, una vez conocidos las causas determinantes de la presencia del DIF puede ser útil para los constructores de tests prevenir con mayor facilidad la presencia de dichos ítems y, en consecuencia, la aparición de DIF en los mismos. En la actualidad, los avances en intentar ofrecer una respuesta a dicha problemática se sitúan en dos direcciones diferentes. La primera de ellas sugiere que los resultados obtenidos en la detección del DIF pueden contribuir a un mayor conocimiento de los procesos psicológicos que subyacen a la hora de responder los diferentes ítems de un determinado test (Hambleton, 1993; 1994; Hambleton, Clauser, Mazor y Jones, 1993; Hambleton y Kanjee, 1995; Hambleton y Jones, 1994; Poortinga, 1995). La identificación de tales factores podría contribuir de forma significativa al desarrollo de instrumentos de evaluación válidos dentro del campo de la Psicología comparativa y transcultural. Por ejemplo, un análisis post hoc de los ítems que presentan DIF podría proporcionar información adicional sobre las posibles causas de la aparición del DIF (Scheuneman, 1987; Skaggs y Lissitz, 1992; Wang y Lane, 1996). Precisamente, los ítems con DIF pueden ser analizados con el fin de identificar la naturaleza de diferencias en procesos de percepción, cognición y contenido en diferentes sectores

grupales u otros factores relacionados con el DIF. De hecho, algunos estudios señalan que la presencia de items con *palabras homógrafas, sinónimos y analogías* tiende a desfavorecer a la mayoría de los grupos minoritarios evaluados. En esta línea queremos destacar que la batería empleada incluye, además de las pruebas analizadas en el presente estudio, una prueba de comprensión verbal compuesta por 30 items referentes a sinónimos, antónimos y analogías verbales, aspectos que en futuras investigaciones pueden ayudarnos a detectar qué características de los items tienden a favorecer a uno u otro grupo, tal y como se ha hecho en algunos estudios realizados (Schmitt y Dorans, 1990; Dorans, Schmitt y Bleistein, 1992).

La segunda aproximación al estudio de las causas del DFIF se basa en el análisis de la multidimensionalidad de la medida (Ackerman, 1992; Oort, 1996; Shealy y Stout, 1993a, 1993b) y por tanto en la consiguiente identificación de qué otras variables mide el test, aparte de la variable principal. Dichas variables son de naturaleza espúrea y posiblemente afectan al nivel conseguido por los sujetos en la variable de interés. Hoy en día, la línea desarrollada con modelos multidimensionales es una de las alternativas que mayor atención está recibiendo y parece ser la opción más prometedora en la posible explicación de las causas del DIF.

Otro línea de investigación que se encuentra en plena expansión es la extensión de los métodos estadísticos desarrollados para items de respuesta dicotómica a situaciones de evaluación psicológica con otros tipos de respuesta, como es el caso de la evaluación de la inteligencia con escalas de respuesta politómica. Así pues, ante estos nuevos avances metodológicos

la presente investigación podría continuar en el estudio empírico del DIF mediante la aplicación de métodos de detección de DIF en ítems politómicos.

Por último y para finalizar, señalaremos que el presente trabajo realizado sugiere nuevos caminos en el campo de la adaptación psicológica referida a la elaboración de instrumentos de medición. La Comunidad Valenciana es una sociedad bilingüe donde el uso y del aprendizaje de la propia lengua es un hecho bastante común, convirtiéndose tanto en objeto de enseñanza como en lengua vehicular en el sistema educativo valenciano. Tal circunstancia hace necesario un instrumento de medida que permita medir el aprovechamiento en los escolares valencianos, por tanto la adaptación y validación de pruebas aptitudinales al valenciano, como la elaboración de pruebas en valenciano, y no menos importante el estudio del DIF entre la versión traducida con respecto a la versión original abre nuevas líneas de investigación por las que poder transitar en estudios posteriores y esperamos que en un futuro muy cercano.



## **Referencias bibliográficas**





- Ackerman, T.A. (1992). A didactic explanation of item bias, item impact, and item validity from a multidimensional perspective. *Journal of Educational Measurement*, 29, 1, 67-91.
- Ackerman, T.A. (1994). Using multidimensional item response theory to understand what items and tests are measuring. *Applied Measurement in Education*, 7, 4, 255-278.
- Ackerman, T.A. y Evans, J.A. (1994). The influence of conditioning scores in performing DIF analyses. *Applied Psychological Measurement*, 18, 4, 329-342.
- Angoff, W.H. (1982). Use of Difficulty and Discrimination indices for detecting Item Bias. In R.A. Berk (Ed.), *Handbook of Methods for Detecting Test Bias*. Baltimore, MD: The Johns Hopkins University Press.
- Angoff, W.H. (1993). Perspectives on differential item functioning. In P.W. Holland y H. Wainer (Eds.), *Differential item functioning* (pp.3-23). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
- Angoff, W.H. y Ford, S.F. (1973). Item-Race Interaction on a Test of Scholastic Aptitude. *Journal of Educational Measurement*, 10, 95-106.
- Arenas, J. (1986). *La immersió lingüística*. Escrits de divulgació. Barcelona: La Llar del Llibre.
- Arenas, J. y Abeyà, M. (1984). *El Programa d'Immersion*. Departament d'Ensenyament de la Generalitat de Catalunya.
- Arnau, J. (1985). Educación en la segunda lengua y rendimiento escolar: una revisión de la problemática general. En M. Siguán (Ed.), *Enseñanza en dos lenguas y resultados escolares*. Barcelona: Publicacions i Edicions de la Universitat de Barcelona.
- Arnau, J.; Comet, C.; Serra, J.M. y Vila, I. (1992). *La Educación Bilingüe*, ICE Universitat de Barcelona, Barcelona, Ed. Horsori.
- Arnberg, L. (1987). *Raising children biligually: the pre-school years*. Clevedon, Multilingual Matters.
- Artigal, J.M. (1991). Uso-adquisición de una lengua extranjera en el marco escolar entre los tres y los seis años. *Lenguaje y Comunicación*, 7-8, 127-144.
- Atucha, K. (1978). Problemática psicopedagógica del bilingüismo en el País Vasco. En Varios (1978): *Bilingüismo y biculturalismo*. Barcelona: CEAC.
- Bain, B. y Yu, A. (1978). Towards an integration of Piaget and Vigotsky: a cross-cultural replication (France, Germany, Canada) concerning cognitive consequences of bilinguality. En Michael Paradis (Ed.), *Aspects of bilingualism*, Columbia, Hornbeam Press.
- Bain, B.C. (1975). Toward an integration of Piaget and Vigotsky: Bilingual considerations. *Linguistics*, 160, 5-20.
- Baker, C. (1993). Bilingual education in Wales. En H. Baetens Beardsmore (Ed.), *European Typologies of Bilingual Education*. Clevedon. Multilingual Matters, 1993.
- Baker, C. y Al-Karni, A (1991). A comparison of two procedures for computing IRT equating coefficients. *Journal of Educational Measurement*, 28, 2, pp.147-162.
- Baker, F.B. (1985). *The Basis of Item Response Theory*. Porstmouth, N. H., Heineman.
- Balkan, L. (1970). *Les effects du bilinguisme français-anglais sur les aptitudes intellectuelles*, Bruxelles, Aimav. (Traducción al castellano, 1979. *Los efectos del bilingüismo sobre las aptitudes intelectuales*. Madrid: Marova.)

- Balkan, L. (1979): *Los efectos del bilingüismo sobre las aptitudes intelectuales*. Madrid: Marova.
- Bel, A.; Serra, J.M. y Vila, I. (1993). Estudio comparativo del conocimiento del catalán y del castellano al final del ciclo superior de E.G.B. En Siguán, M. (coord): *Enseñanza en dos lenguas*. (pp. 97-110). Barcelona: Horsori.
- Bempechat, J. y Ginsburg, H.P. (1989). Underachievement and educational disadvantage: Home and school. (Report No. UD 027 265). New York: ERIC Clearinghouse on Urban Education. (ERIC Document Reproduction Service No. ED 315 485).
- Ben-Zeev, S. (1975). The effect of Spanish-English bilingualism in children from less privileged neighborhoods on cognitive development and cognitive strategy. *Research report to Department of Health, Education and Welfare*.
- Ben-Zeev, S. (1977). *Class, codes and control. Vol. 1. Theoretical Studies Towards a Sociology of Language*. London: Routledge and Kegan Paul.
- Ben-Zeev, S. (1977a). The influence of bilingualism on cognitive strategy and cognitive development. *Child Development*, 48, 1009-1018.
- Ben-Zeev, S. (1977b). The effect of bilingualism in children from Spanish-English low economic neighbourhoods on cognitive development and cognitive strategy. *Working Papers on Bilingualism*, 14, 83-122.
- Bennet, R.E.; Rock, D.A. y Kaplan, B.A. (1987). SAT differential item performance for nine handicapped groups. *Journal of Educational Measurement*, 24, 1, 41-55.
- Bialystok, E. (1987). Influences of bilingual on metalinguistic development. *Second Language Research*, 3, 2, 154-166.
- Bialystok, E. y Ryan, E.B. (1985). Toward a definition of metalinguistic skill. *Merrill-Palmer Quarterly*, 31, 3, 229-251.
- Binet, A. y Simon, T. (1916). *The development of intelligence in children*. New York: Arno. (Original work published, 1916).
- Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability. In F.M. Lord y M.R. Novick, *Statistical theories of mental test scores* (pp. 392- 479). Reading, MA: Addison-Wesley.
- Bishop, Y.M.M.; Fienberg, S.E. y Holland, P.W. (1975). *Discrete multivariate analysis: Theory and practice*. Cambridge, MA: MIT Press.
- Bleistein, C.A. y Wright, D. (1986). *Assessment of unexpected differential item difficulty for Asian-American examinees on the Scholastic Aptitude Test*. Trabajo presentado en el National Council on Measurement in Education annual meeting, San Francisco.
- Bleistein, C.A. y Wright, D. (1987). *Assessment of unexpected differential item difficulty for Asian-American examinees on the Scholastic Aptitude Test*. In A. Schmitt y N. Dorans (Eds.), *Differential item functioning on the Scholastic Aptitude Test* (Research Memorandum No. RM-87-1). Princeton, NJ: Educational Testing Service.
- Bloomfield, L. (1935). *Language*. Londres: Allend and Unwin.
- Boada, H. (1986). Comunicación y bilingüismo. En M. Siguán y cols. (coords), *Estudios de Psicolingüística*. Madrid: Pirámide.
- Bochner, S. (1996). The learning strategies of bilingual versus monolingual students. *British Journal of Educational Psychology*, 66, 1, 83-93.
- Boixaderas, R.; Canal, I. y Fernández, E. (1991). Avaluació del nivells de llengua catalana, castellana i matemàtiques en alumnes que han seguit el programa de immersió lingüística. *Segon Simposi sobre l'Ensenyament del Català a Non-Catalano parlants*.

- Bontempo, R. (1993). Translation fidelity of psychological scales. An item response theory analysis of an individualism-collectivism scale. *Journal of Cross-Cultural Psychology*, 24, 2, 149-166.
- Budgell, G.R.; Raju, N.S. y Quartetti, D.A. (1995). Analysis of differential item functioning in translated assessment instruments. *Applied Psychological Measurement*, 19, 4, 309-321.
- Bueno, V. (1995). *Bilingüisme i educació a la Comunitat Valenciana*. Tesis doctoral no publicada. Universitat de València.
- Camilli, G. (1979). *A critique of the Chi-Square method of assessing Item Bias*. Laboratory of Educational Research, Boulder, CO: University of Colorado.
- Camilli, G. (1992). A conceptual analysis of differential item functioning in terms of a multidimensional item response model. *Applied Psychological Measurement*, 16, 2, 129-147.
- Camilli, G. y Smith, J.K. (1990). Comparison of the Mantel-Haenszel Test with a randomized and a jackknife test for detecting biased items. *Journal of Educational Statistics*, 15, 53-67.
- Camilli, G. Wang, M. y Fesq, J. (1995). The effects of Dimensionality on Equating the Law School Admission Test. *Journal of Educational Measurement*, 32, 1 79-96.
- Camilli, G. y Shepard, L.A. (1994). *Methods for identifying biased test items*. Newbury Park, CA: Sage.
- Candell, G.L. y Drasgow, F. (1988). An alternative procedure for linking metrics and assessing Item Bias in Item Response Theory. *Applied Psychological Measurement*, 12, 3, 253-260.
- Candell, G.L. y Hulin, Ch.L. (1987). Cross-Language and Cross-Cultural comparisons in scale translations: Independent sources of information about item functioning. *Journal of Cross-Cultural Psychology*, 17, 4, 417-440.
- Carlton, S.T. y Harris, A.M. (1989a). *Characteristics of differential item performance on the Scholastic Aptitude Test-selected ethnic group comparisons*. Trabajo presentado en el Annual Meeting of the National Council on Measurement in Education, San Francisco.
- Carlton, S.T. y Harris, A.M. (1989b). *Female/male performance differences on the SAT: Causes and correlates*. Paper presented at the annual meeting of the American Educational Research Association, San Francisco.
- Carlton, S.T. y Harris, A.M. (En preparación). *Characteristics associated with differential item performance on the Scholastic Aptitude Test: Gender and majority/minority group comparisons*.
- Clauser, B.; Mazor, K. y Hambleton, R.K. (1993). The effects of purification of the matching criterion on the identification of DIF using the Mantel-Haenszel procedure. *Applied Measurement in Education*, 6, 4, 269-279.
- Clauser, B.; Mazor, K.M. y Hambleton, R.K. (1994). The effects of score group width on the Mantel-Haenszel procedure. *Journal of Educational Measurement*, 31, 1, 67-78.
- Clauser, B.E.; Mazor, K. y Hambleton, R.K. (1991). Influence of the criterion variable on the identification of differentially functioning test items using the Mantel-Haenszel statistic. *Applied Psychological Measurement*, 15, 4, 353-359.
- Cleary, T.A. y Hilton, T.J. (1968). An investigation of item bias. *Educational and Psychological Measurement*, 5, 115-124.

- Cohen, A.D. (1974). The Culver City Spanish Immersion Programme: The first two years. *The Modern Language Journal*, 58, 95-103.
- Cohen, A.D. (1974). The Culver City Spanish immersion project: The first two years. *Modern Language Journal*, 58, 95-103.
- Cohen, A.D. (1975). *Bilingual Education. A Sociolinguistic approach to Massachusetts*. Rowley: Newbury House Publishers Inc.
- Cohen, A.S.; Kim, S. y Baker, F.B. (1993). Detection of differential item functioning in the graded response model. *Applied Psychological Measurement*, 17, 4, 335-350.
- Cole, N.S. y Moss, P.A. (1989). *Bias in Test Use*. In R.L. Linn (Ed.), Educational Measurement. New York, Macmillan.
- Conselleria de Cultura, Educaci3n y Ciencia (1991). *Un proc3s en marxa*. Val3ncia: Conselleria de Cultura, Educaci3n i Ci3ncia.
- Conselleria de Cultura, Educaci3n y Ciencia (1992). *Compilaci3n legislativa (1980-1991)*. Val3ncia: Conselleria de Cultura, Educaci3n i Ci3ncia.
- Conselleria de Educaci3n y Ciencia (1990a). *Coneixement del valenci3. An3lisi dels resultats del padr3 municipal d'habitants de 1986*. Val3ncia: Conselleria de Cultura, Educaci3n i Ci3ncia.
- Conselleria de Educaci3n y Ciencia (1990b). *Dades sobre la situaci3 sociolingüística de la Comunitat Valenciana*. Val3ncia: Conselleria de Cultura, Educaci3n i Ci3ncia.
- Cuc3, A. (1989). *País i Estat: la qüesti3 valenciana*. Tres i Quatre. Val3ncia.
- Cummins, J. (1978). Metalinguistic development of children in bilingual education programs: Data from Irish and Canadian Ukrainian-English Programs. En M. Paradis (Ed.), *Aspects of Bilingualism*. Columbia, Hornbeam Press.
- Cummins, J. (1979a). Linguistic interdependence and the educational development of bilingual children. *Review of Educational Research*, 49, 2, 222-251.
- Cummins, J. (1979b). Cognitive/academic language proficiency, linguistic interdependence, the optimal age question and some other matters. *Working Papers on Bilingualism*, N3. 19, Toronto, Ontario Institute for Studies in Education.
- Cummins, J. (1980). The Cross Lingual Dimensions of Language Proficiency: The Implications for Bilingual Education and the Optimal Age Issue. *Tesol Quarterly*, 14, 2, 175-187.
- Cummins, J. (1981). The role of primary language development in promoting educational success for language minority students in California State Department of Education. En *Schooling and language minority students: A theoretical framework*, Los Angeles, Evaluation, Dissemination and Assessment Center.
- Cummins, J. (1983). Interdependencia lingüística y desviaciones educativas en los niños bilingües. *Infancia y Aprendizaje*, 21, 37-61.
- Cummins, J. (1983a). *Heritage Language Education: A Literature Review*. Ontario. Ministry of Education.
- Cummins, J. (1983b). Language proficiency, biliteracy and French immersion. *Canadian Journal of Education*, 8, 2, 117-138.
- Cummins, J. (1984). Wanted: A theoretical framework for relating language proficiency to academic achievement among bilingual students. En C. Rivera: *Language proficiency and academic achievement*, Clevedon, Multilingual Matters.
- Cummins, J. (1986a). Bilingual education and anti-racist education. *Interracial Books for Children Bulletin*, 17, (3 y 4), 9-12.

- Cummins, J. (1986b). Empowering minority students: A framework for intervention. *Harvard Educational Review*, 56, 1, 18-36.
- Cummins, J. (1987). *L'éducation bilingue; théorie et mise en oeuvre*. En *l'Education multiculturelle*. París: O.C.D.E., pp. 323-355.
- Cummins, J. (1988). Research and theory in bilingual education: the basque situation in international perspective. En *II Euskal Mundu Biltzarra-II Congreso Mundial Vasco*. Gasteiz-Vitoria, pp. 198.
- Cummins, J. y Gulutsan, M. (1974). Bilingual education and cognition. *The Alberta Journal of Educational Research*, XX, 3.
- Cummins, J. y Mulcahy, R. (1978). Orientation to language in Ukrainian-English bilingual children. *Children Development*, 49, 1239-1242.
- Chang, H.; Mazzeo, J. y Roussos, L. (1996). Detecting DIF for polytomously scored items: An adaptation of the SIBTEST procedure. *Journal of Educational Measurement*, 33, 3, 333-353.
- Darcy, N.T. (1953). A review of the literature on the effects of bilingualism upon the measurement of intelligence. *Journal of Genetic Psychology*, 82, 21-57.
- Díaz, R.M. (1986). Bilingüismo e inteligencia: Una revisión. En Siguán, M. y cols., *Las lenguas minoritarias y la educación*. Barcelona: I.C.E. Universidad de Barcelona.
- Díaz, R.M. (1985). Bilingual Cognitive development. Addressing three gaps in current-research. *Child Development*, 56, 6, 1376-1388.
- Díaz, R.M. y Klinger, C. (1991). Towards an explanatory model of the interaction between bilingualism and cognitive development. En Ellen Bialystock (Ed.): *Language processing in bilingual children*, Cambridge, Cambridge University Press.
- Diez, I. (1995). *Avaluació de l'experiència de l'ús vehicular del valencià en l'E.G.B.* Tesis doctoral no publicada. Universitat de València.
- Domenech, J.Ll. (1990). *Avaluació i seguiment del valencià com a llengua vehicular d'ensenyament*. Tesis doctoral no publicada. Universitat de València.
- Dorans, N.J. (1989). Two new approaches to assessing differential item functioning: Standardization and the Mantel-Haenszel method. *Applied Measurement in Education*, 3, 217-233.
- Dorans, N.J. y Holland, P.W. (1992). DIF detection and description: Mantel-Haenszel and standardization. In P.W. Holland y H. Wainer (Eds.), *Differential item functioning: Theory and practice* (pp. 35-66). Hillsdale, NJ: Erlbaum.
- Dorans, N.J. y Holland, P.W. (1993). DIF Detection and description: Mantel-Haenszel and standardization. In P.W. Holland y H. Wainer (Eds.) *Differential item functioning* (pp. 35-66). Hillsdale, NJ: LEA.
- Dorans, N.J. y Kulick, E.M. (1986). Demonstrating the utility of the standardization approach to assessing unexpected differential item functioning. *Journal of Educational Measurement*, 29, 309-319.
- Dorans, N.J. y Schmitt, A.P. (1993). Constructed response and differential item functioning: A pragmatic approach. In R.E. Bennett y W.C. Ward (Eds.), *Construction vs. choice in cognitive measurement* (pp. 135-165). Hillsdale, NJ: Erlbaum.
- Dorans, N.J.; Schmitt, A.P. y Bleinstein, C.A. (1992). The standarization approach to assessing comprehensive differential item functioning. *Journal of Educational Measurement*, 29, 309-319.
- Dorans, N.J.; Schmitt, A.P.; y Curley, W.E. (1988). *Differential speededness: Some items have DIF because of where they are, not what*

- they are. Trabajo presentado en el Annual Meeting of the National Council on Measurement in Education, New Orleans, LA.
- Drasgow, F. (1987). Study of the measurement bias of two standardized psychological tests. *Journal of Applied Psychology*, 72, 19-29.
- Drasgow, F., y Hulin, C.L. (1986). *Assessing the equivalence of measurement of attitudes and aptitudes cross heterogeneous subpopulations*. Unpublished manuscript. University of Illinois at Champaign-Urbana.
- Duncan, S.E. y De Avila, E.A. (1979). Bilingualism and Cognition: Some Recent Findings. *NABE Journal*, 4, 15-50.
- Edelsky, C. y otros. (1986). Semilingualism and Language Deficit. *Applied Linguistics*, 4, 1-22.
- Eells, K.; Havighurst, R.J.; Herrick, V.E. y Tyler, R.W. (1951). *Intelligence and cultural differences*. Chicago: University of Chicago, Press.
- Elosua, P. (1996). *Adaptación al euskera de una prueba verbal. Estudio del funcionamiento diferencial de los ítems*. Tesis no publicada. Universidad del País Vasco.
- Ellis, B.B. (1989). Differential Item Functioning: Implications for Test Translations. *Journal of Applied Psychology*, 74, 6, 912-921.
- Ervin, S. y Osgood, C. (1954). Second Language Learning and Bilingualism. *Journal of Abnormal and Social Psychology*. Apèndix, 139-146.
- Etxebarria, F. (1992). *Bilingüismo y educación en el País Vasco*. Donostía: Universitat del País Vasco (inèdit).
- Etxeberria, F. (1986). *Bilingüismo escolar e investigación en Euskadi*. Comunicación presentada a las II Jornadas Internacionales de Psicología y Educación. Madrid, 23-27 de junio.
- Everson, H.T.; Millsap, R.E. y Rodríguez, C.M. (1991). Isolating gender differences in test anxiety: A confirmatory factor analysis of the test anxiety inventory. *Educational and Psychological Measurement*, 51, 243-251.
- Fernandez Mauro, J.A. (1981). Aspectos lingüísticos de la educación bilingüe. *Revista de Educación*, 268, 179-199.
- Fidalgo, A.M. (1994). MHDIF: A computer program for detecting uniform and nonuniform differential item functioning with the Mantel-Haenszel procedure. *Applied Psychological Measurement*, 18, 3, 300.
- Fidalgo, A.M. (1996). *Funcionamiento diferencial de los ítems: Procedimiento Mantel-Haenszel y modelos loglineales*. Tesis doctoral no publicada. Universidad de Oviedo.
- Fidalgo, A.M. y Mellenbergh, G.J. (1995). *Evaluación del procedimiento Mantel-Haenszel frente al método logit iterativo en la detección del funcionamiento diferencial de los ítems uniforme y no uniforme*. Comunicación presentada al IV Symposium de Metodología de las Ciencias del Comportamiento, La Manga del Mar Menor, Murcia.
- Fidalgo, A.M. y Paz, M.D. (1995a). Modelos lineales logarítmicos y funcionamiento diferencial de los ítems. *Anuario de Psicología*, 64, 1, 57-66.
- Fidalgo, A.M. y Paz, M.D. (1995b). *Comparación del método logit iterativo frente a los modelos loglineales en la detección del funcionamiento diferencial de los ítems*. Comunicación presentada al IV Symposium de Metodología de las Ciencias del Comportamiento, La Manga del Mar Menor, Murcia.
- Fidalgo, A.M.; Mellenbergh, G.J. y Muñiz, J. (1998). Comparación del procedimiento Mantel-Haenszel frente a los modelos loglineales en la

- detección del funcionamiento diferencial de los ítems. *Psicothema*, 10, 1, 209-218.
- Fishman, J.A. (1976). *Bilingual Education*. Rowley. Newbury House.
- Fishman, J.A. (1977). Language and ethnicity. En H. Giles (ed.), *Language, ethnicity and intergroup relations*. London, Academic Press.
- Fishman, J.A. y Lovas, J. (1970). Bilingual education in sociolinguistic perspective. *TESOL Quarterly*, 4, 215-222.
- Freedle, R. y Kostin, I. (1987). *Semantic and structural factors affecting the performance of matched Black and White examinees on analogies items from the Scholastic Aptitude Test* (Research Final Report, PRC Project). Princeton, NJ: Educational Testing Service.
- Freedle, R. y Kostin, I. (1988). *Relationship between item characteristics and an index of Differential Item Functioning (DIF) for the four GRE verbal item types* (Research Rep. No. 88-29). Princeton, NJ: Educational Testing Service.
- French, A.W. y 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.
- Fuster, J. (1962). *Nosaltres els valencians*. València: Tres i Quatre.
- Gabinet d'Estudis del SEDEC: Alsina, Amau, Bel, Garolera, Grijalvo, Perera y Vila (1983). *Quatre Anys de Català a l'Escola*. Barcelona: Departament d'Ensenyament de la Generalitat de Catalunya.
- Gabiña, J.J.; Gorostidi, R.; Iruretagoiena, R.; Olaziregi, I. y Sierra, J. (1986). *EIFE. La Enseñanza del Euskera: Influencia de los Factores*. Victoria-Gasteiz: Eusko Jaurlaritzaren Argitalpen-Zerbitzu Nagusia.
- Garagorri, X. y Eguillor, E. (1983). Desarrollo y situación actual de la enseñanza en las Ikastolas. En M. Siguán (ed.), *Lenguas y Educación en el ámbito del Estado Español*. Barcelona: Publicacions i Edicions de la Universitat de Barcelona.
- Garmendia, M.C. (1985). *Les Ikastolas en Pays Basque d'Espagne: la Question du Bilinguisme Scolaire*. Doctorado de 3ème Cycle. Université de Strasbourg.
- Generalitat Valenciana (1983). *Estatut d'Autonomia de la Comunitat Valenciana*. València: Generalitat Valenciana.
- Genesse, F. (1992). *The Challenge of modern education: Diversity*. X Congreso Nacional de Pedagogía. Salamanca.
- Genesse, F. y Hamayan, E. (1978). Individual differences in young second language learners. *McGill University. Psychological Department*. Montreal.
- Gilbert, G.N. (1981). *Modelling Society. An introduction to loglinear analysis for social researchers*. London: George Allen y Unwin.
- Gómez, J. (1996). cap del libro de muñiz
- Gómez, J. y Hidalgo, M.D. (1997). Evaluación del funcionamiento diferencial en ítems dicotómicos: Una revisión metodológica. *Anuario de Psicología*, 74, 3, 1997.
- Gómez, J. y Navas, M. J. (1995). *Detección del funcionamiento diferencial de los ítems mediante regresión logística: purificación paso a paso de la habilidad*. Comunicación presentada al IV Symposium de Metodología de las Ciencias del Comportamiento. La Manga del Mar Menor. Murcia.
- Gómez, J. y Navas, M. J. (1996). Detección del funcionamiento diferencial de los ítems mediante regresión logística: purificación paso a paso de la habilidad. *Psicológica*, 17, 2, 397-411.



- González Lorenzo, M. (1985). *Bilingüismo en Galicia. Problemas y Alternativas*. Santiago de Compostela: Universidad de Santiago de Compostela.
- González, V. (1994). A model of cognitive, cultural and linguistic variables affecting bilingual Hispanic children's development of concepts and language. *Hispanic Journal of Behavioral Sciences*, 16, 4, 396-421.
- Gotor, A.; Miralles, J.L. y Perea, M. (1990). *Rendimiento diferencial en función de la lengua de escolarización*. II Congreso del Colegio Oficial de Psicólogos. Madrid: Colegio Oficial de Psicólogos.
- Groosjean, F. y Miller, J. L. (1994). Going in and out of languages: An example of bilingual flexibility. *Psychological Science*, 5, 4, 201-206.
- Grosjean, F. (1982). *Life with two languages. An introduction to bilingualism*. Harvard: Harvard University Press.
- Gulliksen, H. (1950). *Theory of mental tests*. Nueva York. John Wiley and Sons (reimpreso en 1987).
- Hambleton, R.K. (1993). Translation achievement tests for use in cross-national studies. *European Journal of Psychological Assessment*, 9, 1, 57-68.
- Hambleton, R.K. (1994). Guidelines for adapting educational and psychological tests: A progress report. *European Journal of Psychological Assessment*, 10, 3, 39-52.
- Hambleton, R.K. y Swaminathan, H. (1985). *Item response theory*. Boston MA: Kluwer- Nijhoff.
- Hambleton, R.K.; Clauser, B.E.; Mazor, K.M. & Jones, R.W. (1993). Advances in the detection of differentially functioning test item. *European Journal of Psychological Assessment*, 9, 1, 1-18.
- Hambleton, R.K.; Swaminathan, H. y Rogers, H.J. (1991). *Fundamentals of item response theory*. Newbury Park, CA: Sage Publishers.
- Hamers, J.F. y Blanc, M. (1982). Toward a social psychological model of bilingual development. *Journal of Language and Social Psychology*, 1, nº 1, 29-50.
- Hamers, J.F. y Blanc, M. (1983). *Bilinguisme et bilinguisme*. Lieja. Pierre Mardaga.
- Helms-Lorenz, M. y Van de Vijver, F.J.R. (1995). Cognitive assessment in Education in a Multicultural Society. *European Journal of Psychological Assessment*, 11, 3, 158-169.
- Hidalgo, M.D. (1995). *Evaluación del funcionamiento diferencial del ítem en ítems dicotómicos y politómicos: un estudio comparativo*. Tesis doctoral no publicada. Universidad de Murcia.
- Hidalgo, M.D. y López Pina, J.A. (1997). Comparación entre las medidas de área, el estadístico de Lord y el análisis de regresión logística en la evaluación del funcionamiento diferencial de los ítems. *Psicothema*, 9, 417-431.
- Hidalgo-Montesinos, M.D. y Gómez-Benito, J. (1996). The effect of ability purification on the evaluation of differential item functioning with the technique of multinomial logistic regression. *20th Biennial Conference of the Society for Multivariate Analysis in the Behavioral Sciences*, Barcelona (Spain), 8th-10th July, 1996.
- Hills, J.R. (1989). Screening for potentially biased items in testing programs. *Educational Measurement*, 8, 4, 5-10.

- Holmstrand, L.E. (1979). *The effects of general school achievement of early commencement of English Instruction*. Uppsala Reports on Educations, 4, 1-45. Uppsala: University of Uppsala.
- Holland, P.W. y Thayer, D.T. (1985). *An alternative definition of the ETS delta scale of item difficulty*. (Research Rep. No. 85-43). Princeton, NJ: Educational Testing Service.
- Holland, P.W. y Thayer, D.T. (1986). *Differential item functioning and the Mantel-Haenszel procedure*. (Technical Rep. No. 86-69). Princeton, NJ: Educational Testing Service.
- Holland, P.W. y Thayer, D.T. (1988). Differential item performance and the Mantel-Haenszel procedure. In H. Wainer and H.I. Braun (Eds), *Test validity* (pp.129-145). Hillsdale NJ: Erlbaum
- Hosmer, D.W. y Lemeshow, S. (1989). *Applied logistic regression*. New York: Wiley.
- Hulin, C.L.; Drasgow, F. y Komocar, J. (1982). Applications of item response theory to analysis of attitude scale translations. *Journal of Applied Psychology*, 67, 818-825.
- Hulin, C.L.; Drasgow, F. y Parsons, C.K. (1983). *Item response theory: Applications to psychological measurement*. Homewood, IL: Irwin.
- Hulin, Ch.L. y Mayer, L.J. (1986). Psychometric Equivalence of a Translation of the Job Descriptive Index into Hebrew. *Journal of Applied Psychology*, 71, 1, 83-94.
- Husen, T. y Oppen, S. (1984). *Educación multicultural y multilingüe*. Madrid. Narcea.
- Ianco-Worrall, A.D. (1972). Bilingualism and cognitive development. *Child Development*, 43, 1.390.
- Ironson, G.H. y Subkoviak, M.J. (1979). A comparison of several methods of assessing item bias. *Journal of Educational Measurement*, 16, 209-225.
- Isaacs (1976) ??
- Jensen, A.R. (1969). How much can we boast IQ and scholastic achievement? *Harvard Educational Review*, 39, 1-123.
- Jensen, A.R. (1980). *Bias in mental testing*. New York: Free Press.
- Jones, J.E. (1989). Changing needs for a changing future: The need for educational leadership. New York: National Center for children in poverty. (ERIC Document Reproduction Service No. 316 079).
- Jones, W. y Steward, W. (1951). Bilingualism and verbal intelligence. *British Journal of Psychology*, 4, 3-8.
- Joreskog, K.L. y Sörbom, D. (1989). *LISREL VII user's reference guide* [Manual de programa para ordenador]. Mooresville IN: Scientific Software.
- Joreskog, K.L. y Sörbom, D. (1992). amplia al LISREL VIII.
- Jöreskog, K.G. & Sörbom, D. (1993). *LISREL VIII. User's guide*. Mooresville, Ill.: Scientific Software, Inc.
- Kelderman, H. (1989). Item bias detection using loglinear IRT. *Psychometrika*, 54, 681-697.
- Keston, M. y Jimenez, C. (1954). A study of the performance on English and Spanish editions of the Stanford-Binet Intelligence Test by Spanish-American Children. *Journal of Genetical Psychology*, 85, 263-369.
- Kim, S. y Cohen, A.S. (1991). A comparison of two Area Measures for detecting differential item functioning. *Applied Psychological Measurement*, 15, 3, 269-278.
- Kim, S. y Cohen, S. (1992). Effects of Linking Methods on Detection of DIF. *Journal of Educational Measurement*, 29, 1, 51-66.
- Knobe, D. y Burke, P.J. (1980). *Log-linear models*. Beverly Hills: SAGE.

- Kok, F. (1988). Item bias and test multidimensionality. In R. Langeheine y J. Rost (Eds.). *Latent trait and latent class model* (pp. 263-275). New York: Pleum.
- Kok, F.G.; Mellenbergh, G.J. y Van der Flier, H. (1985). Detecting experimentally induced item bias using the iterative logit method. *Journal of Educational Measurement*, 54, 21-31.
- Kulick, E. y Dorans, N.J. (1983). *Assessing unexpected differential item performance of Oriental candidates on SAT form CSA6 and TSWE form E33 (SR-83-106)*. Princeton, NJ: Educational Testing Service.
- Lambert, W.E. (1974). Culture and language as factors in learning and education. En F. Aboud y R.D. Meade (Eds.). *Cultural Factors in learning*, Bellingham, Western Washington State College.
- Lambert, W.E. y Anisfeld, D. (1969). A note on the relationship of bilingualism and intelligence. *Canadian Journal of Behavioral Science*, 1, pp. 123-128.
- Lambert, W.E. y Tucker, C.R. (1972). *Bilingual education of children. The St. Lambert Experiment*. Rowley. Newbury House.
- Lambert, W.E.; Tucker, C.R. y D'Anglejan, A. (1973). Cognitive and attitudinal consequences of bilingual schooling. The St. Lambert project through grade five. *Journal of Educational Psychology*, 1, 44-51.
- Lautenschlager, G.J.; Flaherty, V.L. & Park, D.G. (1994). IRT differential item functioning: An examination of ability scale Purifications. *Educational and Psychological Measurement*, 54, 1, 21-31.
- Lewis, C. (1993). A note on the value of including the studied item in the test score when analyzing test for DIF. En W.P. Holland y H.Wainer (Ed.). *Differential Item Functioning* (pp. 317-319). Hillsdale, NJ: LEA.
- Li, H.H. y Stout, W. (1996). A new procedure for detection of crossing DIF. *Psychometrika*, 61, 4, 647-677. (mirar si es 1994 a ó b)
- Linn, R.L.; Levine, M.V.; Hastings, C.N. y Wardrop, J.L. (1981). Item bias in a Test of Reading Comprehension. *Applied Psychological Measurement*, 5, 159-173.
- Lopez Pina, J.A.; Hidalgo, M.D. y Sánchez Meca, J. (1997). *Error tipo 1 de las pruebas chi-cuadrado en el estudio del sesgo de los items*. Comunicación presentada al III Symposium de Metodología de las Ciencias del Comportamiento. Santiago de Compostela.
- Lord, F.M. (1977). A study of item bias, using item characteristic curve theory. In Y.H. Poortinga (Ed.) *Basic problems in cross-cultural psychology* (pp. 19-29). Amsterdam: Swets y Zeitlinger.
- Lord, F.M. (1980). *Applications of item response theory of practical testing problems*. Hillsdale, NJ: Lawrence Erlbaum.
- Lord, F.M. y Novick, M.R. (1968). *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.
- MacNamara, J. (1967a). The bilingual linguistic performance. A psychological overview. *Journal of Sociological Issues*, 23, 58-77.
- MacNamara, J. (1967b). The linguistic independence of bilinguals. *Journal of Verbal Learning and Verbal Behavior*, 6, 729-736.
- MacNamara, J. (1969). *How can one measure the extent of a person's bilingual dominance?* En L.G. Kelly (ed.). *Description and measurement of bilingualism*. Toronto. University of Toronto, Press.
- Mägiste, E. (1982). The importance of language strategy in simple arithmetic. *Educational Psychology*, 2, 2.

- Mantel, N. y Haenszel, W. (1959). Statistical aspects of the analysis of data from retrospective studies of disease. *Journal of the National Cancer Institute*, 22, 719-748.
- Marco, G.L. (1977). Item Characteristic Curve solutions to three intractable testing problems. *Journal of Educational Measurement*, 14, 2, 139-160.
- Martínez-Arias, M.R. (1994). *Psicometría: teoría de los tests psicológicos y educativos*. Madrid: Síntesis.
- Mazor, K.M.; Clauser, B.E. y Hambleton, R.K. (1992). The effect on sample size on the functioning of the Mantel-Haenszel statistic. *Educational and Psychological Measurement*, 52, 443-451.
- Mazor, K.M.; Clauser, B.E. y Hambleton, R.K. (1994). Identification of nonuniform differential item functioning using a variation of the Mantel-Haenszel Procedure. *Educational and Psychological Measurement*, 54, 2, 284-291.
- Mazor, K.M.; Kanjee, A. y Clauser, B.E. (1995). Using Logistic Regression and the Mantel-Haenszel with multiple ability estimates to detect differential item functioning. *Journal of Educational Measurement*, 32, 2, 131-144.
- McPeck, W.M. y Wild, C.L. (1986). *Performance of the Mantel-Haenszel statistic in a variety of situations*. Trabajo presentado en el Annual Meeting of the American Educational Research Association, San Francisco.
- Mellenbergh, G.J. (1982). Contingency table models for assessing item bias. *Journal of Educational Measurement*, 7, 105-108.
- Mellenbergh, G.J. (1995). Conceptual notes on models for discrete polytomous item response. *Applied Psychological Measurement*, 19, 1, 91-100.
- Meredith, W. y Millsap, R.E. (1992). On the misuse of manifest variables in the detection of measurement bias. *Psychometrika*, 57, 2, 289-311.
- Miller y Hirsch (1992)
- Miller, M.D. y Oshima, T.C. (1992). Effect of sample size, number of biased items, and magnitude of bias on a two-stage item bias estimation method. *Applied Psychological Measurement*, 16, 4, 381-388.
- Miller, T.R. y Spray, J.A. (1993). Logistic discriminant function analysis for DIF identification of polytomously scored items. *Journal of Educational Measurement*, 30, 2, 107-122.
- Miller, T.R., Spray, J.A. y Wilson, A.W. (1992). *A comparison of three methods for identifying nonuniform DIF in polytomously scored test items*. Comunicación presentada en el Psychometric Society Conference, Columbus, OH.
- Millsap, R.E. y Everson, H.T. (1991). Confirmatory measurement model comparison using latent means. *Multivariate Behavioral Research*, 26, 479-497.
- Millsap, R.E. y Everson, H.T. (1993). Methodology Review: Statistical approaches for assessing measurement bias. *Applied Psychological Measurement*, 17, 4, 297-334.
- Millsap, R.E. y Meredith, W. (1992). Inferential conditions in the statistical detection of measurement bias. *Applied Psychological Measurement*, 16, 4, 389-402.
- Miralles, J.L.; Gotor, A.; Cervera, T.; Monterde, H. y Sempere, M.C. (1989). Coneixements lingüístics a primer cicle d'EGB d'alumnes escolaritzats en valencià. En *Suport per a l'Ensenyament en Valencià*,

3. València: Conselleria de Cultura, Educació i Ciència de la Generalitat Valenciana.
- Molla, A. y Palanca, C. (1987). *Curs de sociolingüística*. Vol 1. Alzira: Bromera
- Muñiz, J. (1990). *Teoría de respuesta a los ítems*. Madrid: Pirámide.
- Muñiz, J. (1992). *Teoría clásica de tests*. Madrid: Pirámide.
- Nandakumar, R. (1993a). A Fortran 77 program for detecting differential item functioning through the Mantel-Haenszel statistic. *Educational and Psychological Measurement*, 53, 679-684.
- Nandakumar, R. (1993c). Refinements of Stout's procedure for Assessing Latent Trait Unidimensionality. *Journal of Educational Statistics*, 18, 1, 41-68.
- Nandakumar, R. (1993b). Simultaneous DIF amplification and cancellation: Shealy-Stout's Test for DIF. *Journal of Educational Measurement*, 30, 4, 293-311.
- Nandakumar, R. y Yu, F. (1996). Empirical Validation of DIMTEST on Nonnormal Ability Distributions. *Journal of Educational Measurement*, 33, 3, 369-373.
- Narayanan, P. y Swaminathan, H. (1994). Performance of the Mantel-Haenszel and simultaneous item bias procedures for detecting differential item functioning. *Applied Psychological Measurement*, 18, 4, 315-328.
- Narayanan, P. y Swaminathan, H. (1996). Identification of items that show nonuniform DIF. *Applied Psychological Measurement*, 20, 3, 257-274.
- Navas, M.J. y Gómez, J. (1994). *Comparison of several bias detection techniques*. Comunicación presentada en el 23rd International Congress of Applied Psychology. Madrid. July 17-22.
- Ninyoles, R. (1986). Situació sociolingüística valenciana: factors estructurals. En Azurmendi M.J. y cols (Eds). *Les llengües en l'administració*. València: Conselleria de Cultura, Educació i Ciència.
- O'Neill, K.A.; McPeck, W.M. y Wild, C.L. (1989). *Characteristics of GRE verbal test items that show differential item functioning for Black and White examinees*. Trabajo presentado en el Annual Meeting of the American Educational Research Association, San Francisco.
- O'Neill, K.A.; McPeck, W.M. y Wild, C.L. (en preparación). *Identifying differentially functioning items on the Graduate Management Admission Test*.
- O'Neill, Wild, C.L. y McPeck, W.M. (en preparación). *Identifying differentially functioning items on the Graduate Record Examinations General Test*.
- Olaziregi, I. y Sierra, J. (1993). Diversas alternativas de enseñanza bilingüe. Evaluación y comparación. En Siguán, M. (coord.). *Enseñanza en dos lenguas*. (pp. 97-110). Barcelona: Horsori.
- Oort, F.J. (1992). Using restricted Factor Analysis to detect Item Bias. *Methodika*, VI, 150-166.
- Ortí Testillano, C. (1988). Experiencia lingüística y desarrollo cognitivo: las ventajas del bilingüismo. *Estudios de Psicología*, 33 i 34, 229-236.
- Osterlind, S.J. (1983). *Test item bias*. Beverly Hills: SAGE.
- Pardo, A. y San Martín, R. (1994). *Análisis de datos en psicología II*. Madrid: Pirámide.
- Park, D.G. (1988). *Investigations of item response theory item bias detection*. Unpublished doctoral dissertation, University of Georgia, Athens.

- Park, D.G. y Lautenschlager, G.J. (1990). Improving IRT item bias detection with iterative linking and ability scale purification. *Applied Psychological Measurement*, 14, 163-173.
- Pascual, J.; Domenech, J.L. y Sáez, A. (1989). El valencià com a llengua vehicular d'ensenyament a l'àrea d'experiències. Curs 1986-87. En *Suport per a l'Ensenyament en Valencià*, 3. València: Conselleria de Cultura, Educació i Ciència de la Generalitat Valenciana.
- Pasqual, J. y Sala, V. (1991). Un model educatiu per un sistema escolar amb tres llengües. València: Conselleria de Cultura, Educació i Ciència.
- Peal, E. y Lambert, W.E. (1962). The relation of bilingualism to intelligence. *Psychological Monographs*, 76, 1-23.
- Pérez Pardo, M. (1985). *Efectos del bilingüismo sobre el rendimiento escolar de un municipio de la provincia de Pontevedra*. En M. Sigúan (ed.) *Enseñanza en dos lenguas y resultados escolares*. Barcelona: Publicaciones y Ediciones de la Universitat de Barcelona.
- Pérez Vilaríño, J. (1979). Lengua y discriminación escolar en Galicia. En *Jornadas de Bilingüismo*. Peñíscola 1978. València: ICE de la Universitat Literaria de València.
- Pintner, R. y Arsenian, S. (1937). The relation of bilingualism to verbal intelligence and school adjustment. *Journal of Educational Research*, 31, 255-263.
- Plake, B.S. y Hoover, H.D. (1979). An analytical method of identifying biased test items. *The Journal of Experimental Education*, 48, 153-154.
- Potenza, M.T. y Dorans, N.J. (1995). DIF assessment for polytomously scored items: A framework for classification and evaluation. *Applied Psychological Measurement*, 19, 1, 23-37.
- Raju, N.S. (1990). Determining the significance of estimated signed and unsigned areas between two item response functions. *Applied Psychological Measurement*, 14, 197-207.
- Raju, N.S. (1988). The area between two item characteristic curves. *Psychometrika*, 53, 492-502.
- Raju, N.S.; Bode, R.K. y Larsen, V.S. (1989). An empirical assessment of the Mantel-Haenszel statistic for studying differential item performance. *Applied Measurement in Education*, 2, 1-13.
- Raju, N.S.; Drasgow, F. y Slinde, J.A. (1993). An empirical comparison of the area methods, Lord's chi-square tests, and the Mantel-Haenszel technique for assessing differential item functioning. *Educational and Psychological Measurement*, 53, 301-314.
- Reckase, M.D. (1979). Unifactor latent trait models applied to multi-factor tests: Results and implications. *Journal of Educational Statistics*, 4, 207-230.
- Reckase, M.D. (1985). The difficulty of test items that measures more than one ability. *Applied Psychological Measurement*, 9, 401-412.
- Reckase, M.D. y McKinley, R.L. (1991). The discriminating power of items that measure more than one dimension. *Applied Psychological Measurement*, 15, 4, 361-373.
- Resing, W.C.M.; Bleichrodt, N. y Drenth, P.J.D. (1986). Het gebruik van de RAKIT bij allochtoon etnische groepen. *Nederlands Tijdschrift voor de Psychologie*, 41, 179-188.
- Reynolds, H.T. (1977). *The analysis of cross-classifications*. New York: Free Press.
- Rivera, C. (1984). *Language Proficiency and Academic Achievement*. Clevedon, Multilingual Matters,

## REFERENCIAS BIBLIOGRAFICAS

---

- Roberts, C. (1985). *Teaching and learning commitment in bilingual schools*. Tesis doctoral inédita. University de Wales.
- Rogers, H.J. & Swaminathan, H. (1993). A comparison of Logistic Regression and Mantel-Haenszel Procedures for Detecting Differential Item Functioning. *Applied Psychological Measurement*, 17, 2, 105-116.
- Rogers, H.J. y Hambleton, R.K. (1989). Evaluation of computer simulated baseline statistics for use in item bias studies. *Educational and Psychological Measurement*, 49, 355-369.
- Rogers, H.J. y Kulick, E. (1987). An investigation of unexpected differences in item performance between Blacks and Whites taking the SAT. In A.P. Schmitt y N.J. Dorans (Eds.), *Differential item functioning on the Scholastic Aptitude Test (RM-87-1)*. Princeton, NJ: Educational Testing Service.
- Rogers, H.J. y Swaminathan, H. (1993). A comparison of Logistic Regression and Mantel-Haenszel Procedures for Detecting Differential Item Functioning. *Applied Psychological Measurement*, 17, 2, 105-116.
- Roussos, L.A. y Stout, W.F. (1996b). A Multidimensionality-Based DIF Analysis Paradigm. *Applied Psychological Measurement*, 20, 4, 355-371.
- Roussos, L.A. y Stout, W.F. (1996a). Simulation studies of the effects of small sample size and studied item parameters on SIBTEST and Mantel-Haenszel type I error performance. *Journal of Educational Measurement*, 33, 2, 215-230.
- Roznowski, M. (1987). Use of tests manifesting sex differences as measures of intelligence: Implications for measurement bias. *Journal of Applied Psychology*, 72, 480-483.
- Rudner, L.M. (1977). *An approach to biased item identification using latent trait measurement theory*. Trabajo presentado en el Annual Meeting of the American Educational Research Association, New York.
- Rudner, L.M., Getson, P.R. y Knight, D.L. (1980a). A montecarlo comparison of seven biased item detection techniques. *Journal of Educational Measurement*, 17, 1-10
- Rudner, L.M., Getson, P.R. y Knight, D.L. (1980b). Biased item detection techniques. *Journal of Educational Statistics*, 5, 213-233.
- Ruiz, P. (1974). *El bilingüismo y los tests de inteligencia verbales y no verbales*. Tesis de Licenciatura. València. Universitat de València.
- Saer, D. (1923). The effects of bilingualism on intelligence. *British Journal of Psychology*, 14, 25-38.
- Saer, D.; Smith, F. y Hughes, J. (1924). *The Bilingual Problem*. Wrexham, Hughes y Son.
- Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometric Monograph No 17*. Richmond, VA: William Byrd Press
- Sanchez-Lopez, M.P. y Rodriguez Tembleque, R. (1997). *El bilingüismo. Bases para la intervención psicológica*. Editorial Sintesis. S.A.
- Sánchez-López, M.P. (1980). *Diferencias en la estructura aptitudinal entre sujetos monolingües y bilingües*. Madrid. Editorial de la Universidad Complutense.
- Sánchez, M.P. y Rodriguez Tembleque, R. (1986). La educación bilingüe y el aprendizaje de una segunda lengua: sus características y principios fundamentales. *Infancia y Aprendizaje*, 33, 3-26.
- Sanchis, M. (1978). *La llengua dels valencians*. València: Tres i Quatre.
- Santa-Cruz, R.M. y Senour, M.N. (1989). *Empowering the Hispanic female in the public school setting*. Part II (Evaluation Dept. Rep. No. 617A). San Diego City Schools, Planning, Research and Evaluation Division.

- Scott, S. (1973). *The relation of divergent thinking to bilingualism: Cause or effect*. Informe de investigación. McGill University.
- Scheuneman, J. (1979). A method of assessing bias in test items. *Journal of Educational Measurement*, 16, 3, 143-152.
- Scheuneman, J.D. (1975). *A new method of assessing bias in test items*. Comunicación presentada en el Annual Meeting of the American Educational Research Association, Washington, DC. (ERIC Document Reproduction Service No. ED 106-359).
- Schmeiser, C.B. (1982). *Use of experimental design in statistical item bias studies*. En R. A. Berk (Ed.). *Handbook of methods for detecting test bias* (pp. 64-95). Baltimore: The Johns Hopkins University Press.
- Schmitt, A.P. (1985). *Assessing unexpected differential item performance of Hispanic candidates on SAT form 3FS A08 and TSWE form E47* (SR-85-169). Princeton, NJ: Educational Testing Service.
- Schmitt, A.P. (1986). *Unexpected differential item performance of Hispanic Examinees*. Comunicación presentada en el National Council on Measurement in Education, San Francisco.
- Schmitt, A.P. (1988). Language and cultural characteristics that explain differential item functioning for Hispanic examinees on the Scholastic Aptitude Test. *Journal of Educational Measurement*, 25, 1, 1-13.
- Schmitt, A.P. y Bleistein, C.A. (1987). *Factors affecting differential item functioning for Black examinees on Scholastic Aptitude Test analogy items* (RR-87-23). Princeton, NJ: Educational Testing Service.
- Schmitt, A.P. y Dorans, N. (1987a). *Differential item functioning for minority examinees on the Scholastic Aptitude Test*. Comunicación presentada en el Annual Meeting of the American Psychological Association, New York.
- Schmitt, A.P. y Dorans, N. (Eds). (1987b). *Differential item functioning on the Scholastic Aptitude Test*. (Research Memorandum No. RM-87-1). Princeton, NJ: Educational Testing Service.
- Schmitt, A.P. y Dorans, N.J. (1990). Differential item functioning for minority examinees on the SAT. *Journal of Educational Measurement*, 27, 1, 67-81.
- Secada, W.G. (1991). Degree of bilingualism and arithmetic problem solving in Hispanic first graders. *Elementary School Journal*, 92, 2, 213-231.
- SEDEC (1985). *Prova de català. Llengua 2. Nivell Inicial*. Barcelona: Departament d'Ensenyament de la Generalitat de Catalunya.
- Segall, D.O. (1983). *Test characteristic curves, item bias, and transformation to a common metric in item response theory: A methodological artifact with serious consequences and a simple solution*. Unpublished manuscript, University of Illinois.
- Serra, J.M. (1989). Resultados académicos y desarrollo cognitivo en un programa de inmersión dirigido a escolares de nivel socio-cultural bajo. *Infancia y Aprendizaje*, 47, 55-65.
- Serra, J.M. y Vila, I. (1991). *Estudio comparativo del conocimiento del catalán y el castellano en Octavo de EGB*. Comunicación presentada al I Seminario Internacional de Planificación Lingüística. Santiago de Compostela, 25-28 de Septiembre.
- Shealy, R. y Stout, W. (1993a). An item response theory model for test bias and differential test functioning. En P.W. Holland y H. Wainer (Eds.) *Differential item functioning* (pp.197-239). Hillsdale, NJ: LEA.
- Shealy, R. y Stout, W. (1993b). A model-based standardization approach that separates true bias/DIF from group ability differences and detects



- test bias/DIF as well as item bias/ DIF. *Psychometrika*, 58, 2, 159-194.
- Shealy, R., Stout, W. y Roussos, L. (1995). *SIBTEST (versión 1.3)*. [Programa de ordenador]. Urbana-Champaign: University of Illinois, Department of Applied Statistics.
- Shepard, L., Camilli, G. y Williams, D.M. (1985). Validity of approximation techniques for detecting item bias. *Journal of Educational Measurement*, 22, 77-105.
- Shepard, L.A.; Camilli, G. y Averill, M. (1981). Comparison of procedures for detecting test-item with both internal and external ability criteria. *Journal of Educational Statistics*, 6, 317-375.
- Shepard, L.A.; Camilli, G. y Williams, D.M. (1984). Accounting for statistical artifacts in item bias research. *Journal of Educational Statistics*, 9, 93-128.
- Sierra, J. y Olaziregi, I. (1989). *EIFE 2. La Enseñanza del Euskera: Influencia de los factores*. Gasteiz: Servicio Central de Publicaciones del Gobierno Vasco.
- Sierra, J. y Olaziregi, I. (1990). *EIFE 3. La Enseñanza del Euskera: Influencia de los factores*. Gasteiz: Servicio Central de Publicaciones del Gobierno Vasco.
- Siguán, M. y Mackey, F. (1986). *Educación y bilingüismo*. Madrid: Santillana/Unesco.
- Skutnabb-Kangas, T. (1978). Semilingualism and the education of migrant children as a means of reproducing the caste of assembly line workers. En DIFMAR, N y cols (1978): *Papers for the first Scandanavian-German symposium on the language of immigrant workers and their children*. Røkilde, Denmark: Universetscenter.
- Skutnabb-Kangas, T. y Toukoma, P. (1976). *Teaching Migrants Children's Mother Tongue and Learning the Language of the Host Country in the Context of the Socio-Cultural Situation of the Migrany Family*. Helsinki: The Finnish National Commission for UNESCO.
- Smith, P.C.; Kendall, L.M. y Hulin, C.L. (1969). *Measurement of satisfaction in work and retirement*. Chicago: Rand-McNally.
- Solomon, A. y Orenstein, D. (1991). *A study of the factors which contribute to the academic performance of first grade at-risk Hispanic students at a schoolwide projects facility*. Comunicación presentada en el Annual Meeting of the American Educational Research Association, Chicago, IL.
- Spence, A.G.; Mishra, S.P. y Ghozeil, S. (1971). Home language and performance on standardized tests. *The Elementary School Journal*, 71, 309-313.
- Spray, J.A. (1989). *Performance of three conditional DIF statistics in detecting differential item functioning on simulated tests* (Research Rep. No. 89-7). Iowa City, IA: American College Testing.
- Spray, J.A. y Carlson, J.E. (1986). *Comparison of loglinear and logistic regression models for detecting changes in proportions*. Comunicación presentada al Annual Meeting of the American Educational Research Association. San Francisco.
- Stein, R.F. (1990). *Families as learning environments: Reflections on critical factors affecting differential achievement*. Unpublished manuscript. (ERIC Document Reproduction Service No. 315 498).
- Steinberg, L.; Blinde, P.L. y Chan, K.S. (1984). Dropping out among language minority youth. *Review of Educational Research*, 54, 1, 113-132.

- Stern, W. (1914). *The psychological methods of testing intelligence*. Baltimore: Warwick y York.
- Stocking, M.L. y Lord, F.M. (1983). Developing a common metric in item response theory, *Applied Psychological Measurement*, 7, 201-210.
- Stout, W.; Nandakumar, R.; Junker, B.; Chang, H. y Steidinger, D. (1991). *DIMTEST* [Programa para ordenador]. Champaign IL: Department of Statistics. University of Illinois.
- Stout, W. (1987). A nonparametric approach for assessing latent trait unidimensionality. *Psychometrika*, 52, 589-617.
- Stout, W. (1990). A new item response theory modeling approach with applications to unidimensionality assessment and ability estimation. *Psychometrika*, 55, 293-325.
- Stout, W.; Li, H.; Nandakumar, R. y Bolt, D. (1997). MULTISIB: A procedure to investigate DIF when a test is intentionally two-dimensional. *Applied Psychological Measurement*, 21, 3, 195-213.
- Swain, M. (1975). Writing skills of grade 3 French immersion pupils. *Working Papers on Bilingualism*, 7, 1-38.
- Swain, M. (1978). French immersion: Early, later or partial?. *Canadian Modern Language Review*, 34, 677-585.
- Swain, M. y Lapkin, S. (1991). Heritage language children in an English-French bilingual program. *Canadian Modern Languages Review*, 47, 4, 635-641.
- Swain, M. y Lapkin, S. (1982). *Evaluating bilingual education: A Canadian case study*. Clevedon, Avon: Multilingual Matters Ltd.
- Swaminathan, H. y Rogers, H.J. (1990). Detecting Differential Item Functioning Using Logistic Regression Procedures. *Journal of Educational Measurement*, 27, 4, 361-370.
- Sympson, J.B. (1978). A model for testing with multidimensional items. En D. J. Weiss (Ed.). *Proceedings of the 1977 computerized adaptive testing conference* (pp. 82-98). Minneapolis: University of Minnesota.
- Tabouret-Keller, A. (1985). Lengua materna y lengua extranjera: De la existencia a la comunicación, preguntas de cara a una psicología clínica del bilingüismo. En Siguán, M. y cols (1985). *Enseñanza en dos lenguas y resultados escolares*. Barcelona: I.C.E.
- Thissen, D. y Steinberg, L. (1984). A response model for multiple-choice items. *Psychometrika*, 49, 501-519.
- Thissen, D. y Steinberg, L. (1986). A taxonomy of item response models. *Psychometrika*, 51, 567-577.
- Thissen, D.; Steinberg, L. y Gerrard, M. (1986). Beyond group mean differences: The concept of item bias. *Psychological Bulletin*, 99, 118-128.
- Thissen, D.; Steinberg, L. y Wainer, H. (1988). Use of item response theory in the study of group differences in trace lines. In H. Wainer y H.I. Braun (Eds). *Test validity* (pp.147-169). Hillsdale NJ: Erlbaum.
- Thissen, D.; Steinberg, L. y Wainer, H. (1993). Detection of differential item functioning using the parameters of item response models. En P.W. Holland y H. Wainer (Eds.) *Differential item functioning: Theory and Practice* (pp. 67-113). Hillsdale, NJ: Erlbaum.
- Titone, R. (1976). *Bilingüismo y Educación*. Barcelona, Fontanella (trad. del italiano, 1ª edición, 1972).
- Torrance, E.P.; Gowan, J.C.; Wu, J. y Aliotti, N.C. (1970). Creative functioning of monolingual and bilingual children in Singapore. *Journal of Educational Psychology*, 61, 1, 72-75.

- Torró, T. y Brotons, V. (1990). Models d'ensenyament d'aplicació a la Comunitat Valenciana. *I Jornades de Psicolingüística*. Xest, 1990.
- Tushima, W.T. y Hogan, T.P. (1975). Verbal ability and school achievement of bilingual and monolingual children ages. *Journal of Educational Research*, 68, 349-353.
- Tucker, G.R. (1977). The Linguistic Perspective. En *Bilingual Education: Current Perspectives*. Vol 2. Arlington: Center for Applied Linguistics.
- Uttaro, T. y Millsap, R.E. (1994). Factors Influencing the Mantel-Haenszel Procedure in the Detection of Differential Item Functioning. *Applied Psychological Measurement*, 18, 1, 15-25. 05.
- Van de Rijt, B. (1990). *Reactiesnelheidstest. Een aanvulling voor allochtonen op de bestaande intelligentietests*. Unpublished master's thesis. Tilburg University.
- Van der Flier, H., Mellenbergh, G.J.Ader, H.J. y Wijn, M. (1984). An iterative item bias detection method. *Journal of Educational Measurement*, 21, 131-145.
- van de Vijver, F.J.R. y Poortinga, Y.H. (1991). Testing across cultures. En R.K. Hambleton y J. Zaal (eds.), *Advances in educational and psychological testing* (pp. 277-308). Boston. MA: Klumer Academic Publishers.
- Vila, I. (1983). Reflexiones en torno al bilingüismo y la enseñanza bilingüe. *Infancia y aprendizaje*, 21, 4-22.
- Vila Moreno, I (1986). *Activitats lingüístiques en les aules preescolars d'immersió*. Comunicación presentada al II Congrés Internacional de la Llengua Catalana. Reus-Tarragona.
- Vila, I. (1991). *Els programes d'immersió a Catalunya*. Ponència en el Segon Simposi sobre l'ensenyament del català a no-catalano-parlants. Vic, 4-6 de septiembre.
- Vives Madrigal, M. (1979). EL fracaso escolar y el bilingüismo. En *Jornades de bilingüismo*. València: I.C.E.
- Wainer, H. (1993). Measurement problems. *Journal of Educational Measurement*, 30, 1, 1-21.
- Wainer, H.W.; Sireci, S.G. y Thissen, D. (1991). Differential testlet functioning: Definitions and detection. *Journal of Educational Measurement*, 28, 3, 197-219.
- Weinreich, U. (1953). *Languages in contact*. La Haya: Mouton.
- Welch, C. y Hoover, H.D. (1993). Procedures for extending item bias detection techniques to polytomously scored items. *Applied Measurement in Education*, 6, 1, 1-19.
- Welch, C.J. y Miller, T.R. (1995). Assessing Differential Item Functioning in Direct Writing Assessments: Problems and an Example. *Journal of Educational Measurement*, 32, 2, 163-178. 10.
- Wendler, C.L. y Carlton, S.T. (1987). *An examination of SAT verbal items for differential performance by women and men: An exploratory study*. Comunicación presentada en el Annual Meeting of the American Educational Research Association, Washington. DC.
- Whitely, S. (1980). Multicomponent latent trait models for ability tests. *Psychometrika*, 45, 479-494.
- Wild, C.L. y McPeck, W.M. (en preparación). *Identifying differentially functioning items in the NTE core battery*.
- Wilson, A.W.; Spray, J.A. y Miller, T.R. (1993). *Logistic regression and its use in detecting nonuniform differential item functioning in*

- polytomous items*. Comunicación presentada en el Annual Meeting of the National Council on Measurement in Education, Atlanta, GA.
- Willemse, G. (1989). *Keuze-reactietijdtaken in multi-culturele context*. Unpublished master's thesis. Tilburg University.
- Wright, B.D. y Stone, M.H. (1979). *Best test design*. Chicago, MESA.
- Zieky, M. (1993). Practical questions in the use of DIF statistics in test development. In P.W. Holland y H. Wainer (Eds.), *Differential item functioning* (pp.337-348). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
- Zwick, R. (1990). When do item response function and Mantel-Haenszel definitions of differential item functioning coincide? *Journal of Educational Statistics*, 15, 185-197.
- Zwick, R.; Donoghue, J.R. y Grima, A. (1993). Assessment of differential item functioning for performance tasks. *Journal of Educational Measurement*, 30, 3, 233-251.



## ***Anexos***



H.M.V.

# HABILIDAD MENTAL VERBAL

## ELEMENTAL

Este test consta de 40 preguntas, con seis alternativas de respuesta cada una.

### EJEMPLO T

«Fijaos en el ejemplo T... «*EL PERRO siempre tiene.*» Hay seis posibles respuestas, A, B, C, D, E, F... ¿Cuál es la más correcta?... La E, claro, porque el perro no siempre tiene amo, no siempre tiene collar, ni larga vida, ni hambre, ni instinto para cazar... Luego fijaos en la HOJA DE RESPUESTAS, en la primera columna, debajo de donde pone H. M. VERBAL, junto al ejemplo T marcad con una raya el rectángulo que tiene la letra E, por ser ésa la respuesta correcta...»

### EJEMPLO U

«Observad ahora el ejemplo U... "¿Qué número sigue?: 11 — 9 — 7 — 5..." Se trata de una serie de números que van disminuyendo de dos en dos (11, 9, 7, 5), luego a continuación del 5 debería venir el 3... La respuesta correcta es la B... Pues marcad, en la Hoja de Respuestas, primera columna, junto al ejemplo U, la letra B...»

### EJEMPLO V

«Mirad el ejemplo V... "Ordena las palabras y contesta la frase correcta: TERMOMETRO — UN — MIDE"... Tenéis tres palabras desordenadas, TERMOMETRO, UN, MIDE... Tenéis que ordenar mentalmente esas tres palabras sueltas para formar una frase con sentido y después ver cuál es la respuesta correcta... Bien ordenadas dirían: UN TERMOMETRO MIDE... La respuesta correcta es la D, la temperatura... Marcad en la Hoja de Respuestas, en el ejemplo V, la respuesta D...»

### EJEMPLO X

«"¿Qué palabra *sobra*?: TRIANGULO — CUADRADO — ROMBO — PENTAGONO — EXAGONO — SECTOR CIRCULAR..." En ese grupo de seis palabras tenéis que ver cuál es la más diferente, la más distinta, la que sobra... Como todas son figuras geométricas con tres, cuatro, cinco o seis lados, sobra SECTOR CIRCULAR, que es una parte de un círculo, una superficie... Marcad, pues, en el ejemplo X, en la Hoja de Respuestas, la letra F, que es la de la respuesta correcta...»

### EJEMPLO Y

«En el ejemplo Y hay otra serie de números... ¿Qué número debería *seguir* en esa serie?... Como van subiendo de dos en dos, debería continuar lógicamente el 10... La respuesta correcta es la A... Marcadla en la Hoja de Respuestas...»

### EJEMPLO Z

«"Ordena las palabras y contesta la frase correcta: DOS — TIERRA — TIENE — LA..." Si ordenamos bien esas cuatro palabras formarían la frase: LA TIERRA TIENE DOS... Y la respuesta correcta sería la B... Pues marcadla en la Hoja de Respuestas...»

¿Alguna duda antes de empezar? (Se contestan las posibles dudas y se completan las normas leyendo los puntos que hay debajo de los ejemplos)... Pues empezad... Ya...»



## EJEMPLOS:

**T** El *PERRO* siempre tiene:

- |               |           |                        |
|---------------|-----------|------------------------|
| A) Larga vida | C) Collar | E) Patas               |
| B) Amo        | D) Hambre | F) Instinto para cazar |

**U** ¿Qué número sigue?: 11 — 9 — 7 — 5 ...

- |      |      |      |
|------|------|------|
| A) 7 | C) 6 | E) 4 |
| B) 3 | D) 2 | F) 1 |

**V** Ordena las palabras y contesta la frase correcta: **TERMOMETRO — UN — MIDE:**

- |                |                   |                     |
|----------------|-------------------|---------------------|
| A) De mercurio | C) El médico      | E) 30 grados        |
| B) El calor    | D) La temperatura | F) De 0 a 40 grados |

**X** ¿Qué palabra *sobra*?:

**TRIANGULO — CUADRADO — ROMBO — PENTAGONO — HEXAGONO — SECTOR CIRCULAR**

- |              |              |                    |
|--------------|--------------|--------------------|
| A) Cuadrado  | C) Triángulo | E) Rombo           |
| B) Pentágono | D) Hexágono  | F) Sector circular |

**Y** ¿Qué número sigue?: 2 — 4 — 6 — 8 ...

- |       |       |       |
|-------|-------|-------|
| A) 10 | C) 11 | E) 12 |
| B) 16 | D) 9  | F) 14 |

**Z** Ordena las palabras y contesta la frase correcta: **DOS — TIERRA — TIENE — LA**

- |                    |                 |               |
|--------------------|-----------------|---------------|
| A) Polos centrales | C) Continentes  | E) Satélites  |
| B) Polos           | D) Vías lácteas | F) Atmósferas |

- Todas las PREGUNTAS TIENEN SOLO UNA RESPUESTA BUENA.
- Procura responder por orden, de manera que COINCIDA el número de la PREGUNTA con el número de la RESPUESTA. Fíjate que en las preguntas 1, 10, 20, 30 y 40 hay un signo para que te fijas si coincide en la pregunta y en la respuesta.
- Usa como **BORRADOR** para lo que necesites la parte de **ATRAS DE LA HOJA DE RESPUESTAS**.
- **TRABAJA LO MAS DEPRISA QUE PUEDES, PERO SIN EQUIVOCARTE.**
- Si **ALGUNA PREGUNTA TE RESULTA MUY DIFICIL, LA SALTAS**, sin entretenerte demasiado, y si te da tiempo al final, la repasas.
- **SI NO TERMINAS, NO TE PREOCUPES, QUE ES LO CORRIENTE.**

TIEMPO DE LA PRUEBA: 4.º E.G.B. (9 años), 20 minutos.  
5.º E.G.B. (10 años), 16 minutos.  
6.º E.G.B. (11 años), 12 minutos.

1 Ordena las palabras y contesta la frase correcta: TIENE — COCHE — EL

- |                  |                      |                        |
|------------------|----------------------|------------------------|
| A) Dos ventanass | C) Muchos kilómetros | E) Cinco plazas        |
| B) Ruedas        | D) Tejado            | F) Carrocería de acero |

2 ¿Qué palabra *sobra*? PENSAR — IMAGINAR — LEER — ESTUDIAR — REPASAR — PASEAR

- |           |            |             |
|-----------|------------|-------------|
| A) Pasear | C) Repasar | E) Imaginar |
| B) Pensar | D) Leer    | F) Estudiar |

3 ¿Qué número *sigue*? 70 — 60 — 50 — 40 ...

- |       |       |       |
|-------|-------|-------|
| A) 38 | C) 30 | E) 10 |
| B) 35 | D) 20 | F) 40 |

4 Ordena las palabras y contesta la frase correcta: SIRVEN — PARA — LIBROS — LOS

- |                |             |                    |
|----------------|-------------|--------------------|
| A) Dialogar    | C) Escribir | <del>E) Leer</del> |
| B) Pasar hojas | D) Jugar    | F) Aprobar         |

5 Sandra y Andrés son hermanos gemelos. Sumando la edad que tienen los dos juntos salen 22 años. ¿Cuántos años tiene Sandra?

- |            |            |            |
|------------|------------|------------|
| A) 11 años | C) 10 años | E) 12 años |
| B) 22 años | D) 8 años  | F) 16 años |

6 El PAN siempre se hace con:

- |                           |            |            |
|---------------------------|------------|------------|
| A) Azúcar                 | C) Cebada  | E) Piensos |
| B) Harina de algún cereal | D) Centeno | F) Maíz    |

7 Ordena las palabras y contesta la frase correcta: SILLA — USA — SE — PARA — LA

- |                 |                  |              |
|-----------------|------------------|--------------|
| A) Sentarse     | C) Adornar       | E) El sillón |
| B) En un mueble | D) Tiene 4 patas | F) Es cómoda |

8 ¿Qué palabra *sobra*? ARBOL — TOMATE — PEPINO — ZANAHORIA — LECHUGA — CEBOLLA

- |           |              |            |
|-----------|--------------|------------|
| A) Pepino | C) Tomate    | E) Lechuga |
| B) Arbol  | D) Zanahoria | F) Cebolla |

9 ¿Qué número *sigue*? 27 — 29 — 31 — 33 ...

- |       |       |       |
|-------|-------|-------|
| A) 31 | C) 37 | E) 30 |
| B) 34 | D) 35 | F) 36 |

▲ 10 La MESA siempre debe tener:

- |   |            |                  |
|---|------------|------------------|
| A) Madera                                 | C) Cajones | E) Libros encima |
| B) Algo que sirva para apoyar en el suelo | D) Hierros | F) Adornos       |

11 ¿Qué palabra *sobra*? CEREZO — GUINDO — PERAL — MATORRAL — NARANJO — MANZANO

- |            |           |             |
|------------|-----------|-------------|
| A) Cerezo  | C) Guindo | E) Naranja  |
| B) Manzano | D) Peral  | F) Matorral |

SIGUE 

**12** ¿Qué palabra *sobra*? TAZON — COMIDA — JARRA — CUCHARA — TENEDOR — VASO

- |            |            |          |
|------------|------------|----------|
| A) Tazón   | C) Comida  | E) Jarra |
| B) Cuchara | D) Tenedor | F) Vaso  |

**13** Si me encuentro en una fila de 15 personas en el último lugar, y adelanto 7 puestos, ¿en qué lugar me encontraré al final?

- |              |            |           |
|--------------|------------|-----------|
| A) Séptimo   | C) Octavo  | E) Noveno |
| B) Penúltimo | D) Primero | F) Décimo |

**14** ¿Qué número *sigue*? 4 — 13 — 4 — 12 — 4 — 11 — 4 ...

- |      |       |       |
|------|-------|-------|
| A) 1 | C) 9  | E) 10 |
| B) 4 | D) 11 | F) 13 |

**15** María tiene ocho años y su hermana Belén tres menos que ella. Su primo Roberto tiene tantos como las dos juntas. ¿Cuántos años tiene Roberto?

- |               |                    |                        |
|---------------|--------------------|------------------------|
| A) Cinco años | C) Menos que Belén | E) El doble que María  |
| B) Trece años | D) Once años       | F) El triple que Belén |

**16** Un *EDIFICIO* siempre tiene:

- |                        |             |             |
|------------------------|-------------|-------------|
| A) Portero             | C) Ascensor | E) Amplitud |
| B) Escaleras de mármol | D) Terrazas | F) Vigas    |

**17** Un *MINERAL* siempre tiene:

- |                  |                   |                |
|------------------|-------------------|----------------|
| A) Peso          | C) Hierro         | E) Gran tamaño |
| B) Tierra pegada | D) Aristas agudas | F) Sílice      |

**18** ¿Qué número *sigue*? 15 — 30 — 45 — 60 ...

- |       |       |       |
|-------|-------|-------|
| A) 95 | C) 70 | E) 85 |
| B) 65 | D) 75 | F) 55 |

**19** Una *COMPETICION* siempre tiene:

- |                        |              |                        |
|------------------------|--------------|------------------------|
| A) Competidores        | C) Problemas | E) Animación e interés |
| B) Muchos espectadores | D) Ganancias | F) Un único ganador    |

● **20** Los *LIBROS* siempre tienen:

- |                        |                      |                      |
|------------------------|----------------------|----------------------|
| A) Dibujos de personas | C) Tapas de piel     | E) Lomo              |
| B) Menos de 50 páginas | D) Cincuenta páginas | F) Más de 50 páginas |

**21** El anterior al segundo es el primero. ¿Y el anterior al último?

- |                     |                      |                 |
|---------------------|----------------------|-----------------|
| A) El antepenúltimo | C) El posterior      | E) Ninguno      |
| B) El penúltimo     | D) El último también | F) El siguiente |

**22** ¿Qué palabra *sobra*? FANTA — VINO — ANIS — LICOR — BRANDY — GINEBRA

- |            |          |           |
|------------|----------|-----------|
| A) Vino    | C) Anís  | E) Brandy |
| B) Ginebra | D) Licor | F) Fanta  |

- 23** De Madrid a Segovia hay 100 kilómetros. ¿Cuántas horas tarda en llegar a Segovia un coche que sale de Madrid a 50 kilómetros por hora?
- A) Una hora                      C) Dos horas                      E) Una hora y media  
 B) Media hora                    D) Cuatro horas                  F) Tres cuartos de hora
- 24** ¿Qué palabra sobra?: FRANCIA — ESPAÑA — ITALIA — INGLATERRA — MEJICO — PORTUGAL
- A) Portugal                      C) Inglaterra                      E) España  
 B) Méjico                        D) Italia                            F) Francia
- 25** ¿Qué palabra sobra?: ESCRIBIR — LIBRETA — BOLIGRAFO — LAPIZ — LIBRO — PAPEL
- A) Lápiz                        C) Libreta                        E) Bolígrafo  
 B) Libro                        D) Papel                        F) Escribir
- 26** ¿Qué número sigue?: 12 — 14 — 28 — 30 — 60 ...
- A) 62                            C) 64                            E) 70  
 B) 66                            D) 65                            F) 63
- 27** Un TREN siempre tiene:
- A) Un vagón para carga        C) Dos vagones                  E) Inseguridad  
 B) Vía ancha                    D) Sistema de arranque        F) Vía estrecha
- 28** ¿Qué número sigue?: 80 — 40 — 20 — 10 ...
- A) 8                            C) 2                            E) 15  
 B) 5                            D) 4                            F) 6
- 29** El humo de una locomotora se desplaza hacia el sur. ¿En qué dirección va la locomotora?
- A) Hacia el norte                C) Hacia el sur                  E) Hacia atrás  
 B) Hacia el este                D) En la dirección del viento    F) Hacia adelante
- 30** Ordena las palabras y contesta la frase correcta:  
 UNA — EL — CERRADA — DE — PORCION — ES — LAGO
- A) Océano                      C) Agua                      E) Tierra  
 B) Lago                        D) Continente                F) Mar
- 31** Luis acabó la carrera antes que Antonio y Antonio antes que Andrés, pero detrás de Gerardo. ¿Quién llegó el último?
- A) Antonio                      C) Luis                      E) Andrés  
 B) Gerardo                      D) No se sabe                F) Ninguno
- 32** ¿Cuánto es una sexta parte del doble de 12?
- A) 25                            C) 4                            E) 5  
 B) 300                        D) 10                        F) 6
- 33** Una TIENDA siempre tiene:
- A) Alimentos para vender    C) Muchos compradores    E) Electrodomésticos  
 B) Mercancías para vender    D) Buena venta                F) Tejidos y prendas de vestir

**34** Ordena las palabras y contesta la frase correcta: **DE — UNA — CONJUNTO — PALABRA — ES — UN**

- |             |                |              |
|-------------|----------------|--------------|
| A) Frases   | C) Signos      | E) Oraciones |
| B) Palabras | D) Vocabulario | F) Letras    |

**35** El *COCHE* siempre tiene:

- |                        |              |                     |
|------------------------|--------------|---------------------|
| A) Embrague automático | C) Mal trato | E) Carrocería       |
| B) Frenos eléctricos   | D) Seguridad | F) Faros antiniebla |

**36** En una estantería de cuatro estantes, en el de abajo hay diez libros; en el siguiente, el doble de libros; en el tercero, como en los dos anteriores juntos, y en el cuarto, la mitad que en el tercero. ¿Cuántos libros tiene la estantería completa?

- |              |              |              |
|--------------|--------------|--------------|
| A) 95 libros | C) 39 libros | E) 75 libros |
| B) 65 libros | D) 55 libros | F) 80 libros |

**37** Ordena las palabras y contesta la frase correcta: **SON — Y — BLANCO — PALABRAS — NEGRO**

- |                     |                      |                 |
|---------------------|----------------------|-----------------|
| A) De distinta raza | C) Contradictorias   | E) Del lenguaje |
| B) Sinónimas        | D) La noche y el día | F) Opuestas     |

**38** Un avión tarda 12 horas en hacer un recorrido. ¿Cuánto tardará en hacer ese mismo recorrido otro avión que va a mitad de velocidad?

- |             |             |             |
|-------------|-------------|-------------|
| A) 18 horas | C) 2 días   | E) 6 horas  |
| B) 24 horas | D) 12 horas | F) 22 horas |

**39** ¿Qué palabra sobra?: **OIR — TOCAR — CORRER — GUSTAR — MIRAR — OLFATEAR**

- |           |             |           |
|-----------|-------------|-----------|
| A) Correr | C) Mirar    | E) Gustar |
| B) Oír    | D) Olfatear | F) Tocar  |

**40** ¿Qué número sobra: **21 — 30 — 32 — 39 — 48 — 57**

- |       |       |       |
|-------|-------|-------|
| A) 32 | C) 39 | E) 30 |
| B) 57 | D) 21 | F) 48 |

**FIN DE LA PRUEBA**

Si te ha sobrado tiempo, repasa lo que has hecho en esta prueba

---

**NO PASES A LA PAGINA SIGUIENTE HASTA QUE TE LO INDIQUEN**

H.M.Nv.

# HABILIDAD MENTAL NO VERBAL

## ELEMENTAL

Este test consta de 40 preguntas, con seis alternativas de respuesta cada una.

### EJEMPLO V

«Observad el recuadro del ejemplo V... A la figura que está dentro de él le falta una parte. ¿Cuál de los dibujos A, B, C, D, E, o F LE COMPLETARA CORRECTAMENTE?... No puede ser el A porque tiene una zona rayada, ni el C o el F porque están en posición distinta, ni el B o D porque no falta más que la cuarta parte de un círculo... La respuesta correcta es la E... Entonces mirad en la Hoja de Respuestas, en la segunda columna, debajo de donde pone H. M. No VERBAL... Al lado del ejemplo V marcad con una raya horizontal la letra E, por ser la respuesta correcta...»

### EJEMPLO X

«Mirad el Ejemplo X... Observa los dibujos que hay en ese recuadro, cómo están ordenados SIGUIENDO UNA CIERTA LEY O REGLA... Tu tarea consiste en BUSCAR CUAL ES LA LEY O REGLA QUE ORDENA ESAS FIGURAS para después saber qué figura falta donde está la interrogación... Date cuenta que abajo se va repitiendo la misma figura que arriba, pero metida dentro de un círculo..., entonces en la zona rayada faltará la figura D... Una cruz metida dentro de un círculo... Marcad la respuesta D en la hoja de respuestas, junto al ejemplo X...»

### EJEMPLO Y

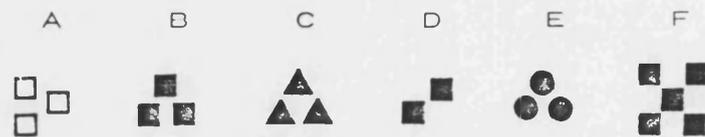
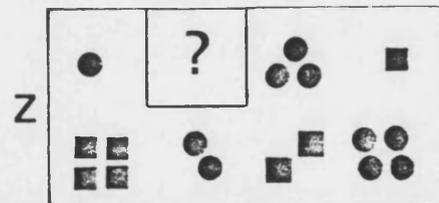
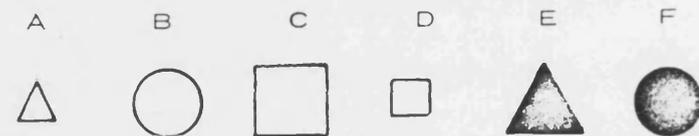
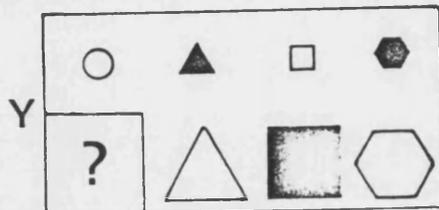
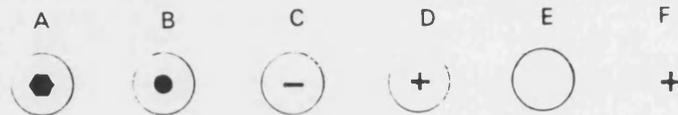
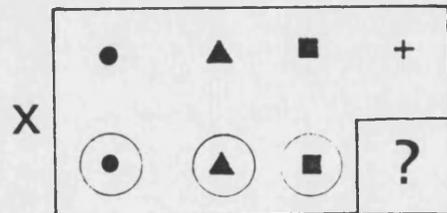
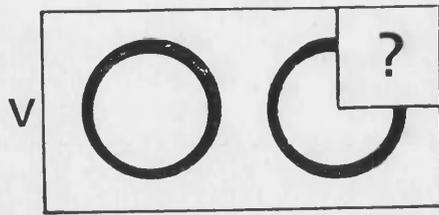
«Observad ahora el recuadro Y... Dentro de él hay triángulos, cuadrados, círculos y exágonos, blancos o negros, pequeños o grandes... Cuando la figura de abajo es blanca, la de arriba es negra, y al revés... Además, las figuras de arriba tienen la misma forma que las de abajo, pero son más pequeñas... Entonces, ¿qué figura faltará en la zona rayada?... Claro, la F, un círculo grande y negro, porque arriba hay un círculo blanco y pequeño... Marcad la F en la Hoja de Respuestas, ejemplo Y...»

### EJEMPLO Z

«Fijaos en el recuadro Z... Observa esos dibujos y estudia QUE LEY O REGLA ORDENA ESAS FIGURAS y piensa qué figura de las de la derecha debería ir donde está la interrogación... Ved que los círculos van aumentando de uno a cuatro y saltando de arriba abajo... Los cuadrados también van aumentando, pero empezando por el lado opuesto... Falta el dibujo que debería tener tres cuadros negros, es decir, la respuesta B, que es la correcta... pues marcadla en la Hoja de Respuestas...»

«¿Alguna duda antes de empezar?... (Se responde a las dudas planteadas explicando si fuera preciso otra vez los ejemplos, pero sólo una vez más... Después se leen los puntos indicados en la otra página a continuación de los ejemplos)... Pues empezad... Ya...»

**EJEMPLOS:**

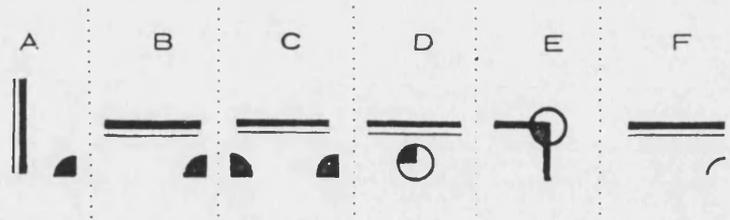
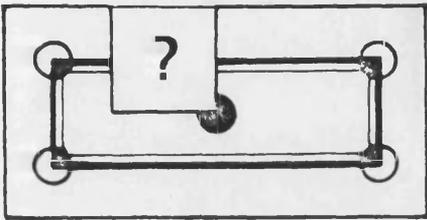


- Todas las PREGUNTAS TIENEN SOLO UNA RESPUESTA BUENA.
- Procura responder por orden, de manera que COINCIDA el número de la PREGUNTA con el número de la RESPUESTA. Fíjate que en las preguntas 1, 10, 20, 30 y 40 hay un signo para que te fijas si coincide en la pregunta y en la respuesta.
- Si ALGUNA PREGUNTA TE RESULTA MUY DIFÍCIL, LA SALTAS, sin entretenerte demasiado, y si te da tiempo al final, la repasas.
- SI NO TERMINAS, NO TE PREOCUPES, QUE ES LO CORRIENTE

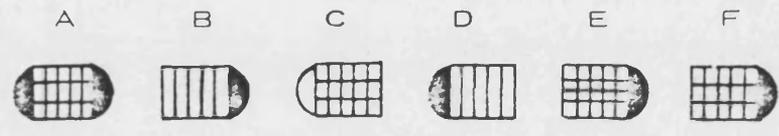
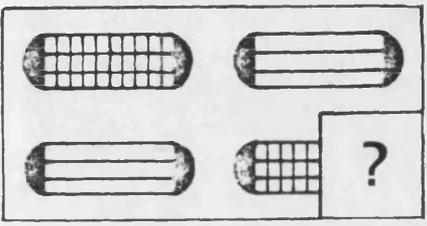
TIEMPO DE LA PRUEBA: 4.º E.G.B. (9 años), 20 minutos.  
 5.º E.G.B. (10 años), 16 minutos.  
 6.º E.G.B. (11 años), 12 minutos.

**NO PASES LA PAGINA HASTA QUE TE LO INDIQUEN**

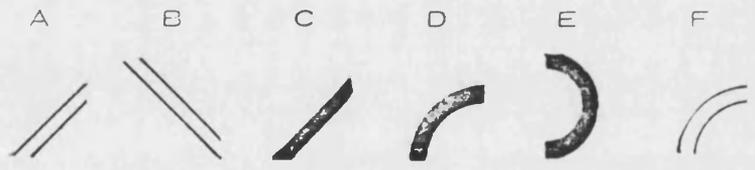
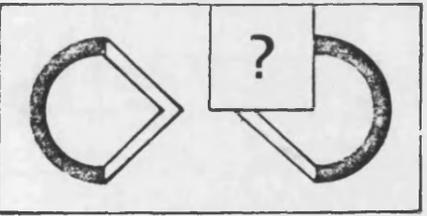
1



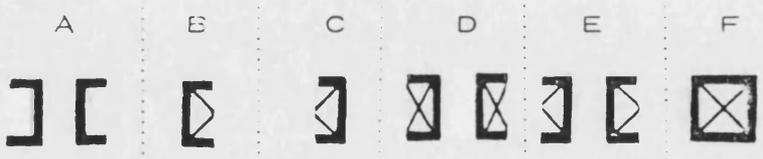
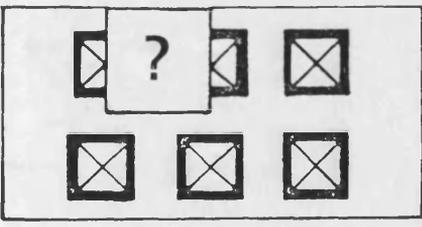
2



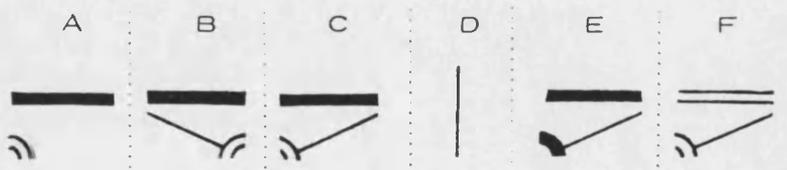
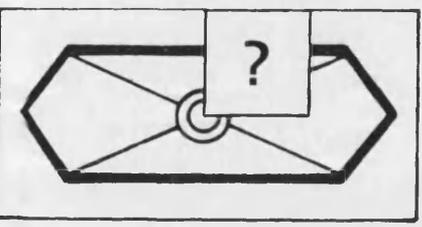
3



4

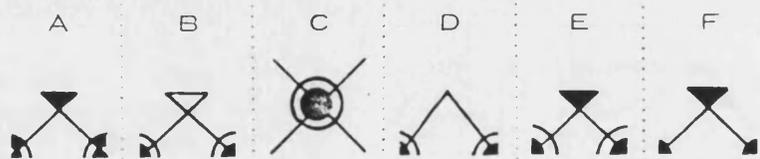
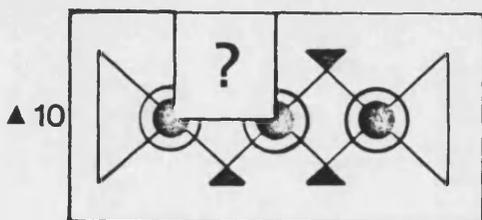
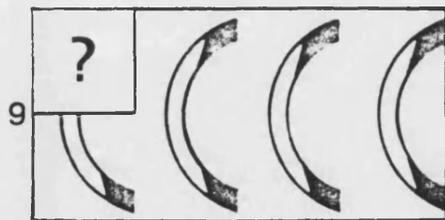
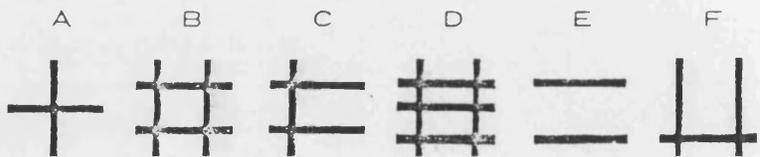
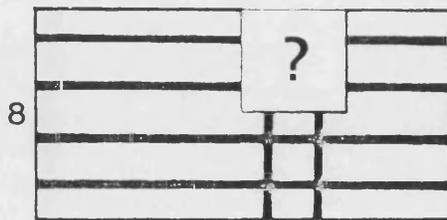
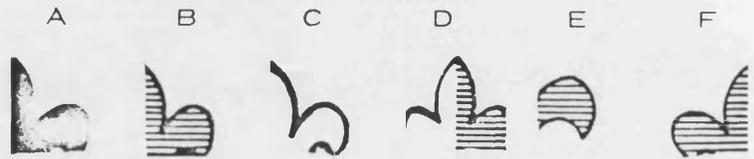
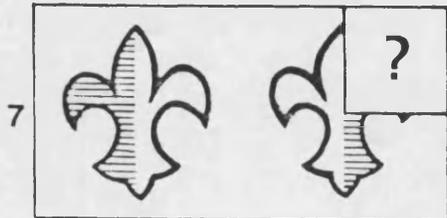
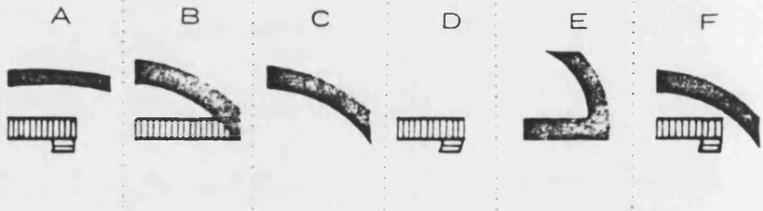
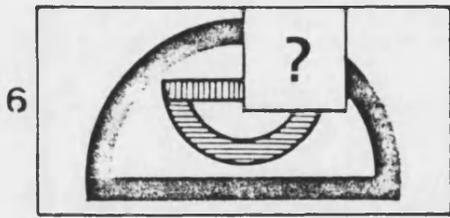


5



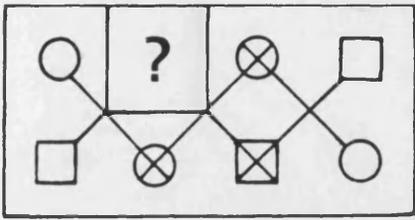
SIGUE ➡





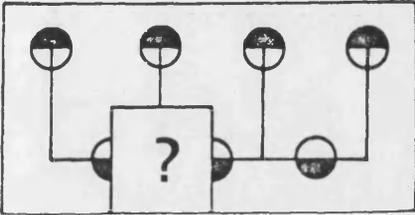
SIGUE

11



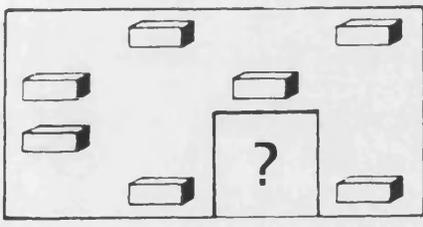
- A
- B
- C
- D
- E
- F

12



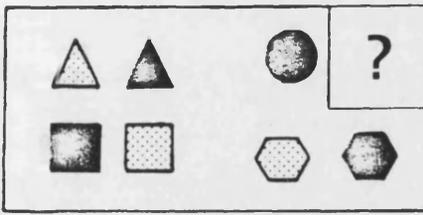
- A
- B
- C
- D
- E
- F

13



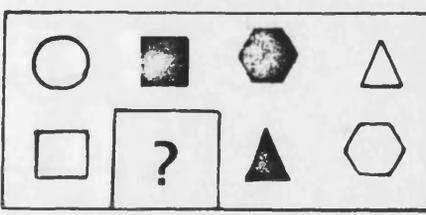
- A
- B
- C
- D
- E
- F

14



- A
- B
- C
- D
- E
- F

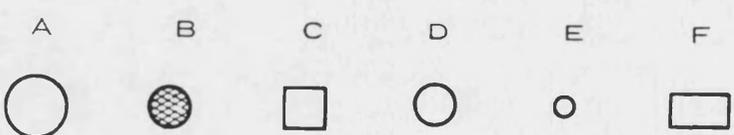
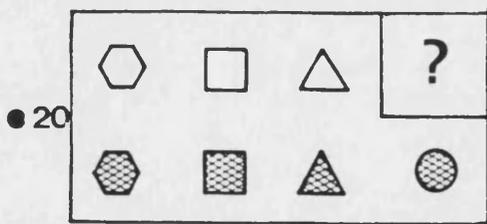
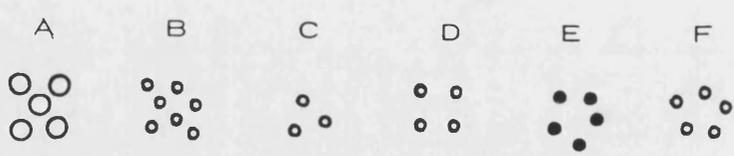
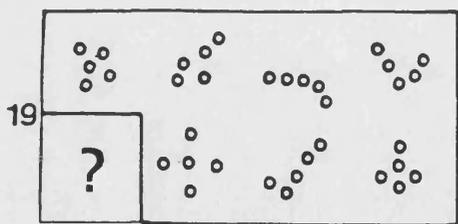
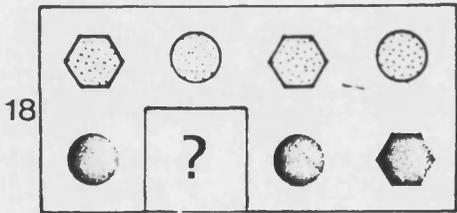
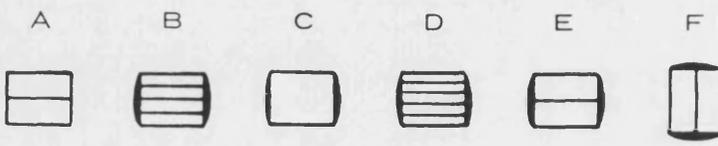
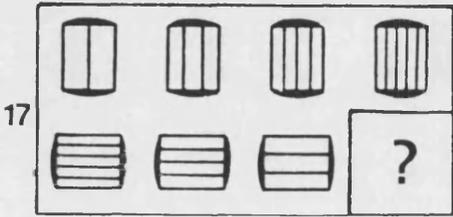
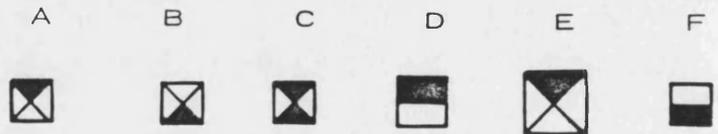
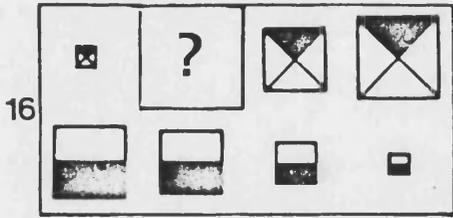
15

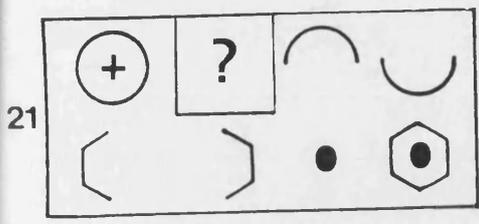


- A
- B
- C
- D
- E
- F

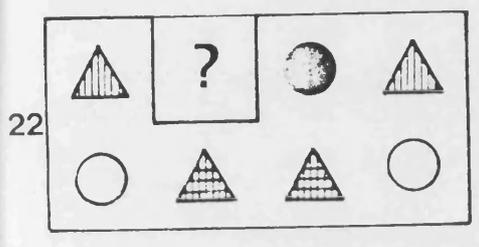
SIGUE

12

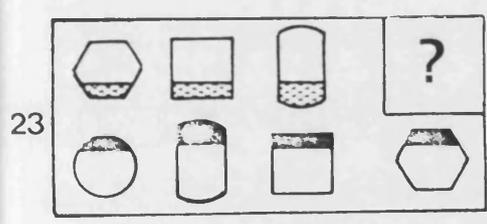




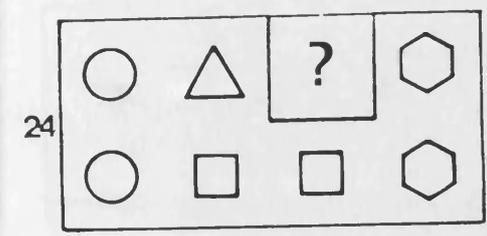
- A
- B
- C
- D
- E
- F



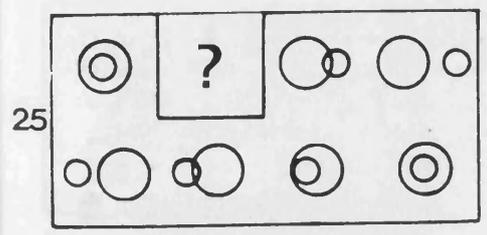
- A
- B
- C
- D
- E
- F



- A
- B
- C
- D
- E
- F

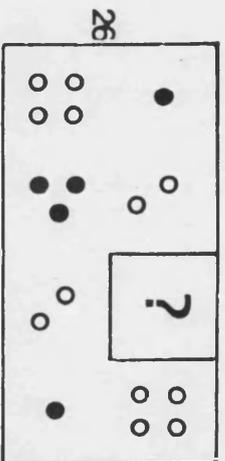


- A
- B
- C
- D
- E
- F

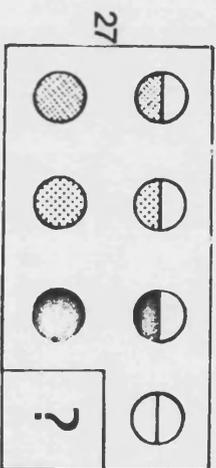


- A
- B
- C
- D
- E
- F

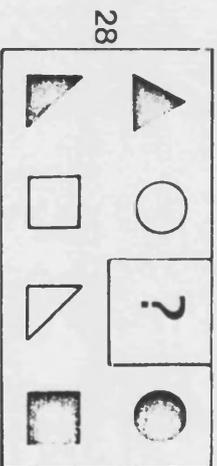
**SIGUE** ➔



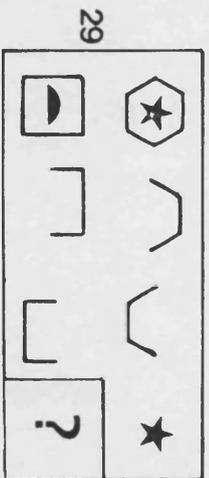
- A
- B
- C
- D
- E
- F



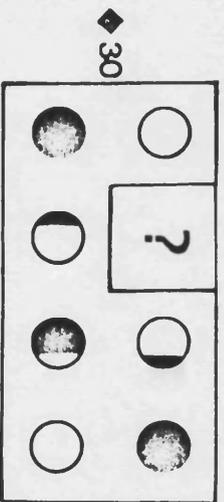
- A
- B
- C
- D
- E
- F



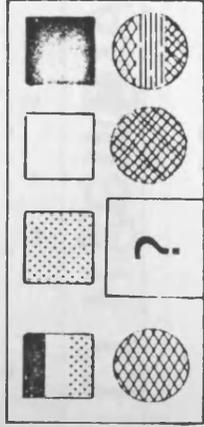
- A
- B
- C
- D
- E
- F



- A
- B
- C
- D
- E
- F

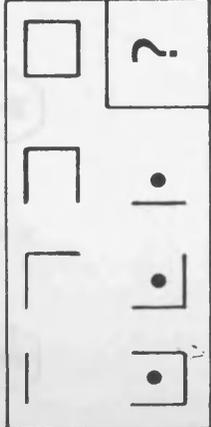


- A
- B
- C
- D
- E
- F



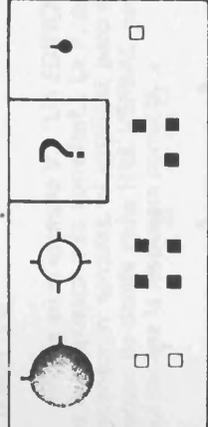
31

- A
- B
- C
- D
- E
- F



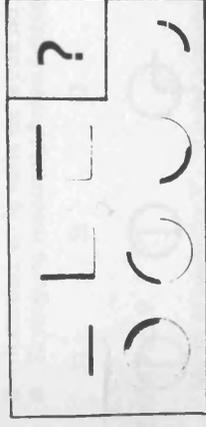
32

- A
- B
- C
- D
- E
- F



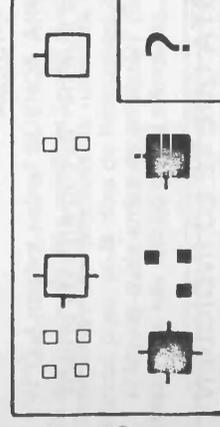
33

- A
- B
- C
- D
- E
- F



34

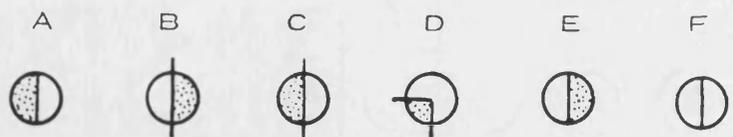
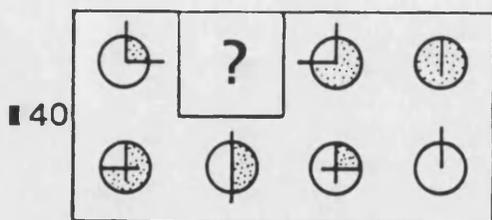
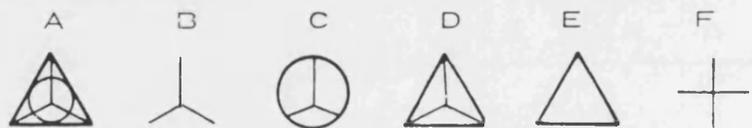
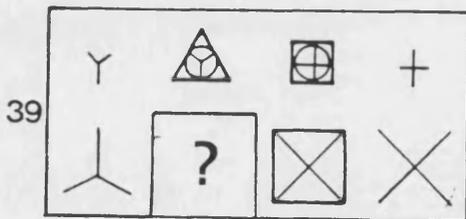
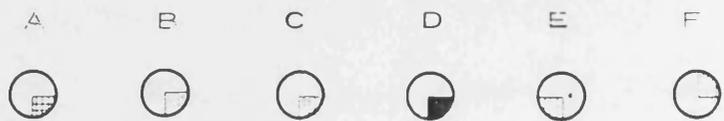
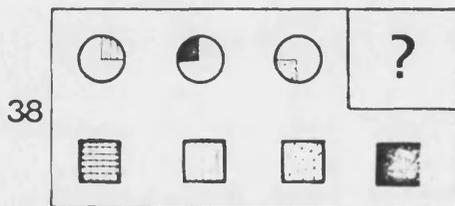
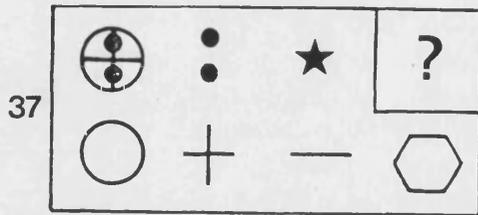
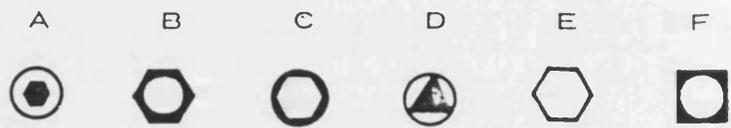
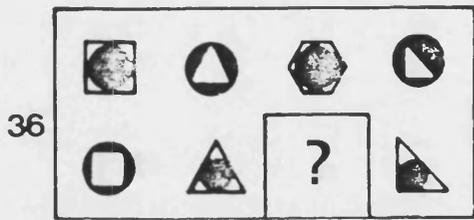
- A
- B
- C
- D
- E
- F



35

- A
- B
- C
- D
- E
- F

SIGUE



FIN DE LA PRUEBA

Si te ha sobrado tiempo, repasa lo que has hecho en esta prueba

NO PASES A LA PAGINA SIGUIENTE HASTA QUE TE LO INDIQUEN

H.M.V.]

# HABILIDAD MENTAL VERBAL

M E D I O

Este test consta de 40 preguntas con seis alternativas de respuesta cada una.

## EJEMPLO V

«Fijaos en el ejemplo V... "Un *EDIFICIO* siempre tiene:" ...Hay seis posibles respuestas, A, B, C, D, E, F... ¿Cuál es la respuesta más correcta?... La F, porque un edificio no siempre tiene portero, ni escaleras de mármol, ni ascensor, ni terrazas, ni amplitud, pero sí vigas... Luego fijaos en la Hoja de Respuestas, en la primera columna, debajo de donde pone H.M. VERBAL, junto al ejemplo V marcad con una raya el rectángulo que tiene la letra F, por ser ésa la respuesta correcta...»

## EJEMPLO X

«Observad ahora el ejemplo X... "¿Qué letra o grupo de letras *continúa* en la serie siguiente...?" Se trata de una serie de grupos de letras de las que se repite siempre la primera, es decir la A y la otra va siguiendo la serie alfabética... Luego el grupo siguiente será el AE, es decir, que la respuesta correcta es la alternativa B... Pues marcad, en la Hoja de Respuestas, la primera columna, junto al ejemplo X, la respuesta B...»

- Recordad: en la serie de letras no se tendrán en cuenta las compuestas: ch, ll, rr, w.
- Si ninguna respuesta fuese correcta debería elegirse la F (otra no indicada).

## EJEMPLO Y

Mirad el ejemplo Y... "*Ordena* estas palabras sueltas en una frase correcta y *contéstala*. CIEN - SIGLO - TIENE - EL." Tenéis cuatro palabras desordenadas: CIEN, SIGLO, TIENE, EL... Tenéis que ordenar mentalmente esas palabras sueltas para formar una frase con sentido y después ver cuál es la respuesta correcta... Bien ordenadas dirían: EL SIGLO TIENE CIEN... Entonces la respuesta correcta es la D, años... Marcad en la hoja de respuestas en el ejemplo Y, la respuesta D...»

## EJEMPLO Z

«¿"Qué palabra *sobra*? BUENOS AIRES - BOGOTA - MONTEVIDEO - PERU - CARACAS - BRASILIA..." En ese grupo de seis palabras tenéis que ver cuál es la más diferente, la más distinta, la que sobra en esa serie... Como todas son capitales de naciones, menos PERU, que es una nación, la respuesta correcta es la F, Perú... Marcad, pues, en la Hoja de Respuestas la letra F, ejemplo Z...»

«¿Alguna duda antes de empezar? (Se contestan las posibles dudas y se completan las normas leyendo los puntos que hay debajo de los ejemplos)... Pues empezad... Ya...»



**EJEMPLOS:**

**V** Un *EDIFICIO* siempre tiene:

- |             |                        |             |
|-------------|------------------------|-------------|
| A) Portero  | B) Escaleras de mármol | C) Ascensor |
| D) Terrazas | E) Amplitud            | F) Vigas    |

**X** ¿Qué *letra* o grupo de letras *continúa* en la serie siguiente?: **ab - ac - ad...**

- |       |       |                     |
|-------|-------|---------------------|
| A) af | B) ae | C) a                |
| D) e  | E) be | F) Otra no indicada |

(En las series de letras no se tienen en cuenta las compuestas ch, ll, rr, w.)  
Si ninguna respuesta fuese correcta, debería elegirse la F.)

**Y** *Ordena* estas palabras sueltas en una frase correcta y contéstala: **CIEN - SIGLO - TIENE - EL...**

- |              |          |                |
|--------------|----------|----------------|
| A) Problemas | B) Días  | C) Bisiesto    |
| D) Años      | E) Meses | F) Un centenar |

**Z** ¿Qué *palabra* *sobra* en la serie siguiente?:

**BUENOS AIRES - BOGOTA - MONTEVIDEO - PERU - CARACAS - BRASILIA**

- |                 |             |               |
|-----------------|-------------|---------------|
| A) Buenos Aires | B) Bogotá   | C) Montevideo |
| D) Caracas      | E) Brasilia | F) Perú       |

- Todas las PREGUNTAS TIENEN SOLO UNA RESPUESTA BUENA.
- Procura responder por orden, de manera que COINCIDA el número de la PREGUNTA con el número de la RESPUESTA. Fijate que en las preguntas 1, 10, 20, 30 y 40 hay un signo para que te fijes si coincide en la pregunta y en la respuesta.
- Usa como **BORRADOR** para lo que necesites la parte de ATRAS DE LA HOJA DE RESPUESTAS.
- TRABAJA LO MAS DEPRISA QUE PUEDAS, PERO SIN EQUIVOCARTE.
- Si ALGUNA PREGUNTA TE RESULTA MUY DIFICIL, LA SALTAS, sin entretenerte demasiado, y si te da tiempo al final, la repasas.
- SI NO TERMINAS, NO TE PREOCUPES, QUE ES LO CORRIENTE.

TIEMPO DE LA PRUEBA: 1.º ESO → 20 minutos
2.º ESO → 18 minutos
3.º ESO → 16 minutos

**NO PASES LA PAGINA HASTA QUE TE LO INDIQUEN**

■ 1 ¿Qué *palabra sobra* en esta serie?:

TRIANGULO - CUADRADO - ROMBO - PENTAGONO - EXAGONO - CIRCUNFERENCIA

- A) Circunferencia                      B) Exágono                      C) Pentágono  
D) Rombo                                  E) Triángulo                      F) Cuadrado

2 Mi hermano Andrés tiene 4 años más que yo y yo el doble que mi hermana Luisa. Luisa tiene 4 años. ¿Cuántos años tiene mi hermano Andrés?

- A) 10    B) 14    C) 8  
D) 12    E) 16    F) 24

3 ¿Qué *letra* o grupo de letras *continúa* en la serie siguiente?: ah - bi - cj...

- A) di    B) dk    C) fk  
D) kd    E) d    F) Otra no indicada

4 Si un caballo tarda 2 horas en recorrer 20 kms, ¿cuántas horas tardará en recorrer esa misma distancia otro caballo que corre al doble de velocidad?

- A) 3 horas                                  B) Media hora                                  C) 4 horas  
D) 45 minutos                                  E) 2 horas                                  F) 1 hora

5 ¿Qué *palabra sobra* en la serie siguiente?:

AVISPA - ARAÑA - MOSCA - LAGARTO - GUSANO - ABEJA

- A) Gusano                                  B) Araña                                  C) Avispa  
D) Mosca                                  E) Lagarto                                  F) Abeja

6 *Ordena* estas palabras sueltas en una frase correcta y contéstala:

FRANCIA - NACIONES - DEL - Y - CONTINENTE - ESPAÑA - SON.

- A) Distinta lengua                      B) Colindantes                      C) Europeo  
D) Vecinos                                  E) Pirenaicas                      F) De lengua romance

7 ¿Qué *palabra sobra* en la serie siguiente?:

PENSAR - QUERER - ANDAR - IMAGINAR - RAZONAR - REFLEXIONAR

- A) Pensar                                  B) Querer                                  C) Andar  
D) Razonar                                  E) Imaginar                                  F) Reflexionar

8 ¿Qué *letra* o grupo de letras *continúa* en la serie siguiente?: hpls - ah - bl - cp...

- A) sd    B) dp    C) es  
D) cs    E) ds    F) Otra no indicada

9 Juan tiene 4 años, Pedro 2 más que Juan, y Ana como Juan y Pedro juntos. ¿Cuántos años tiene Ana?

- A) 12 años                                  B) 6 años                                  C) 8 años  
D) 10 años                                  E) 14 años                                  F) 4 años

▲ 10 ¿Qué *letra* o grupo de letras *continúa* en la serie siguiente?: ab - b - cd - d - ef...

- A) g    B) h    C) e  
D) f    E) fg    F) Otra no indicada

11 ¿Qué *palabra sobra* en esta serie?: Tenedor - PASTELES - UVAS - ARROZ - SALSA - ENTREMESES

- A) Tenedor                                  B) Pasteles                                  C) Uvas  
D) Arroz    E) Entremeses                                  F) Salsa

**SIGUE** 

12 Si con un duro compro 2 cromos, ¿cuántos compraré con 10 duros?

- |       |       |       |
|-------|-------|-------|
| A) 40 | B) 2  | C) 15 |
| D) 20 | E) 10 | F) 5  |

13 ¿Qué *palabra sobra* en esta serie?: **SALIR - OIR - TOCAR - GUSTAR - MIRAR - OLFATEAR**

- |             |           |          |
|-------------|-----------|----------|
| A) Olfatear | B) Oír    | C) Tocar |
| D) Salir    | E) Gustar | F) Mirar |

14 El *RUIDO* siempre produce...

- |                 |                        |                     |
|-----------------|------------------------|---------------------|
| A) Temor        | B) Dolor de cabeza     | C) Roturas          |
| D) Electricidad | E) Una onda vibratoria | F) Fuerte conmoción |

15 Si una persona recorre 4 kilómetros en una hora, ¿cuántas horas tardará en recorrer los 4 kms otra persona que va a la mitad de su velocidad?

- |            |            |               |
|------------|------------|---------------|
| A) 3 horas | B) 2 horas | C) Media hora |
| D) 4 horas | E) 1 hora  | F) 35 minutos |

16 Un *SER VIVO* siempre tiene...

- |                 |                          |                            |
|-----------------|--------------------------|----------------------------|
| A) Visión       | B) Cabeza, tronco y cola | C) Colores vivos           |
| D) Buen aspecto | E) Existencia            | F) Extremidades superiores |

17 ¿Qué *letra* o grupo de letras *continúa* en la serie siguiente?: **b - e - cd - f - i...**

- |       |       |                     |
|-------|-------|---------------------|
| A) gh | B) f  | C) ih               |
| D) hj | E) gj | F) Otra no indicada |

18 Un *CABALLO* siempre tiene...

- |                 |               |           |
|-----------------|---------------|-----------|
| A) Montura      | B) Herraduras | C) Pelo   |
| D) Gran reprise | E) Esbeltez   | F) Rendas |

19 ¿Qué número dividido entre 5 nos dará lo mismo que la mitad de 8 por 2?

- |       |       |       |
|-------|-------|-------|
| A) 20 | B) 40 | C) 30 |
| D) 5  | E) 4  | F) 10 |

● 20 ¿Qué *letra continúa* en la serie siguiente?: **a - c - b - d - f...**

- |      |      |                     |
|------|------|---------------------|
| A) g | B) h | C) i                |
| D) f | E) j | F) Otra no indicada |

21 ¿Qué *letra* o grupo de letras *continúa* en la serie siguiente?: **ck - dj - ei...**

- |       |       |                     |
|-------|-------|---------------------|
| A) f  | B) hi | C) gh               |
| D) fg | E) hj | F) Otra no indicada |

22 ¿Qué *palabra sobra* en la serie siguiente?:

**TRIANGULO - CIRCULO - CILINDRO - CURVA - ESFERA - CIRCUNFERENCIA**

- |             |              |                   |
|-------------|--------------|-------------------|
| A) Círculo  | B) Esfera    | C) Circunferencia |
| D) Cilindro | E) Triángulo | F) Curva          |

**SIGUE** 

23 ¿Qué número es la mitad de la cuarta parte de 200?

- A) 80                                      B) 25                                      C) 24  
D) 50                                      E) 100                                    F) 12 y medio

24 ¿Qué letra o grupo de letras continúa en la serie siguiente?: a - b - d - g...

- A) kj                                        B) h                                        C) k  
D) i                                        E) l                                        F) Otra no indicada

25 ¿Qué palabra sobra en esta serie?:

**FELINOS - COLEOPTEROS - BALLENAS - SAURIOS - OFIDIOS - ARACNIDOS**

- A) Arácnidos                              B) Ballenas                              C) Felinos  
D) Coleópteros                            E) Ofidios                                F) Saurios

26 Ordena estas palabras sueltas en una frase correcta y contéstala: VIVIO - EN - VEINTE - SIGLO - EL

- A) Hitler                                    B) Napoleón                              C) Cervantes  
D) Descartes                              E) Julio César                            F) Homero

27 El GRANIZO siempre tiene que ir acompañado de...

- A) Rayos y truenos                      B) Fuertes vientos                      C) Altas temperaturas  
D) Nubes rojas                            E) Tormentas eléctricas              F) Bajas temperaturas

28 ¿Cuánto es un tercio del cuádruple de 30?

- A) 60                                        B) 30                                        C) 40  
D) 9                                         E) 80                                        F) 100

29 Ordena estas palabras sueltas en una frase correcta y contéstala: SALE - POR - SOL - EL

- A) El mar                                    B) El norte                                C) La tierra  
D) El oeste                                E) El este                                 F) Las montañas

◆ 30 ¿Qué número si le multiplicamos por 5 nos dará lo mismo que la mitad de la tercera parte de 120?

- A) 4                                         B) 6                                         C) 5  
D) 2                                         E) 8                                         F) 10

31 Un nadador tarda 30 segundos en recorrer una distancia. ¿Cuánto tardará otro nadador que va a doble velocidad, en recorrer una distancia la mitad de larga?

- A) 12 segundos                            B) 7 segundos y medio                C) 15 segundos  
D) 90 segundos                            E) 10 segundos                            F) 5 segundos

32 Con 15 duros, ¿cuántas hojas de papel de 0,10 ptas. cada una puedo comprar?

- A) 800                                        B) 15                                        C) 750  
D) 1.500                                    E) 380                                      F) 150

33 ¿Qué letra o grupo de letras continúa en la serie siguiente?: c - b - a - f - e...

- A) c                                         B) b                                        C) f  
D) g                                         E) ed                                        F) Otra no indicada

**SIGUE** 



# HABILIDAD MENTAL NO VERBAL

## M E D I O

Este test consta de 40 preguntas con seis alternativas de respuesta cada una.

### EJEMPLO V

«Observad el recuadro del ejemplo V... Los dibujos de este recuadro están ORDENADOS SIGUIENDO UNA CIERTA LEY O REGLA... Tu tarea consiste en BUSCAR CUAL ES LA LEY O REGLA QUE ORDENA ESAS FIGURAS para después saber qué figura falta en la zona con una interrogación... Date cuenta que todas las figuras de abajo son rayadas, y arriba se repite la misma figura pero en blanco... Luego la respuesta correcta será la E..., falta un círculo blanco en esa zona... Ahora mirad en la Hoja de Respuestas, en la segunda columna, debajo de donde pone H.M.No VERBAL... Al lado del ejemplo V marcad con una raya horizontal la letra E, por ser la respuesta correcta...»

### EJEMPLO X

«Mirad el ejemplo X... Busca la LEY O REGLA QUE ORDENA ESAS FIGURAS, para deducir cuál falta en esa zona donde está la interrogación... Vemos que en todos los cuadros hay tres figuras siempre distintas... Luego la respuesta correcta será la D, porque tiene un grupo de tres figuras distintas a las demás que ya están en el recuadro... Marca la letra D en el ejemplo X de la Hoja de Respuestas...»

### EJEMPLO Y

«Observad ahora el recuadro Y... Pensad un poco cuál puede ser la respuesta correcta... Es la B, porque vemos que el círculo pequeño va entrando en el grande por el lado opuesto a la serie de abajo, quedando tangente por dentro a la circunferencia grande... Marcad la respuesta B en la Hoja de Respuestas, ejemplo Y...»

### EJEMPLO Z

«Mirad el recuadro Z... Observad esos dibujos... Veréis que el exágono de la zona de arriba va perdiendo cada vez el lado más marcado, y abajo van apareciendo los lados que arriba pierden. Luego la respuesta correcta será la F... Marcadla en la Hoja de Respuestas...»

«¿Alguna duda antes de empezar?... (Se responde a las dudas planteadas explicando si fuera preciso otra vez los ejemplos, pero sólo una vez más... Después se leen los puntos indicados en la otra página a continuación de los ejemplos)... Pues empezad... Ya...»

**EJEMPLOS:**

**V**

			?

A B C D E F

**X**

?			

A B C D E F

**Y**

	?		

A B C D E F

**Z**

			?

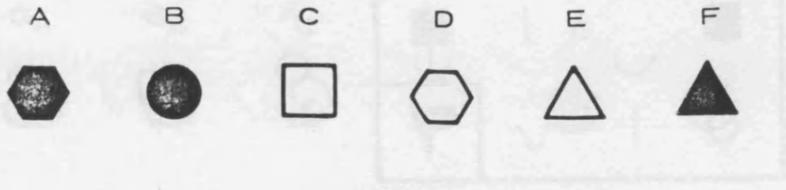
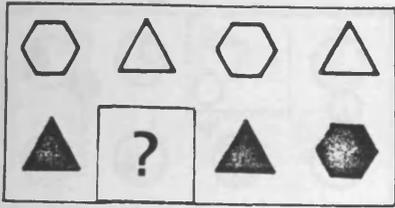
A B C D E F

- Todas las PREGUNTAS TIENEN SOLO UNA RESPUESTA BUENA.
- Procura responder por orden, de manera que COINCIDA el número de la PREGUNTA con el número de la RESPUESTA. Fíjate que en las preguntas 1, 10, 20, 30 y 40 hay un signo para que te fijes si coincide en la pregunta y en la respuesta.
- Si ALGUNA PREGUNTA TE RESULTA MUY DIFÍCIL, LA SALTAS, sin entretenerte demasiado, y si te da tiempo al final, la repasas.
- SI NO TERMINAS, NO TE PREOCUPES, QUE ES LO CORRIENTE.

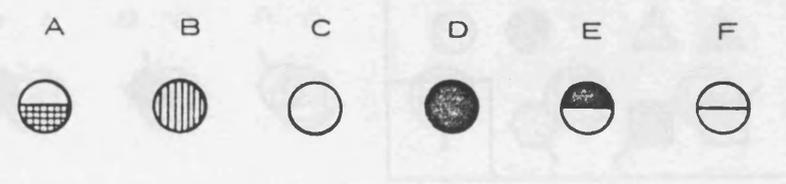
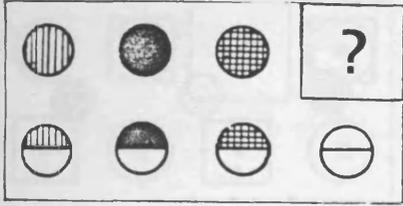
TIEMPO DE LA PRUEBA: 1.º ESO → 22 minutos  
 2.º ESO → 20 minutos  
 3.º ESO → 18 minutos

**NO PASES LA PAGINA HASTA QUE TE LO INDIQUEN**

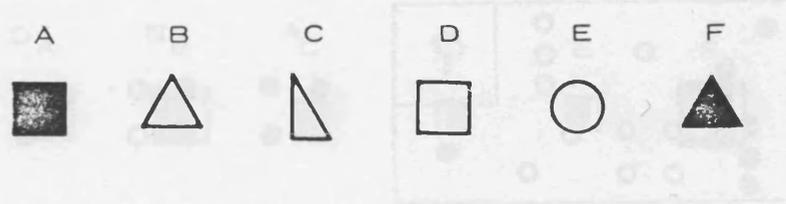
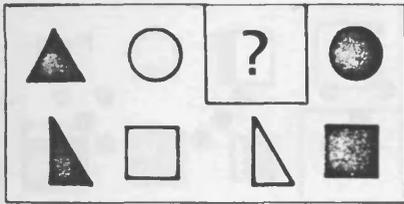
1



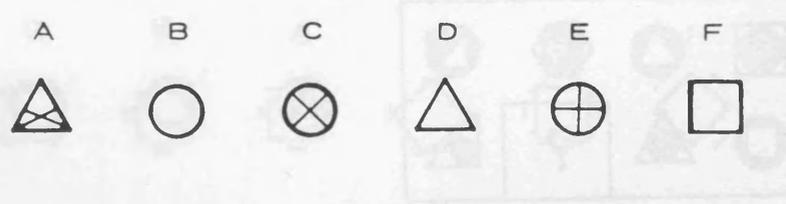
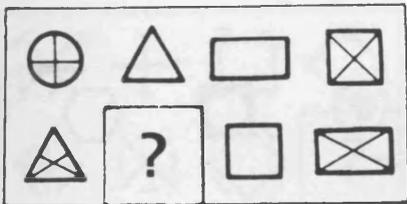
2



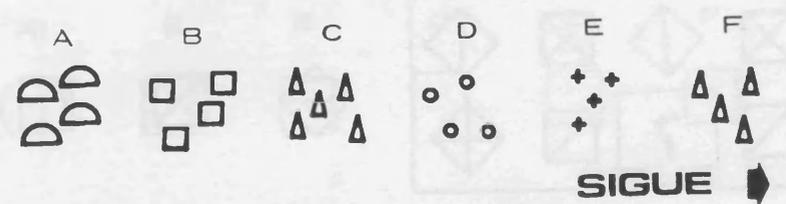
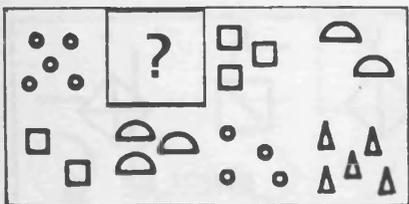
3



4

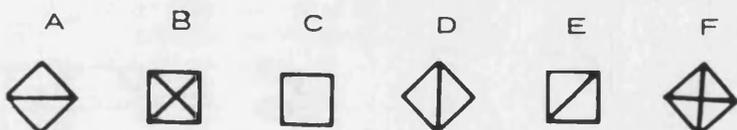
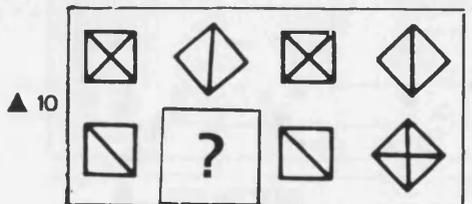
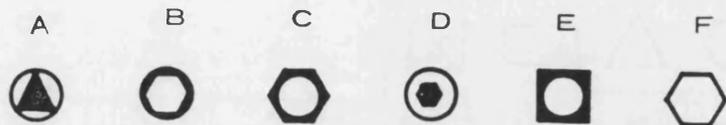
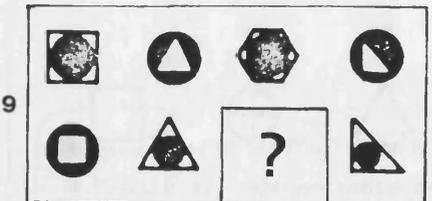
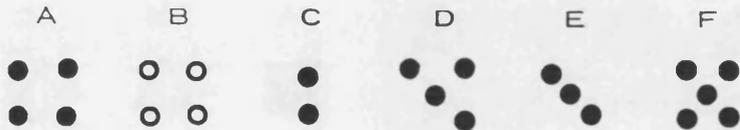
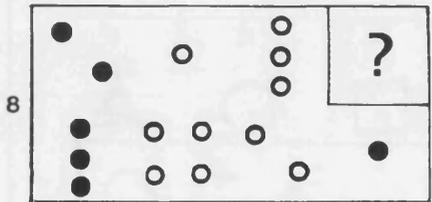
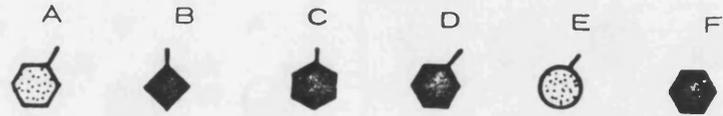
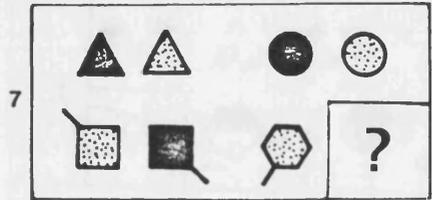
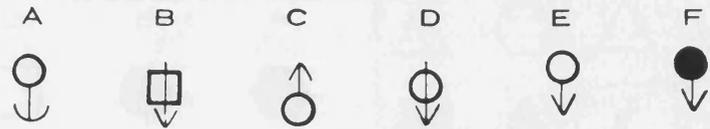
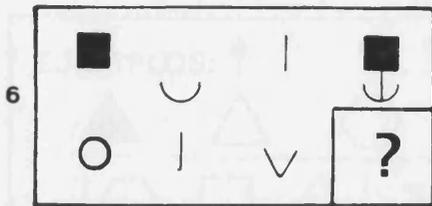


5



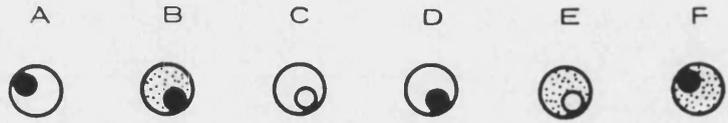
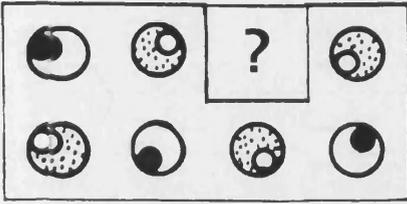
SIGUE ➡



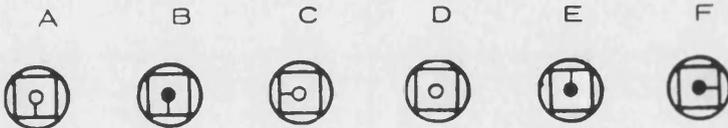
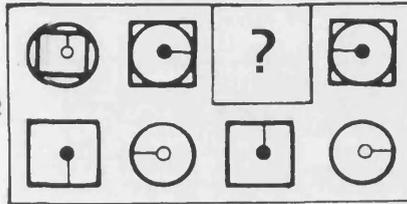


SIGUE

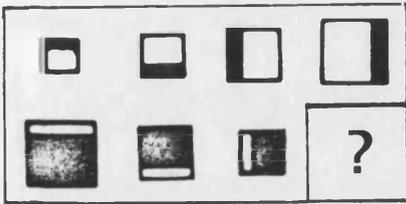
11



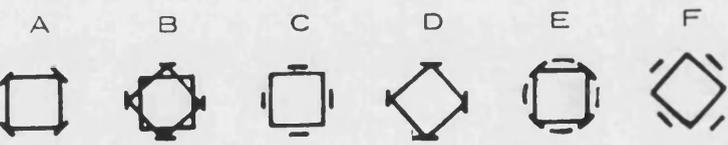
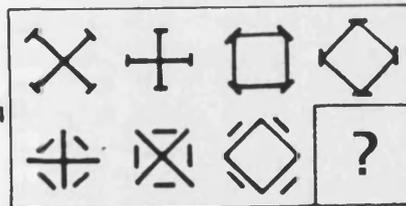
12



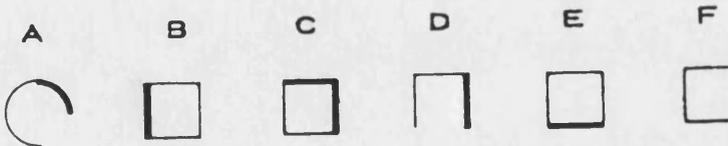
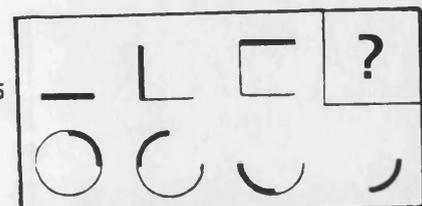
13



14



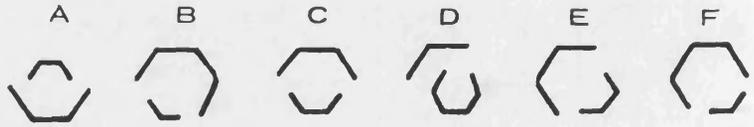
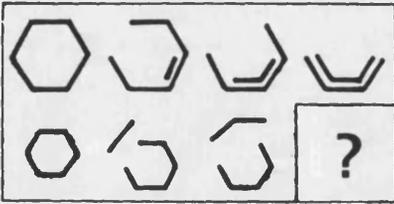
15



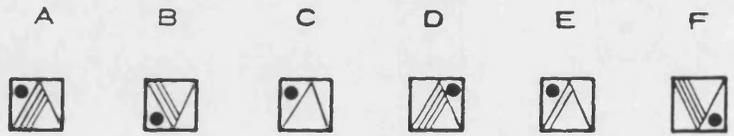
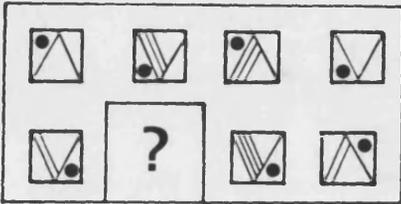
12

SIGUE ➡

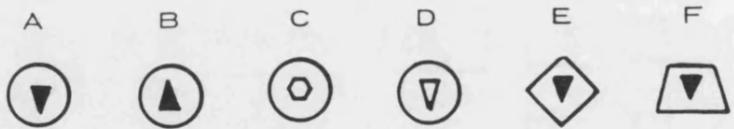
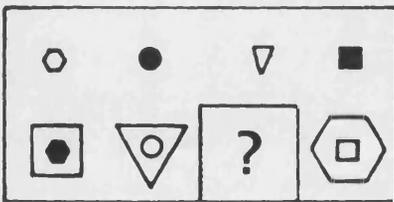
16



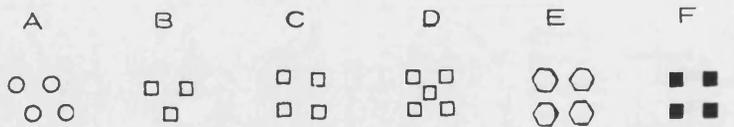
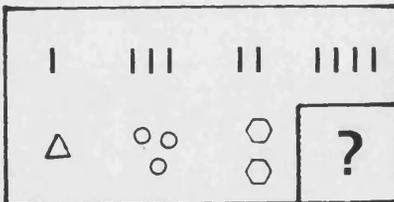
17



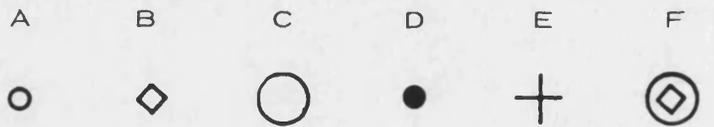
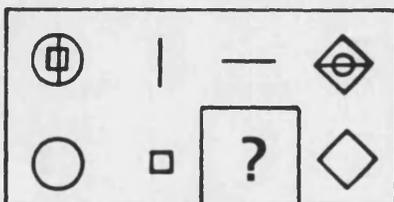
18



19

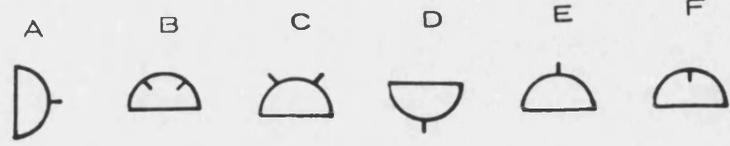
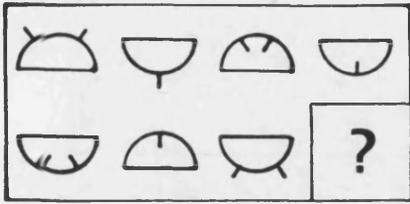


20

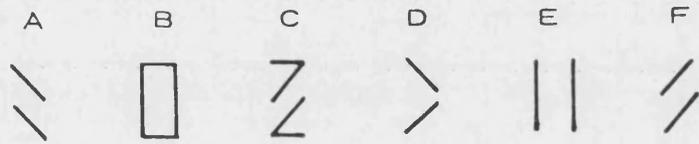
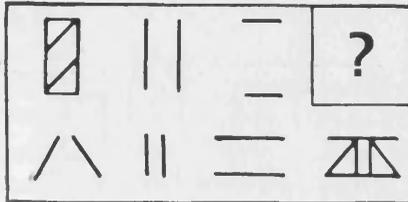


SIGUE

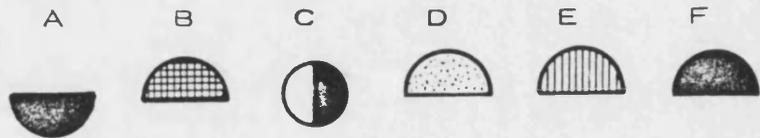
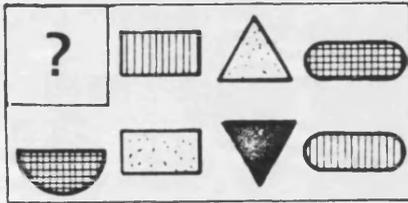
21



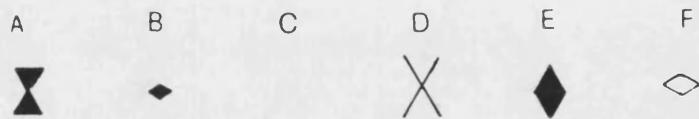
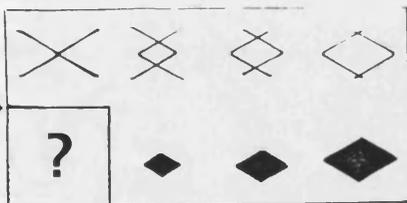
22



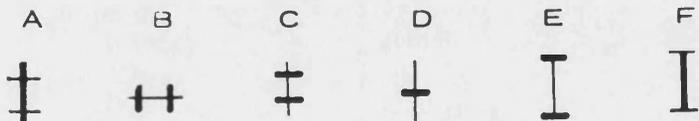
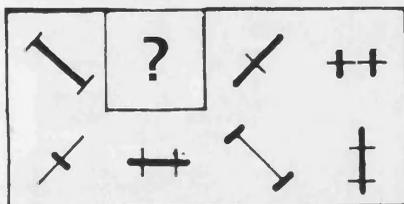
23



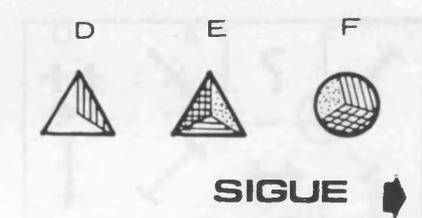
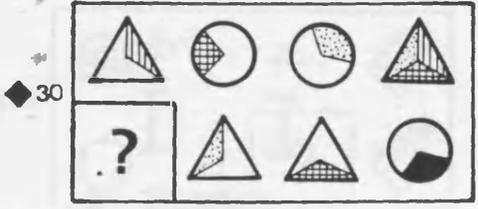
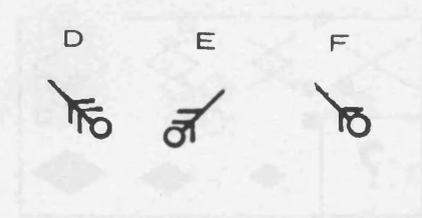
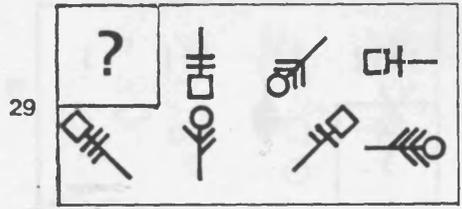
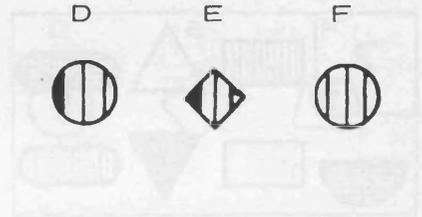
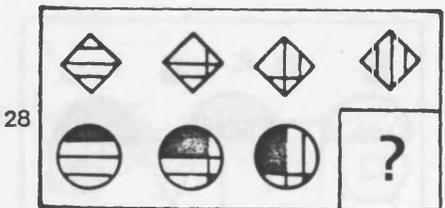
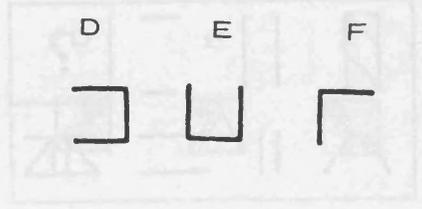
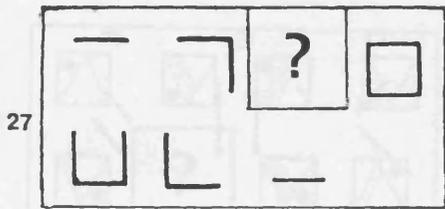
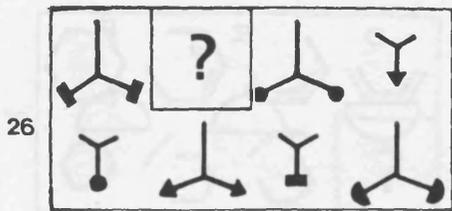
24



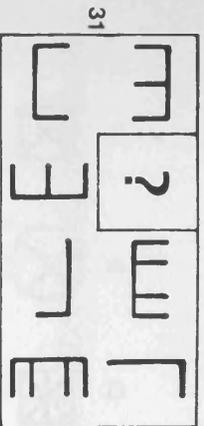
25



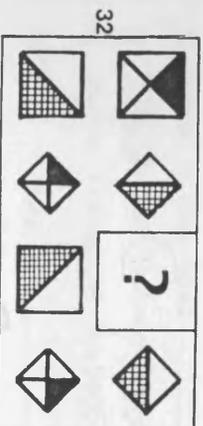
SIGUE ➡



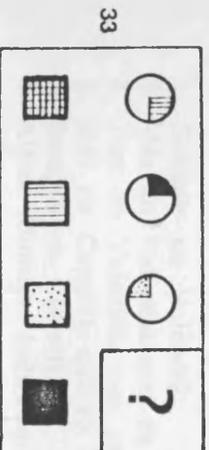
SIGUE



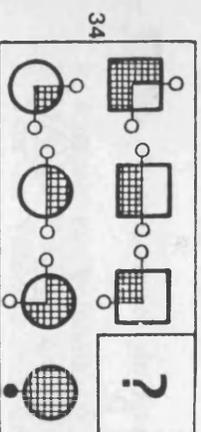
- A
- B
- C
- D
- E
- F



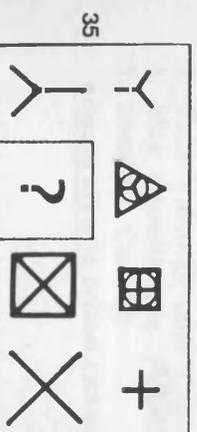
- A
- B
- C
- D
- E
- F



- A
- B
- C
- D
- E
- F

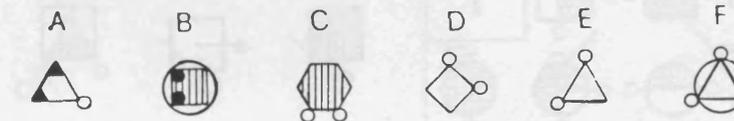
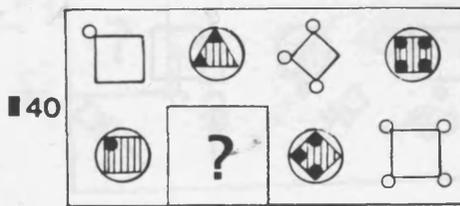
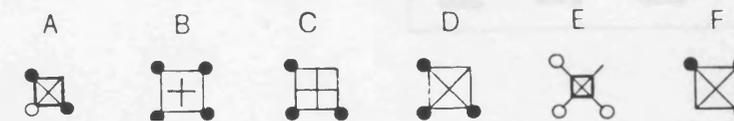
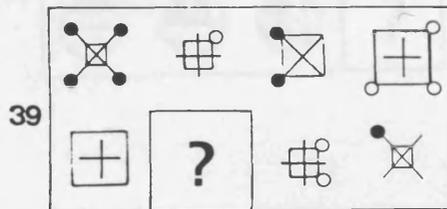
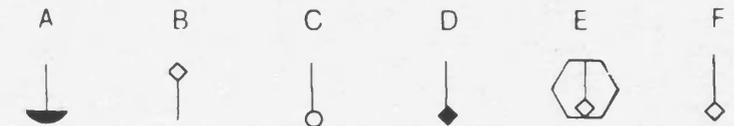
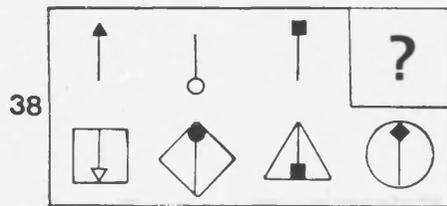
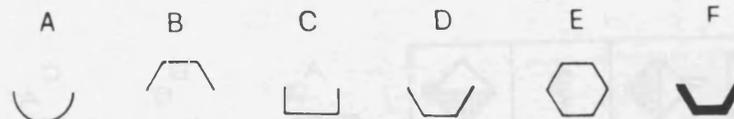
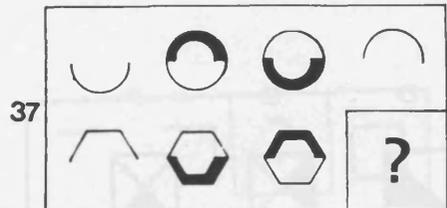
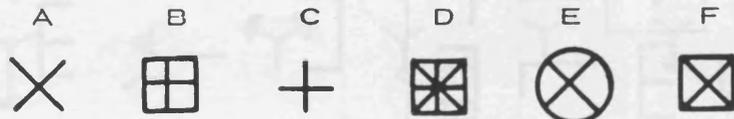
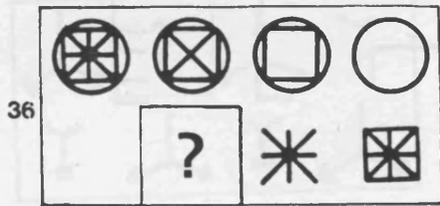


- A
- B
- C
- D
- E
- F



- A
- B
- C
- D
- E
- F

SIGUE



FIN DE LA PRUEBA

Si te ha sobrado tiempo, repasa lo que has hecho en esta prueba

NO PASES A LA PAGINA SIGUIENTE HASTA QUE TE LO INDIQUEN

ENQUESTA SOCIO-LINGÜÍSTICA.

COGNOMS \_\_\_\_\_ NOM \_\_\_\_\_  
COL·LEGI \_\_\_\_\_ CURS \_\_\_\_\_  
POBLACIO \_\_\_\_\_

1. Indica el teu sexe: 1. Home. 2. Dona.
2. Indica la teua edad: ..... anys.
3. Indica el teu lloc de naixement: .....
4. Indica l'ocupació (treball) dels teus pares:  
Pare: .....  
Mare: .....

A continuació, contesta les següents preguntes marcant amb una creu la resposta correcta.

5. Els que viviu a casa teua parleu:
  1. Sempre en Valencià.
  2. Més en Valencià que en Castellà.
  3. Tant en Valencià com en Castellà.
  4. Més en Castellà que en Valencià.
  5. Sempre en Castellà.
  6. Altres idiomes: Indica quin .....
6. Amb els companys al pati parles:
  1. Sempre en Valencià.
  2. Més en Valencià que en Castellà.
  3. Tant en Valencià com en Castellà.
  4. Més en Castellà que en Valencià.
  5. Sempre en Castellà.
7. Amb els teus amics de fora de l'escola parles:
  1. Sempre en Valencià.
  2. Més en Valencià que en Castellà.
  3. Tant en Valencià com en Castellà.
  4. Més en Castellà que en Valencià.
  5. Sempre en Castellà.
8. Quan vec la televisió, l'idioma que escolte és:
  1. Sempre Valencià.
  2. Més Valencià que Castellà.
  3. Tant Valencià com Castellà.
  4. Més Castella que Valencià.
  5. Sempre Castellà.



ENCUESTA SOCIOLINGÜÍSTICA.

APELLIDOS \_\_\_\_\_ NOMBRE \_\_\_\_\_  
COLEGIO \_\_\_\_\_ CURSO \_\_\_\_\_  
POBLACION \_\_\_\_\_

1. Indica tu sexo:            1. Varón.            2. Mujer.
2. Indica tu edad: ..... años.
3. Indica tu lugar de nacimiento: .....
4. Indica la profesión de tus padres:  
Padre: .....  
Madre: .....

A continuación, contesta las siguientes preguntas marcando con una cruz la respuesta correcta.

5. En tu casa, ¿en qué idioma habláis?
  1. Siempre en Valenciano.
  2. Más en Valenciano que en Castellano.
  3. Tanto en Valenciano como en Castellano.
  4. Más en Castellano que en Valenciano.
  5. Siempre en Castellano.
  6. En otros idiomas. Di cuál.....
6. Con tus compañeros en el patio, ¿en qué idioma hablas?
  1. Siempre en Valenciano.
  2. Más en Valenciano que en Castellano.
  3. Tanto en Valenciano como en Castellano.
  4. Más en Castellano que en Valenciano.
  5. Siempre en Castellano.
7. Con tus amigos fuera de la escuela, ¿en qué idioma hablas?
  1. Siempre en Valenciano.
  2. Más en Valenciano que en Castellano.
  3. Tanto en Valenciano como en Castellano.
  4. Más en Castellano que en Valenciano.
  5. Siempre en Castellano.
8. Cuando veo la televisión, el idioma que escucho es:
  1. Siempre Valenciano
  2. Más Valenciano que Castellano
  3. Tanto Valenciano como Castellano
  4. Más Castellano que Valenciano
  5. Siempre en Castellano