

## VAVEL inicial: evidencias de validez<sup>1</sup>

Antonio Ferrer, Vicenta Ávila, Mario Brancal, Inés Tomás y David Carreres

Grupo COMULEN (COMUNicación y LENGuaje)- Universitat de València-Estudi General

En la construcción de pruebas psicométricas destinadas a la medición de aptitudes resulta ineludible contar con un apartado dedicado a explorar su validez. En este trabajo se presentan evidencias de validez relacionadas con VAVEL Inicial (Ferrer *et al.*, 2006), prueba creada por nuestro Grupo de Investigación que explora la comprensión de vocabulario en niños de 6 a 9 años. Con un grupo de estudio de 89 niños y niñas, siguiendo las directrices de la última versión de las normas de la American Psychological Association (AERA, APA, NCME, 1999), se proporcionan, desde una primera aproximación, datos sobre *evidencia convergente*, estableciendo la relación entre pruebas que, a priori, están destinadas a medir el mismo constructo (Vavel Inicial-Test de Vocabulario en Imágenes Peabody - Adaptación Española; Dunn, 1981). En una segunda aproximación, centrada en *relaciones test-criterio*, se aportan datos relativos a la competencia léxica mostrada en VAVEL Inicial según el rendimiento general de los niños establecido a través del juicio subjetivo de los tutores de aula.

### Introducción

Con las pruebas de Valoración de Vocabulario Español destinadas a población de entre 3-6 años (VAVEL Infantil: Brancal *et al.*, 2005) y 6-9 años (VAVEL Inicial: Ferrer *et al.*, 2006a) nuestro grupo continúa con el trabajo iniciado tiempo atrás (Brancal *et al.*, 1998) dirigido a construir herramientas psicométricas útiles tanto en el diagnóstico como en el seguimiento del tratamiento logopédico llevado a cabo sobre población en edad escolar. Al elaborar estas herramientas siempre se ha tratado de contar con fuentes de evidencia que constaten su estatus como pruebas válidas para conseguir el objetivo que se proponen; en el caso de las pruebas VAVEL, la evaluación de la comprensión de vocabulario.

---

<sup>1</sup> Este trabajo ha sido realizado en el marco del proyecto “Elaboración de una prueba de evaluación de repertorio léxico para población escolar de 3 a 12 años” (UV-AE-20050973) financiado por la Universitat de València – Estudi General.

En este trabajo se resumen los hallazgos más significativos relacionados con el análisis de evidencias de validez de la prueba VAVEL Inicial destinada a evaluar la competencia en comprensión de léxico en niños de entre 6-9 años. Dicha prueba consta de 70 ítems más tres de demostración, cada uno de los cuales contiene 4 dibujos, siendo la tarea propuesta la de “Señala X” y utilizando como reactivos sustantivos, verbos o adjetivos. Véase en el ejemplo de la Imagen 1 el ítem 56 en el que se solicita la identificación de “soga”.

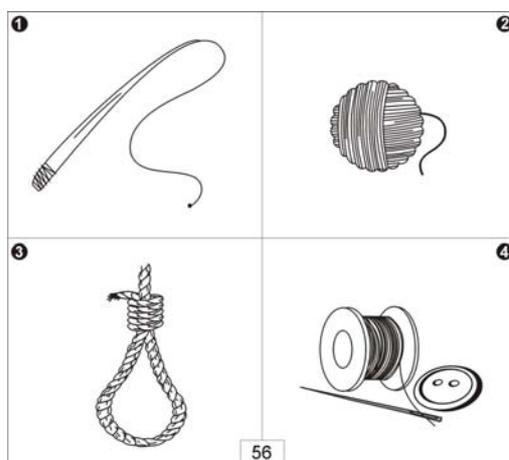


Imagen 1. Ejemplo de elemento de Vavel Inicial: Item 56 “Soga”

Siguiendo las directrices de la última versión de las normas de la *American Psychological Association* (AERA, APA, NCME, 1999) se ha explorado diversas fuentes de evidencia de validez de las que aquí se tratan dos en particular. La primera de ellas intenta establecer evidencia convergente a través del enfoque correlacional, analizando las relaciones entre las puntuaciones del test y la medida de otro que presumiblemente mide el mismo constructo, en este caso el Test de Vocabulario en Imágenes Peabody (TVIP en adelante) (Dunn, 1981). Este tipo de estudios incrementan nuestro conocimiento acerca de cómo se comportan de forma comparada pruebas que miden a través de idéntica tarea la misma competencia psicolingüística. Resulta difícil encontrar investigaciones de esta naturaleza en instrumentos destinados a población hispanoparlante dada la escasez de pruebas existentes en nuestra lengua que puedan entrar en competencia, lo cual contrasta con la situación vivida en idiomas como el inglés donde la variedad de tests permite realizar tales estudios comparativos. De

estos, sabemos que resulta dudoso suponer que medir lo mismo con tareas idénticas por parte de pruebas diferentes implica obtener indicadores de competencia equivalente. Los trabajos realizados en población infantil estadounidense comparando la nueva edición del TVIP (PPVT-III: Dunn y Dunn; 1997) con la versión revisada (PPVT-R: Dunn y Dunn; 1981), documentan diferencias de hasta 10 puntos más por término medio en puntuaciones típicas a favor de la última versión (Williams, 1998; Ukrainetz y Duncan, 2000). En otras investigaciones como la de Pankratz, Morrison y Plante (2004), con los mismos referentes de evaluación, pero utilizando población adulta (18 a 53 años), se observa que los sujetos con una competencia lingüística inferior también obtenían puntuaciones significativamente mayores en el PPVT-III frente a las observadas en PPVT-R.

Por tanto, además de explorar la relación que pudiera existir en pruebas semejantes, también resultaría interesante añadir, tal y como aquí haremos, un análisis de las divergencias por cuanto parece que variables como las características de los ítems o de la muestra, junto con los procedimientos empleados en la tipificación, son factores que inciden sobre la puntuación final que recoge el rendimiento del evaluado, y no parecen ser equivalentes cuando se utilizan pruebas diferentes aun empleando el mismo tipo de tareas. Si esto sucede con reformulaciones de la misma prueba, no sería de extrañar encontrar una situación similar al emplear tests diferentes, como así ocurría al comparar la evaluación del rendimiento en vocabulario en niños de entre 3-6 años con VAVEL Infantil y TVIP (Ferrer *et al.*, 2006b).

La segunda aproximación descansa en las relaciones test-criterio, enfoque que informa sobre la eficacia del test en la predicción de algún criterio relevante. En este caso, la evidencia de validez viene dada por la confirmación de la hipótesis de la existencia de relación entre las puntuaciones obtenidas en el test y las medidas de algún criterio que intenta predecir, aportando aquí datos relativos a la relación mantenida entre VAVEL Inicial y el juicio subjetivo de las maestras respecto al rendimiento general de los niños. Con este análisis se intenta también responder a un enfoque de carácter ecológico (Sbordone, 1996), que destaca la necesidad de realizar investigaciones que acoten la relación que la competencia medida en tests de evaluación pudiera tener sobre aspectos funcionales, en nuestro caso respecto a su incidencia sobre el rendimiento escolar. Este tipo de trabajos ya empieza a abrirse camino en nuestro contexto, encontrando antecedentes similares en el estudio de Mendoza, Muñoz, Fresneda y Carballo (2005), en que se obtienen datos que permiten afirmar que la competencia para la comprensión de estructuras gramaticales se relaciona con la evaluación del rendimiento

académico de los estudiantes, considerado éste a partir de la Escala de Valoración del Lenguaje Observable del *CELF-3 Spanish Edition* (Semel, Wiig y Secord; 1997).

### Objetivos

El objetivo fundamental de este trabajo consiste en establecer evidencias de validez en relación con la prueba VAVEL Inicial. Para ello se explora por un lado el grado de relación/discrepancia a la hora de determinar la comprensión de vocabulario entre VAVEL Inicial y el TVIP; por otro se analiza el grado de acuerdo a la hora de juzgar el nivel de dominio en comprensión léxica por parte de TVIP y de considerar el nivel de rendimiento general como alumnos por parte de los tutores.

Como objetivos específicos contemplamos los siguientes:

1. Investigar la relación entre VAVEL Inicial y TVIP, pormenorizando el análisis según edad.
2. Presentar datos comparativos acerca de cómo estiman la comprensión de vocabulario ambas pruebas de evaluación estandarizadas para su uso en población española.
3. Explorar el nivel de acuerdo a la hora de determinar el rendimiento para los mismos casos, según la valoración subjetiva del rendimiento realizada por las maestras tutoras en relación con la calificación obtenida según dominio léxico en VAVEL Inicial y TVIP.
4. Aportar elementos de juicio que permitan establecer la bondad de una prueba frente a otra, a la hora de estimar la comprensión de vocabulario.

### Método

#### *Grupo de Estudio*

El estudio se desarrolló con un grupo de 89 niños y niñas, distribuyéndose según grupo de edad y sexo tal como muestra la tabla 1.

Edad	Niño	Niña	N	%
<b>6 a 7 años</b>	14	13	27	30,3
<b>7 a 8 años</b>	14	14	28	31,5
<b>8 a 9 años</b>	16	18	34	38,2
<b>TOTAL</b>	44	45	89	100

Tabla 1: Distribución según edad y sexo

Procedían de seis escuelas diferentes, cinco de la provincia de Valencia y una de la de Castellón. Cuatro de los centros eran públicos y aportaban al grupo de estudio 45 alumnos, mientras que los 44 restantes procedían de dos centros concertados. Todos los niños y niñas tenían como lengua familiar y escolar el castellano.

### *Procedimiento*

Al ser parte de la muestra empleada para la tipificación de VAVEL Inicial, todos los casos fueron primero evaluados con tal prueba. Posteriormente se les aplicó el TVIP. No se balanceó el orden de pase por cuanto los resultados que aquí se presentan fueron extraídos de la fase de estudio para la validez, posterior a la de tipificación. Entre la aplicación de una y otra transcurrió un promedio de dos meses, siendo poco probable que el orden influyera en los resultados obtenidos. En cualquier caso, los indicadores usados no son puntuaciones directas sino puntuaciones típicas y edades de desarrollo, ambas derivadas de acuerdo con la edad cronológica en el momento de la aplicación, permitiendo establecer comparaciones sin sesgos potenciales debidos al tiempo y, por tanto, edad a la hora de efectuarse el pase de una y otra prueba.

En paralelo al proceso de aplicación de VAVEL Inicial se invitaba a las tutoras a completar un formulario que contenía la cuestión “¿Cuál es su consideración acerca del rendimiento general como alumno de estos/as niños/as respecto a su clase?”. A continuación contaban con el listado de sus alumnos, debiendo señalar en una columna adyacente su valoración de acuerdo con la siguiente escala: 1= rendimiento bajo, 2= rendimiento medio y 3= rendimiento alto. Participaron en esta evaluación los 20 maestros tutores de las diferentes aulas con que se trabajó en los 6 centros escolares. Hay que hacer notar que siempre evaluaban sin conocer el resultado obtenido en las pruebas de vocabulario.

De esta manera contábamos, de acuerdo con el rendimiento, con una clasificación de los estudiantes establecida por parte de las profesoras en tres niveles (alto, medio y bajo). Con los resultados obtenidos en VAVEL Inicial y TVIP también se estableció una clasificación de los niños en tres niveles de vocabulario según los siguientes parámetros:

1. Alto: Cuando se obtuvo una Puntuación Típica Derivada (PTD en adelante) por encima de una desviación típica en relación con el promedio (las puntuaciones estándar ofrecidas por ambas pruebas tienen una media de 100 y una desviación típica de 15).

2. Medio: PTD entre más-menos una desviación típica de la media
3. Bajo: PTD por debajo de una desviación típica respecto al promedio

También se consideraron en los análisis ofrecidos las edades de desarrollo de vocabulario estimadas en ambas pruebas: Edad Equivalente en términos de TVIP y Edad de Desarrollo de Vocabulario en términos de Vavel Inicial.

## Resultados

### *Relación entre VAVEL Inicial y TVIP*

Considerando en su totalidad el grupo de estudio, se observa una relación lineal positiva significativa entre ambas pruebas. El análisis pormenorizado por grupo de edad corrobora dicha relación en todos los niveles con una magnitud bastante similar.

Edad	N	r	Significación
<b>6-9 años (total)</b>	89	0,63	0,000
<b>6-7 años</b>	27	0,61	0,001
<b>7-8 años</b>	28	0,65	0,000
<b>8-9 años</b>	34	0,67	0,000

Tabla 2: Correlación entre TVIP y VAVEL Inicial

### *Estimación de la comprensión de vocabulario en VAVEL Inicial y TVIP*

Aun observando a través de la correlación establecida en el punto anterior, que existe una clara relación entre una y otra prueba, la comparación del rendimiento obtenido según se utilice una u otra revela claras diferencias respecto a cómo ambas estiman el dominio léxico.

Como podemos observar en la tabla 3, considerando el conjunto del grupo existe una diferencia de 11 puntos en el rendimiento promedio observado, siendo mayor la puntuación en TVIP, lo cual supone una diferencia próxima a una desviación típica, siendo tal disparidad significativa desde el punto de vista estadístico. En los tres niveles de edad se observan diferencias significativas, siendo siempre mayor la puntuación de los grupos cuando se valora con TVIP, alcanzando el mayor grado de discrepancia en el grupo de mayor edad.

Edad		N	Media	Sx	t	Gl	Sig. (bilateral)
6-9 años (Total)	TVIP	89	111,43	12,37	-8,777	88	0,000
	VAVEL		100,25	15,12			
6-7 años	TVIP	27	110,96	14,09	-4,085	26	0,000
	VAVEL		100,93	14,86			
7-8 años	TVIP	28	110,07	12,90	-3,991	27	0,000
	VAVEL		100,18	17,14			
8-9 años	TVIP	34	112,91	10,55	-7,326	33	0,000
	VAVEL		99,76	13,95			

Tabla 3: Pruebas t para muestras relacionadas entre puntuaciones típicas para TVIP y VAVEL Inicial

*Determinación del nivel de rendimiento en vocabulario según VAVEL Inicial y TVIP en relación con la consideración del rendimiento como alumnos según valoración de las maestras*

En términos generales, existe un grado semejante de asociación entre la comprensión léxica valorada en ambas pruebas y el rendimiento percibido por las maestras, alcanzando niveles de significación estadística sólo para el caso de VAVEL Inicial ( $\phi=0.324$ ;  $p=0.05$ ), no así en TVIP ( $\phi=0.301$ ;  $p=0.09$ ).

A la hora de concretar el nivel de rendimiento destacan ciertas discrepancias notables: mientras TVIP sólo considera 2 casos con rendimiento bajo, VAVEL Inicial delimita 17 casos y las profesoras sitúan en dicha categoría 14. En el otro extremo, VAVEL Inicial ubica con un rendimiento alto 18, TVIP eleva este guarismo hasta 38 casos, cifra muy cercana a la opinión mostrada por los maestros que sitúan en este nivel a 39 casos.

En este apartado volvemos a observar diferencias entre TVIP y VAVEL Inicial a pesar de su grado de relación. La tabla 4 compara la competencia en vocabulario entre TVIP y VAVEL Inicial agrupando los niños según nivel de rendimiento general de acuerdo con la opinión de las maestras. Son estos datos los que contribuyen a esclarecer el objetivo de establecer relaciones entre los tests y el rendimiento. Así, los 14 niños de bajo rendimiento según las maestras, obtienen una Puntuación Típica Derivada promedio de 88 en VAVEL, esto es, próxima a una desviación por debajo del promedio (85), mientras que por el contrario de lo que cabría esperar, este conjunto obtiene con TVIP un promedio de 102, ligeramente superior a la media de estas

puntuaciones estándar (100). Tanto para los niños de rendimiento medio como alto según las maestras, ambas pruebas delimitan puntuaciones similares, en torno a 100 en VAVEL Inicial y cercanas a 115 (una desviación por encima de la media) en TVIP. En todos los grupos de rendimiento según las maestras, las dos pruebas difieren significativamente en términos estadísticos.

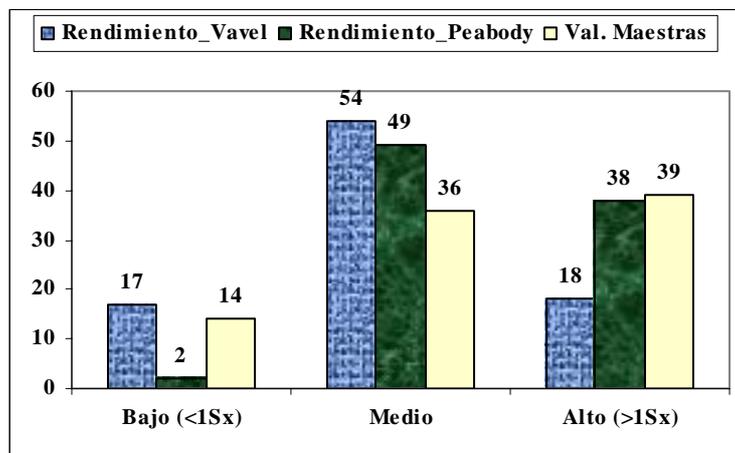


Gráfico 1: Frecuencia de Nivel de Rendimiento Léxico en TVIP, VAVEL Inicial y Valoración de Rendimiento General por parte de las maestras.

Puntuaciones Típicas Derivadas	Valoración Maestras	N	Media	Desv. típ.	t	Gl	sig. (bilateral)
VAVEL Ini.	Total	89	100,25	15,12	-8,777	88	0,000
TVIP		89	111,43	12,37			
VAVEL Ini.	Bajo	14	88,64	13,25	-4,561	13	0,001
TVIP		14	102,07	12,65			
VAVEL Ini.	Medio	36	101,84	14,84	-5,222	35	0,000
TVIP		36	113,28	11,44			
VAVEL Ini.	Alto	39	102,95	14,38	-5,534	38	0,000
TVIP		39	113,08	11,86			

Tabla 4: Pruebas t para muestras relacionadas entre puntuaciones típicas para TVIP y VAVEL Inicial.

		VAVEL Inicial			Total		
		Baja (<1Sx)	Media	Alta (>1Sx)	f	%	
TVIP	Baja (<1Sx)	f	2	0	0	2	2,25
		% del total	2,25	0,00	0,00		
	Media	f	12	33	4	49	55,06
		% del total	13,48	37,08	4,49		
	Alta (>1Sx)	f	3	21	14	38	42,70
		% del total	3,37	23,60	15,73		
<b>Total</b>		<b>f</b>	17	54	18		
		<b>%</b>	19,10	60,67	20,22		

Tabla 5: Tabla de Contingencia para el Nivel de Rendimiento Léxico en TVIP y VAVEL Inicial.

Dadas tales diferencias, conviene profundizar en el grado de acuerdo/discrepancia de la calificación de nivel de rendimiento por parte de ambas pruebas. Así, la tabla 5 nos ayuda a determinar que un 55% de los casos son clasificados de la misma forma por ambas pruebas, alcanzando el mayor grado de acuerdo en el nivel medio de rendimiento donde ambas coinciden en 33 de los 89 casos (37,08%). En cambio, mientras que TVIP consideraría sólo 2 casos con un rendimiento bajo, VAVEL Inicial sumaría a estos 15 más. En el extremo del rendimiento alto, mientras que TVIP destaca 38 casos, VAVEL Inicial sólo considera como tal a 18.

Si se analiza de acuerdo con la edad, encontramos el mayor grado de acuerdo al clasificar los niños más pequeños: entre 6-7 años hay una correspondencia correcta del 74,1%, mientras que para los otros intervalos (7-8; 8-9) es respectivamente del 39,3% y 52,9%.

*Respecto a la bondad de estimación en la comprensión de vocabulario de VAVEL Inicial y TVIP*

Responder a esta cuestión resulta complejo, pues carecemos de referentes con los que juzgar un mejor ajuste a la hora de establecer la competencia en vocabulario, más allá de los baremos establecidos por estas mismas pruebas.

De este modo, para comparar una y otra e intentar informar acerca de su mejor o peor comportamiento, tomamos como punto de referencia tres factores:

- a) el grado de ajuste en la estimación de la edad de vocabulario,

- b) alejamiento/aproximación de la distribución normal en el establecimiento del nivel de rendimiento, y
- c) la relación que mantiene cada prueba respecto al juicio de rendimiento general establecido por las maestras.

En relación con el primer factor, la tabla 6 nos expresa el promedio de diferencia en meses entre la edad cronológica y la estimada en aptitud léxica por cada test. Así, mientras VAVEL Inicial promedia para los 89 niños y niñas una edad estimada de desarrollo de vocabulario con una diferencia de 1,2 meses por debajo de la cronológica, TVIP promedia una diferencia de 14,7 meses (un año y dos meses) por encima de la edad cronológica con una desviación mucho más acusada. Tales diferencias se mantienen de forma similar en términos cuantitativos para los tres grupos de edad.

Edad	VAVEL		TVIP	
	Media	Sx	Media	Sx
<b>6-9 (total)</b>	-1,2	9,6	14,7	17,3
<b>6-7</b>	1,6	11,9	11,3	14,7
<b>7-8</b>	-3,4	10,1	12,6	14,9
<b>8-9</b>	1,0	6,2	19,1	20,3

Tabla 6: Diferencia entre Edad Cronológica y Edad Estimada de Vocabulario en ambas pruebas.

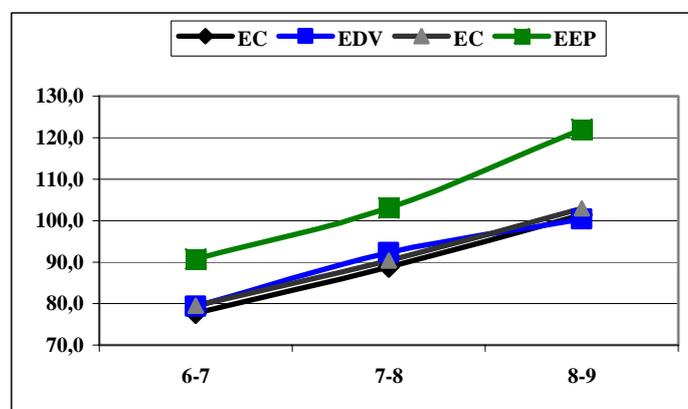


Gráfico 2: Comparación entre Edad Cronológica y Edad de Vocabulario obtenida en las distintas pruebas aplicadas.

El gráfico 2 refleja la diferencia entre Edad Cronológica (EC)<sup>2</sup>, Edad de Desarrollo de Vocabulario (EDV) obtenida en VAVEL Inicial y Edad Equivalente en TVIP (EEP). Los datos reflejan un mayor ajuste entre edad cronológica y la edad de desarrollo estimada con VAVEL en todos los grupos de edad; mientras que en TVIP la Edad Equivalente tiende a distanciarse en torno a 10 meses más para los grupos de entre 6-7 y 7-8 años, produciéndose un mayor alejamiento en 8-9 años.

Obviamente, no tiene por qué corresponderse la edad cronológica con la edad estimada: algunos niños pueden que tengan un dominio léxico muy superior al comúnmente mostrado a su edad, mientras que otros pueden caracterizarse por una competencia típica de edades inferiores. Si bien, conforme aumentamos la cuantía de niños estudiados en cualquier aptitud humana, tal es el caso de la comprensión de vocabulario, cabe esperar una aproximación a una distribución normal del rendimiento. Así, en este intento por aportar datos respecto a la bondad de una y otra prueba, añadimos información respecto a cómo se distribuyen los niños en niveles de rendimiento en comparación con una distribución teórica normal. Así, contrastamos la distribución observada de los 89 niños para los tres niveles de rendimiento respecto a la distribución teórica que se correspondería con la normalidad (14 casos de rendimiento alto y 14 de bajo -15,87% del grupo en cada extremo- y 61 con rendimiento medio -68,26% del total-). El análisis realizado nos permite asumir una distribución normal según los tres niveles de rendimiento para el caso de VAVEL Inicial (Chi cuadrado=2,589 p=0,274), afirmación que no podemos sostener a la luz de tales resultados para el caso de TVIP (Chi cuadrado =53,789 p=0,000). En este caso, se produce una clara asimetría en la distribución con mayor acumulación de casos en el promedio y por encima del mismo, y práctica inexistencia de niños con bajo rendimiento.

---

<sup>2</sup> En la gráfica se contemplan dos líneas referentes a EC, puesto que la aplicación de una y otra prueba se distanció en torno a 2 meses por término medio tal como queda explicado en el apartado de procedimiento. En negro la edad cronológica en el pase de VAVEL Inicial, en gris para el momento de pase de TVIP.

			Valoración Maestras			Total	
			Baja	Media	Alta	f	%
TVIP	Baja	f	1	1	0	2	2,25
		% del total	1,1	1,1	0,0		
	Media	f	11	16	22	49	55,06
		% del total	12,4	18,0	24,7		
	Alta	f	2	19	17	38	42,70
		% del total	2,2	21,3	19,1		
Total	f	14	36	39			
	%	15,73	40,45	43,82			

Tabla 7: Tabla de Contingencia para el Nivel de Rendimiento Léxico en TVIP y Valoración de los maestros.

			Valoración Maestras			Total	
			Baja	Media	Alta	f	%
VAVEL	Baja	f	6	6	5	17	19,1
		% del total	6,7	6,7	5,6		
	Media	f	8	20	26	54	60,7
		% del total	9,0	22,5	29,2		
	Alta	f	0	10	8	18	20,1
		% del total	0,0	11,2	9,0		
Total	f	14	36	39			
	%	15,7	40,4	43,8			

Tabla 8: Tabla de Contingencia para el Nivel de Rendimiento Léxico en VAVEL y Valoración de los maestros.

Respecto al tercer factor, que se apoya en el criterio externo de la valoración de los maestros y la relación que el rendimiento general puede tener con la comprensión de vocabulario, descubrimos a través de las tablas 7 y 8 que ambas pruebas tienen un grado de coincidencia en la clasificación del nivel de rendimiento respecto al juicio de las maestras del 38,2% (34 casos), si bien con una distribución dispar. Así, algunos niños de rendimiento bajo para las maestras son situados por TVIP con un nivel de rendimiento medio e incluso en dos casos alto, mientras que VAVEL Inicial no identificaría ninguno de estos con rendimiento alto en vocabulario. En cambio, de los 39 considerados como de rendimiento alto por las maestras, VAVEL Inicial llega a calificar 5 de ellos con un rendimiento en léxico bajo, mientras que TVIP no ubica ninguno en este nivel de rendimiento.

### **Conclusiones**

Dados los resultados obtenidos de relacionar TVIP con VAVEL Inicial, estaríamos en condiciones de otorgar validez convergente a esta última. Encontramos una relación positiva significativa de VAVEL Inicial con el Test de Vocabulario en Imágenes Peabody, prueba estandarizada para uso con población española que, desde la década de los 80, ha sido una de las más utilizadas en la evaluación psicolingüística.

Respecto a las relaciones test-criterio, también obtenemos evidencias de validez, existiendo relación entre el rendimiento académico informado por las maestras y la competencia léxica establecida por VAVEL Inicial. También TVIP observa un grado de relación semejante aunque sin alcanzar significación estadística. Somos conscientes de que el rendimiento escolar de un niño es consecuencia de la interacción de múltiples factores, siendo difícil encontrar resultados determinantes en investigaciones como esta en que se relaciona con variables aisladas. Si bien, la exigencia de que la evaluación deba responder también a criterios lo más funcionales posibles, propicia la incorporación de este tipo de criterios en los estudios sobre validez de instrumentos psicométricos.

A pesar de las evidencias obtenidas partiendo de parámetros estadísticos basados en coeficientes de correlación, resultan notables las diferencias observadas entre cómo juzgan la competencia léxica VAVEL Inicial y TVIP. Las distancias observadas en el promedio de rendimiento tanto para el conjunto del grupo de estudio como para cada uno de los grupos de edad, así como el hecho de que un número notable de niños obtenga asignaciones de nivel de rendimiento diferentes según se aplique una u otra, confirma que los índices derivados de ambas no se pueden intercambiar, lo cual es congruente con estudios ya citados que reflejaban diferencias con población estadounidense infantil (Ukrainetz y Duncan, 2000) y adulta (Pankratz *et al.*, 2004) en versiones distintas de Peabody (PPVT-R y PPVT-III) y según nuestra propia experiencia a la hora de comparar VAVEL Infantil y TVIP para niños de entre 3-6 años (Ferrer *et al.*, 2006b).

De forma generalizada, existe una consideración de la comprensión de vocabulario superior cuando se utiliza TVIP. Por ello, cabe plantearse la cuestión acerca de si esta prueba estaría sobrestimando en la actualidad el nivel de vocabulario o, por el contrario, VAVEL Inicial estaría subestimándolo dado el sentido y significación de sus diferencias. Indudablemente, tratándose de tests basados en la misma tarea, parece claro que el conjunto de reactivos (incluyendo en éste no sólo el léxico empleado sino también las imágenes) recogido en

VAVEL Inicial acaba por ser de mayor dificultad, al menos para este grupo de estudio.

Consideramos que, en la actualidad, con los criterios manejados en este estudio, Vavel Inicial es una prueba que ofrece una mejor aproximación para evaluar la comprensión de vocabulario en edades de 6 a 9 años. Su mejor ajuste en relación con la estimación sobre la edad de desarrollo en vocabulario respecto a edad cronológica y su mayor aproximación a una distribución normal del rendimiento en el grupo estudiado así nos lo hacen pensar. El uso de un criterio externo como el rendimiento académico juzgado por las maestras, no aporta elementos de juicio determinantes a favor de una u otra.

A ello se une la actualidad de sus baremos frente a los ofrecidos por TVIP; quedaría por estudiar en el futuro el comportamiento de VAVEL Inicial en relación con la recién aparecida versión española del PPVT-III (Dunn, Dunn y Arribas, 2006). También, en la línea de mostrar evidencias de validez comparadas, sería aconsejable realizar en el futuro estudios de este tipo con poblaciones clínicas. Con ello, testaríamos la sensibilidad para la detección de déficits en la comprensión de vocabulario que pudieran apuntar a trastornos más amplios en el desarrollo del lenguaje. No en vano, estudios como el de Gray *et al.* (1999) han puesto en duda la utilidad de este tipo de pruebas en niños al testar la eficacia de cuatro pruebas, entre ellas, el PPVT-III.

### Referencias bibliográficas

- American Educational Research Association, American Psychological Association y National Council on Measurement in Education (AERA, APA, NCME) (1999): *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
- Brancal, M.; Ferrer, A.; Alcántud, F.; Quiroga, M. (1998): *EDAF. Evaluación de la Discriminación Auditiva y Fonológica*. Barcelona: Ediciones Lebon.
- Brancal, M., Ferrer, A., Carreres, D., Tomás, I. y Ávila, V. (2005): *VAVEL Infantil. Prueba de valoración del Vocabulario Español 3-6 años*. Barcelona: Lebon.
- Dunn, L.I.M. (1981). *Test de Vocabulario Imágenes Peabody. Adaptación Española*. Madrid: MEPSA (Traducción de Santiago Pereda Marín)
- Dunn, L.I.M. y Dunn, L.I.M. (1981): *Peabody Picture Vocabulary Test – Revised Version*. Circle Pines, MN: American Guidance Service.
- Dunn, L.I.M. y Dunn, L.I.M. (1997): *Peabody Picture Vocabulary Test - Third Edition*. Circle Pines, MN: American Guidance Service.
- Dunn, L.I. M. y Dunn, L.I. M. y Arribas, D. (2006): *PPVT-III Peabody, Test de Vocabulario en Imágenes*. Madrid: TEA Ediciones.

Beatriz Gallardo, Carlos Hernández y Verónica Moreno (Eds): *Lingüística clínica y neuropsicología cognitiva. Actas del Primer Congreso Nacional de Lingüística Clínica. Vol 2: Lingüística y evaluación del lenguaje*, coord. por C. Hernández y M. Veyrat. ISBN: 84-370-6576-3.

Antonio FERRER, Vicenta ÁVILA, Mario BRANCAL, Inés TOMÁS y David CARRERES  
*VAVEL inicial: evidencias de validez*

- Ferrer, A.; Brancal, M.; Ávila, V.; Tomás, I. y Carreres, D. (2006a): *VAVEL Inicial. Prueba de valoración del Vocabulario Español 6-9 años*. Barcelona: Lebón.
- Ferrer, A.; Ávila, V.; Tomás, I.; Brancal, M.; Carreres, D.(2006b). Rendimiento comparado del léxico receptivo a través de dos pruebas de valoración del vocabulario español. *Actas 25 Congreso Internacional de AELFA* (pp. 545-552). Granada: Editorial Universidad de Granada.
- Gray, S., Plante, E., Vance, R. y Henrichsen, M. (1999): The Diagnostic Accuracy of Four Vocabulary Tests Administered to Preschool-Age Children. *Language Speech and Hearing Services in Schools*, 30(2), 196-206.
- Mendoza, E., Muñoz, J., Fresneda, M. D. y Carballo, G. (2005): *Revista de Logopedia, Foniatría y Audiología*, 25(2), 62-71.
- Pankratz, M., Morrison, A. y Plante, E. (2004): Difference in Standard Scores of Adults on the Peabody Picture Vocabulary Test (Revised and Third Edition). *Journal of Speech Language and Hearing Research*, 47(3), 714-718.
- Sbordone, R. J. (1999): Ecological validity: some critical issues for the neuropsychologist. En R. J. Sbordone y C. J. Long (eds), *Ecological validity of Neuropsychological Testing* (pp. 15-41). Delray Beach: FL: St. Lucie Press.
- Semel, E.; Wiig, E. H. y Secord, W. A. (1997): *Clinical Evaluation of Language Fundamentals-Third Edition (CELF-3 Spanish Edition)*. San Antonio, TX: The Psychological Corporation.
- Ukrainetz, T. A. y Duncan, D. S. (2000): From Old to New: Examining Score Increases on the Peabody Picture Vocabulary Test-III. *Language Speech and Hearing Services in Schools*, 31(4), 336-339.
- Williams, K. T. (1998): Peabody Picture Vocabulary Test-III: What is new and different? *Clinical Connection*, 11(3), 6-8.

Beatriz Gallardo, Carlos Hernández y Verónica Moreno (Eds): *Lingüística clínica y neuropsicología cognitiva. Actas del Primer Congreso Nacional de Lingüística Clínica. Vol 2: Lingüística y evaluación del lenguaje*, coord. por C. Hernández y M. Veyrat. ISBN: 84-370-6576-3.