

Didáctica

LAS RÚBRICAS EN LA EVALUACION ESCOLAR: SU CONSTRUCCIÓN Y SU USO

José Guillermo Martínez-Rojas*
Universidad Nacional de Colombia, Colombia

Cuando se habla de evaluación en educación generalmente se hace referencia a herramientas de corte cuantitativo como las pruebas objetivas, o de corte cualitativo como los portafolios, los ensayos, los informes descriptivos, por ejemplo. Es común encontrar razones a favor y en contra de estos dos modelos de evaluación, tal vez porque en el ámbito educativo los expertos suelen tomar posiciones un tanto polarizadas que generalmente los llevan a defender una y atacar la otra. El propósito de este escrito no es hacer un abordaje sobre el debate entre estos dos modelos o paradigmas evaluativos sino más bien presentar, de manera teórica, un modelo de evaluación denominado rúbricas o matrices de evaluación que permiten una integración de dichos modelos, y a su vez proporciona indicaciones a quienes deseen emplearlo para su construcción y su uso.

Aunque las rúbricas son una herramienta aplicable a diferentes campos de los procesos educativos, en este artículo se hace una presentación de la misma aplicable única y exclusivamente al proceso de aprendizaje o al proceso de construcción del saber por parte de los estudiantes con el propósito de dar una perspectiva clara y precisa a aquellos educadores que estén interesados en implementar o usar dicha herramienta o mejorar una práctica que han introducido en su propio quehacer pedagógico.

Antes de ahondar en el tema de las rúbricas propiamente dicho conviene recordar el concepto de evaluación que maneja el Ministerio de Educación Colombiano, pues la rúbrica, como herramienta de evaluación, es válida entenderla en referencia a dicho concepto. En términos generales, evaluar significa emitir juicios sobre un asunto determinado e implica un proceso de investigación que permita, de acuerdo con unos criterios, saber qué tanto se ha avanzado en un determinado proceso. En el campo educativo se puede entender la evaluación como “La acción permanente por medio de la cual se busca apreciar, estimar y emitir juicios sobre procesos de desarrollo del alumno o sobre los procesos pedagógicos o administrativos, así como sobre sus resultados con el fin de elevar y mantener la calidad de los mismos. El campo de la evaluación educativa es muy amplio: los objetos, los procesos, las acciones, las relaciones, todos pueden ser estimados, apreciados o valorados según determinadas exigencias, necesidades, intereses, expectativas o aspiraciones” (MEN, 1997, p. 17).

En el contexto educativo, una *rúbrica* es un conjunto de criterios o de parámetros desde los cuales se juzga, valora, califica y conceptúa sobre un determinado aspecto del proceso educativo. Las rúbricas también pueden ser entendidas como pautas que permiten aunar criterios, niveles de logro y descriptores cuando de juzgar o evaluar un aspecto del proceso educativo se trata (Vera, 2004). Según Díaz Barriga (2005) las rúbricas son guías o escalas de evaluación donde se establecen niveles progresivos de dominio o pericia relativos al desempeño que una persona muestra respecto de un proceso o producción determinada. También se puede decir que las rúbricas integran un amplio rango de criterios que cualifican de modo progresivo el tránsito de un desempeño incipiente o novato al grado del experto. Son escalas

* Departamento de Psicología, Universidad Nacional de Colombia, Bogotá, Colombia. E-mail: josememo22@hotmail.com

ordinales que destacan una evaluación del desempeño centrada en aspectos cualitativos, aunque es posible el establecimiento de puntuaciones numéricas.

Las rúbricas como instrumento de evaluación son perfectibles en tanto que las mismas constituyen una herramienta que se puede ir ajustando con la práctica hasta encontrar el valor justo de las metas de la evaluación a las cuales se espera llegar o se quiere que los estudiantes lleguen.

También se puede afirmar que una rúbrica es una descripción de los criterios empleados para valorar o emitir un juicio sobre la ejecutoria de un estudiante en algún trabajo o proyecto. O dicho de otra manera, una rúbrica es una *matriz* que puede explicarse como un listado del conjunto de criterios específicos y fundamentales que permiten valorar el aprendizaje, los conocimientos o las competencias logrados por el estudiante en un trabajo o materia particular.

¿Por qué usar rúbricas en la evaluación educativa?

Algunas de las ventajas que trae a los procesos educativos el uso de las rúbricas de acuerdo con Goodrich (documento electrónico), citado por Díaz Barriga (2005) son, entre otras.

1. Son una poderosa herramienta para el maestro que le permite evaluar de una manera más objetiva, pues los criterios de la medición están explícitos y son conocidos de antemano por todos, no se los puede cambiar arbitrariamente y con ellos se hace la medición a todos los casos sobre los cuales se ofrezca emitir juicios.

2. Promueven expectativas sanas de aprendizaje en los estudiantes pues clarifican cuáles son los objetivos del maestro respecto de un determinado tema o aspecto y de qué manera pueden alcanzarlos los estudiantes.

3. Enfocan al profesor para que determine de manera específica los criterios con los cuales va a medir y documentar el progreso del estudiante.

4. Permiten al maestro describir cualitativamente los distintos niveles de logro que el estudiante debe alcanzar.

5. Permiten que los estudiantes conozcan los criterios de calificación con que serán evaluados, previamente al momento mismo de la evaluación.

6. Aclaran al estudiante cuáles son los criterios que debe utilizar al evaluar su trabajo y el de sus compañeros.

7. Permiten que el estudiante evalúe y haga una revisión final a sus trabajos, antes de entregarlos al profesor.

8. Indican con claridad al estudiante las áreas en las que tiene falencias o deficiencias y con esta información, planear con el maestro los correctivos a aplicar.

9. Proveen al maestro información de retorno sobre la efectividad del proceso de enseñanza que está utilizando.

10. Proporcionan a los estudiantes retroalimentación sobre sus fortalezas y debilidades en las áreas que deben mejorar.

11. Reducen al mínimo la subjetividad en la evaluación.

12. Promueven la responsabilidad.

13. Ayudan a mantener el o los logros del objetivo de aprendizaje centrado en los estándares de desempeño establecidos y en el trabajo del estudiante.

14. Proporcionan criterios específicos para medir y documentar el progreso del estudiante.
15. Son fáciles de utilizar y de explicar.

Clases de rúbricas

Las rúbricas pueden ser de dos tipos: globales u holísticas y analíticas (Mertler, 2001).

La rúbrica global u holística es aquella que permite hacer una valoración de conjunto del desempeño del estudiante sin determinar o definir los aspectos fundamentales que corresponden al proceso o tema evaluado (Mertler, 2001). Como su nombre lo indica, permite hacer una valoración general o de conjunto de un determinado aspecto (Vera, 2004). Este tipo de rúbrica sólo tiene descriptores correspondientes a unos niveles de logro que son globales. A manera de ejemplo, si se va a evaluar el nivel de inglés de un estudiante en sexto grado en lo que se refiere a fluidez en el manejo de la lengua, se puede construir la rúbrica que se ilustra en la Tabla 1.

Tabla 1
Rúbrica para evaluar fluidez en el manejo de inglés

Puntos	Nombre	Descriptor
0 a 2	Básico / Nivel de inicio / Insuficiente	El estudiante tiene un vocabulario muy básico que le permite decir palabras y algunas breves frases en inglés pero no puede construir un discurso con coherencia, consistencia, claridad y profundidad.
3 a 4	Básico-Avanzado / Nivel básico / Deficiente	El estudiante tiene un vocabulario un poco más amplio que le permite construir frases con cierta coherencia y consistencia pero que no le permiten construir un discurso que sea lo suficientemente coherente, consistente, claro y profundo.
5 a 6	Intermedio / Nivel Medio / Aceptable	El estudiante posee un vocabulario amplio que le permite construir frases completas y con sentido, y que le permite construir discursos breves y con sentido, pero con poca claridad y más bien superficial.
7 a 8	Intermedio-Avanzado / Sobresaliente	El estudiante posee un vocabulario amplio y suficiente que no sólo le permite construir frases completas y con sentido sino que pueden ser articuladas en párrafos y discursos con una cierta coherencia, sentido y con alguna claridad.
9 a 10	Avanzado / Nivel de alcance / Nivel superior / Excelente	El estudiante posee un vocabulario amplio y suficiente que no sólo le permite construir frases completas y con sentido sino que pueden ser articuladas en párrafos y discursos un grado óptimo de coherencia, consistentes, con claridad y profundidad suficiente para transmitir un pensamiento propio o de algún autor o aspecto estudiado.

Lo clave de este tipo de rúbrica es poder graduar o gradar los diferentes niveles teniendo suficientes criterios o elementos que definen un determinado nivel. Se puede y debe poner un nombre o título para que los estudiantes puedan identificar lo que significa cada nivel. Pero además de lo anterior, Díaz Barriga (2005), afirma que los descriptores de una rúbrica deben: a) ser sensibles a los objetivos educativos perseguidos, b) ser apropiados para la etapa de desarrollo de los alumnos, c) requieren credibilidad ante los diversos agentes involucrados en la evaluación, d) ser claramente comunicables y e) hacer explícita la dimensión ética de la evaluación: tienen que ser justos y libres de sesgos.

A diferencia de la rúbrica global, **la rúbrica analítica** es mucho más compleja, amplia y definitiva para la evaluación (Vera, 2004; Rodríguez, documento electrónico). En este tipo de rúbricas es necesario dar los siguientes pasos:

1. Definir cada uno de los criterios que se va a evaluar. Se trata de “pensar” cuáles son los aspectos que se deben tener en cuenta cuando se evalúa este o aquel aspecto. Si el ejercicio es el mismo anterior,

evaluar el nivel de inglés de un estudiante en 6 de bachillerato en lo que se refiere a fluidez en el manejo de la lengua, habría que hacer el siguiente ejercicio: ¿qué aspectos se deben tener en cuenta para que haya fluidez en el manejo de una lengua en sexto grado? Tal vez “vocabulario”, “construcción gramatical de frases”, “uso adecuado del lenguaje” y todos aquellos que el grupo o el docente considere que son necesarios, esenciales y suficientes (tres, cuatro, cinco, los que sean).

2. Supóngase que los tres nombrados anteriormente son los criterios necesarios y suficientes para poder evaluar la fluidez de un estudiante, lo que sigue es ponderarlos, es decir, definir qué tanto cada uno de ellos influye en la decisión global o cuanto “pesa” cada uno de ellos a la hora de determinar el nivel de fluidez de un estudiante. Por ejemplo, se puede decir que de 1 a 100, el vocabulario puede tener una importancia del 60% pues es importante pero no el más importante y determinante; el de construcción gramatical de frases un 80%, puesto que éste podría ser un poco más importante que el de vocabulario; y el de uso adecuado del lenguaje, por ejemplo un 100%, dado que fluidez dependería en mucho de este criterio. Igual, esta ponderación debe tener unas razones que permitan definir por qué se le da ese valor y no otro, pero esto sólo lo puede hacer un “experto” en la materia.

3. El tercer paso es realizar para estos tres criterios los descriptores para cada uno de los niveles de logro, tal y como se mencionó anteriormente en la exposición de la rúbrica global. Habría que elaborar un esquema como el de la tabla 2 y completarlo con los descriptores de cada uno de los niveles de logro.

Tabla 2.
Esquema para la construcción de una rúbrica analítica

Puntos	Criterio 1 Vocabulario	Criterio 2. Construcción Gramatical de Frases	Criterio 3. Uso adecuado del Lenguaje
0 a 2	Nivel 1.	Nivel 1.	Nivel 1.
3 a 4	Nivel 2.	Nivel 2.	Nivel 2.
5 a 6	Nivel 3.	Nivel 3.	Nivel 3.
7 a 8		Nivel 4.	Nivel 4.
9 a 10			Nivel 5.

4. Una vez se llena el cuadro con los descriptores de cada una de las categorías o aspectos evaluados, se puede pasar a mirar cómo se debe organizar la calificación con esta rúbrica si se quisiese dar una nota global luego de aplicar dicha rúbrica, y además si dicha nota global se quisiera expresar, para el caso colombiano, en la escala valorativa que contempla el artículo 5 del Decreto 230 (MEN, 2002). Obviamente el uso de este tipo de rúbricas tiene todas las ventajas ya descritas y un poco más, puesto que son más abarcentes, complejas e integrales. Un ejemplo para la nota sería el de la tabla 3.

Tabla 3.
Modelo de tabla de puntuaciones y nivel correspondiente

Puntos	Nota
21 a 24	Excelente
16 a 20	Bien
11 a 15	Aceptable
6 a 10	Insuficiente
0 a 5	Deficiente

Otro aspecto importante en las rúbricas es que aquello que determina el nivel que un estudiante alcanza un aspecto determinado son los elementos involucrados en los criterios y no un número por el número mismo, es decir, es una evaluación más de corte cualitativo que cuantitativo, pues de lo que se trata es de determinar el grado de comprensión o de manejo o de desempeño de un estudiante en un aspecto particular. Lo que cuenta es el nivel de desarrollo y desde allí se define qué valoración dar. Un ejemplo de rúbrica analítica es el de la tabla 4.

Tabla 4.
Ejemplo de rúbrica analítica (Parques Naturales de Colombia)

Categoría	Excelente	Bueno	Satisfactorio	Requiere mejoras
Fuentes	Registros cuidadosos y precisos son mantenidos para documentar el origen de 95-100% de la información e imágenes en la presentación.	Registros cuidadosos y precisos son mantenidos para documentar el origen de 94-85% de la información e imágenes en la presentación.	Registros cuidadosos y precisos son mantenidos para documentar el origen de 84-75% de la información e imágenes en la presentación.	Las fuentes no son documentadas en forma precisa ni son registradas en mucha de la información o en las imágenes.
Atractivo y Organización	La presentación tiene un formato excepcionalmente atractivo y una información bien organizada.	La presentación tiene un formato atractivo y una información bien organizada.	La presentación tiene la información bien organizada.	El formato de la presentación y la organización del material es confuso para el asistente.
Mapas y Fotografías	Los mapas y fotografías van bien con el texto y hay una buena combinación de texto y gráficos.	Los mapas y fotografías van bien con el texto, pero hay muchos que se desvían del mismo.	Las mapas y fotografías van bien con el título, pero hay muy pocos y el folleto parece tener un "texto pesado" para leer.	Los mapas y fotografías no van con el texto pero aparentan haber sido escogidas sin ningún orden.
Ortografía y Revisión	No quedan errores ortográficos después de que otra persona además del mecanógrafo lee y corrige la presentación.	No queda más que 1 error ortográfico después de que otra persona además del mecanógrafo lee y corrige la presentación.	No quedan más que 3 errores ortográficos después de que otra persona además del mecanógrafo lee y corrige la presentación.	Quedan varios errores de ortografía en la presentación.
Claridad	Cada sección en la presentación tiene una introducción, un desarrollo y una conclusión claro.	Casi todas las secciones de la presentación tienen una introducción, un desarrollo y una conclusión claros.	La mayor parte de las secciones en la presentación tienen una introducción, un desarrollo y una conclusión claros.	Menos de la mitad de las secciones de la presentación tienen una introducción, un desarrollo y una conclusión claros.
Argumentación	La presentación impacta, es convincente e invita a la acción de los asistentes.	La presentación es convincente e invita a la acción de los asistentes.	La presentación es convincente pero no invita a la acción de los asistentes.	La presentación no es convincente ni invita a la acción de los asistentes.
Pertinencia	Toda la información, mapas y fotografías de la presentación se refieren a los parques a estudiar y sus respectivos ecosistemas.	Entre el 95-100% de la información, mapas y fotografías de la presentación se refieren a los parques a estudiar y sus respectivos ecosistemas.	Entre el 90-94% de la información, mapas y fotografías de la presentación se refieren a los parques a estudiar y sus respectivos ecosistemas.	Menos del 90% de la información, mapas y fotografías de la presentación se refieren a los parques a estudiar y sus respectivos ecosistemas.
Escritura-Gramática	No hay errores gramaticales en la presentación.	No hay errores gramaticales en la presentación después de la revisión por parte de un adulto.	Hay 1-2 errores gramaticales en la presentación aún después de la revisión por parte de un adulto.	Hay varios errores gramaticales en la presentación aún después de la revisión por parte de un adulto
Contenido-Precisión	Toda la información en la presentación es correcta.	99-90% de la información en la presentación es correcta.	89-80% de la información en la presentación es correcta.	Menos del 80% de la información en la presentación es correcta.
Conocimiento Ganado	Todos los estudiantes en el grupo pueden contestar adecuadamente todas las preguntas relacionadas con la información y el proceso técnico usado para crearlo.	Todos los estudiantes en el grupo pueden contestar adecuadamente la mayoría de las preguntas relacionadas con la información y el proceso técnico usado para crearlo.	Algunos estudiantes en el grupo pueden contestar adecuadamente algunas de las preguntas relacionadas con la información en la presentación y el proceso técnico usado para crearlo.	Varios estudiantes en el grupo parecen tener poco conocimiento sobre la información y procesos técnicos usados en la presentación.

Categoría	Excelente	Bueno	Satisfactorio	Requiere mejoras
Culminación	Se completó el informe con todas las actividades.	Se completó el informe pero quedaron actividades deficientes.	El informe quedó incompleto. Faltaron 2 actividades o menos.	El informe quedó incompleto. Faltaron más de 2 actividades.
Esfuerzo	El trabajo final demuestra que los estudiantes se esforzaron al máximo.	El trabajo final demuestra que los estudiantes no alcanzaron a poner todo su esfuerzo.	El trabajo final demuestra que faltó esfuerzo.	El trabajo final demuestra que los estudiantes no pusieron ningún esfuerzo.

Fuente: <http://www.eduteka.org/proyectos/RubricParques.php3>

Finalmente se puede afirmar también que cuando se evalúa con rúbricas, el estudiante entiende porqué razón obtiene una determinada nota, qué es capaz de hacer y qué le falta para ir al siguiente nivel o al más superior. Se gana en objetividad y sobre todo, se incluye un aspecto que es importante en la evaluación y que tiene que ver con proporcionar la información suficiente o retroalimentar para que el estudiante sepa qué puede hacer para avanzar en su proceso.

Estas consideraciones pueden ayudar a un docente a emplear las rúbricas en su proceso de evaluación y además le dan pistas para que pueda construir sus propias matrices de valoración. Hay que tener en cuenta que una rúbrica siempre será un documento o un texto perfectible que se encuentra en construcción y que como tal la práctica del maestro hará que la misma esté mejor o peor construida. De lo que se trata es de incursionar en el uso de esta herramienta de evaluación para mejorar nuestra práctica evaluativa y desde ésta, por su puesto, lo más importante: La enseñanza.

Referencias

- Díaz Barriga, Frida (2005). *Enseñanza situada: Vínculo entre la escuela y la vida*. México: McGraw Hill.
- Goodrich, H. *Understanding Rubrics*. Recuperado del sitio <http://learnweb.harvard.edu/alps/thinking/docs/rubricar.htm>
- Rodríguez, D. *Rúbricas: Avalúo y retroalimentación efectiva en el salón de clase*. Recuperado del sitio <http://www.uprm.edu/ideal/rubricas.pdf>
- Mertler, Craig A. (2001). Designing scoring rubrics for your classroom. *Practical Assessment, Research & Evaluation*, 7(25). Recuperado del sitio <http://pareonline.net/getvn.asp?v=7&n=25>
- Ministerio de Educación Nacional (2002). *Decreto 230 de 2002*. Santafé de Bogotá: Imprenta Nacional.
- Ministerio de Educación Nacional. (1997). *La evaluación en el aula y más allá de ella: Lineamientos para la educación preescolar, básica y media*. Santafé de Bogotá: Ministerio de Educación Nacional.
- WebQuest. *Colombia y sus Bellezas Naturales para grados de 6-8*. Recuperado del sitio <http://www.eduteka.org/proyectos/RubricParques.php3>
- Vera, L. (2004, Octubre). *Rúbricas y listas de cotejo*. Recuperado del sitio <http://ponce.inter.edu/cai/reserva/vera/RUBRICAS.pdf>

Reseña

Prueba: ESCALA DE VALORACIÓN CUALITATIVA DEL DESARROLLO INFANTIL

Angélica Rocío Becerra Rojas*
Universidad Nacional de Colombia, Colombia

Torrado, M. Concha, M et al (1995) *Escala de Valoración cualitativa del desarrollo Infantil (desde el nacimiento hasta los seis años)* Bogotá, ICBF- UNICEF

La *Escala de Valoración Cualitativa del Desarrollo Infantil* fue el resultado de la investigación por varios años de la población infantil por parte del Instituto Colombiano de Bienestar Familiar, con el apoyo de entidades entre las que se encuentran Unicef, así como de profesionales en psicología y psicometría de las Universidades Nacional y Javeriana de la ciudad de Bogotá, y de esta forma responder necesidades prácticas y urgentes de esta población cuyo rango de edad se encuentra desde el nacimiento hasta los seis años. Para el desarrollo de la escala se tuvo en cuenta la concepción que el ICBF tiene sobre el desarrollo infantil. Ellos se enmarcan dentro de un modelo explicativo integrador de los procesos psicológicos, donde existe una interdependencia con la realidad, por lo que se puede decir –y como señalan explícitamente- se encuentran dentro de los modelos interaccionistas expuestos por Piaget, Bruner y Vigotsky. Por lo tanto se puede señalar que los procesos psicológicos son “indisociables del contexto socio-cultural” donde es muy importante la calidad de la relación que existe entre el niño y el adulto que es responsable de éste.

Como argumentación teórica, se considera que el psiquismo es una construcción progresiva que pone de manifiesto una secuencia ordenada que lleva a etapas específicas. De esta forma, los autores se aproximan a la clasificación de la escala por *etapas* (Primera edad, infancia temprana, edad preescolar, y transición), *dimensiones* (Relación con el entorno, consigo mismo y con los demás), y procesos psicológicos relacionados con estas dimensiones que a su vez se presentan en *indicadores* (que hacen las veces de ítems) con distintos niveles de complejidad, indicadores que son evaluados de manera cualitativa dentro del contexto natural del niño, y permiten que cualquier niño menor de seis años pueda ser ubicado dentro de un perfil, a fin de reorientar las acciones pedagógicas y las relaciones y prácticas a nivel familiar, más no para establecer un diagnóstico clínico de desarrollo.

Un factor importante que cabe mencionar dentro de la presentación de esta escala es que comparte algunos aspectos importantes con los instrumentos psicométricos clásicos que le permiten tener una mayor credibilidad en cuanto a sus contenidos. Dentro de la metodología de la construcción de la escala se tuvo en la cuenta la validez entendida por ellos como el proceso de acumular evidencia para lograr de este modo hacer inferencias, asimismo como parte de los procesos que se hicieron para lograr esta validez se encuentran la aplicación de una prueba piloto de una muestra representativa a nivel nacional, la validación de contenido por parte de expertos para darle de este modo un alto nivel de representatividad de los indicadores que necesariamente son operacionalizadas para que puedan ser observables.

Por lo dicho anteriormente, podemos ver que aunque la escala no está referida a puntajes o patrones estándares ajustados a un grupo de edad o a la norma establecida, ocupa un lugar importante dentro de las escalas de evaluación del desarrollo infantil y se ajusta a la valoración de los criterios de desarrollo de la población infantil colombiana, para de esta manera realizar acciones que puedan ser emprendidas tanto por los maestros de los niños como por los adultos responsables de ellos de forma clara.

* Laboratorio de Psicometría. Universidad Nacional de Colombia. Correo electrónico: abecerraro@unal.edu.co

Reseña

TEST DE PERCEPCIÓN VERBAL NO MOTRIZ

César Merino Soto*

Universidad Privada San Juan Bautista, Perú

Marisol Angulo Ramos **

Universidad Nacional Federico Villarreal, Lima-Perú

Collarusso, R. P. & Hammill, D. D. (1980). *Test de Percepción Visual No Motriz* [TPVNM]. Buenos Aires: Panamericana.

Generalidades

El *Test de Percepción Verbal no Motriz* [TPVNM] fue creada por Ronald P. Colarusso y Donald D. Hammill (1972) y publicada en español en 1980, y se dirige a evaluar el funcionamiento de la percepción visual mediante tareas de rápida aplicación, es un instrumento fácil de corregir, de interpretación normativa y muy aceptable por el examinado. Las tareas que contiene la prueba representan cinco áreas, y fueron seleccionadas esencialmente desde un trabajo de revisión de las investigaciones realizadas en los Estados Unidos de Norteamérica en los años 60 sobre las disfunciones en el procesamiento central en niños (Chalfant & Scheffelin, 1969).

La versión presentada aquí se aplica desde los 4 años hasta los 8 años y 11 meses. Estas edades son consideradas críticas en el desarrollo de la habilidad perceptual visual, y la evaluación de ésta con otras habilidades como la integración visomotora (Koppitz, 1994). Esta versión del TPVNM puede ser aplicada en adultos con problemas de lesión cerebral u otras condiciones neurológicas comprometidas, y como se verá más adelante, su utilidad en estos grupos clínicos se orienta al diagnóstico y descripción (York & Cermak, 1995; Mazer, Korner-Bitensky & Sofer, 1998; Su et al., 2000)

Esta prueba enfatiza que la respuesta motora es reducida y no influencia en el desempeño del niño, por lo tanto se diferencia de otras pruebas que, publicadas alrededor de esa década, contenían tareas con algún tipo de desempeño motor. Junto a ello, la practicidad, la falta de información adecuada respecto a la validez y confiabilidad y la pobre estandarización normativa promovieron su creación y lanzamiento comercial. Aunque principalmente destinado para psicólogos, el manual informa su utilidad para profesores, especialistas en educación especial y otros profesionales vinculados a la exploración de la percepción visual mediante un instrumento válido.

Siguiendo la clasificación de Navas (2001), el TPVNM se puede identificar como una prueba cognitiva, de ejecución máxima, de potencia, psicométrica, individual, oral y de contenido no verbal. Respecto a estudios hispanos sobre la prueba casi no hay; aunque recientemente hay dos iniciativas en Latinoamérica, una de tipo exploratoria (Merino, Sotelo & Ramos, 2007), y otra normativa (Leal & Alarcón, 2000). A continuación, describiremos las características estructurales y técnicas de este instrumento.

* Dirección postal: Enrique Palacios 430, Chorrillos, Lima 9, Perú. E-mail: sikayax@yahoo.com.ar

** Email: marisolanguloramos@yahoo.es

Descripción

Materiales. El material se compone de un manual, hoja de registro y un cuadernillo de láminas, y complementariamente, bolígrafo para el examinador. Aunque el cronómetro no está incluido como material para el examinador, hará falta para el registro del tiempo total de desempeño del niño. Las láminas se presentan horizontalmente en cuadernillo anillado, un ítem por lámina. Los ítems se califican con 1 ó 0. El formato es de opción múltiple; cada ítem tiene 4 opciones de respuesta.

Administración. La administración es individual, por lo tanto se siguen todas las condiciones estándar respecto al ambiente físico y relacional de evaluación. El tiempo de respuesta aconsejable para cada ítem es 15 segundos. El test toma aproximadamente 15 minutos o menos, aunque esto depende de la edad del niño. En ocasiones el niño puede dar una respuesta que indica que no discrimina las demás opciones o que responde impulsivamente; o que señala la opción de la misma posición en cada ítem. Es recomendable enfatizar que vea todas las opciones antes de dar la respuesta final.

Cada una de las cinco partes tiene una instrucción moderadamente diferente, y un ejemplo que obligatoriamente el examinador aplica. En el primer subtest (ítems del 1 al 8) se pide que encuentre entre las opciones la que es igual al modelo que se ubica sobre ellas en la lámina, desde la lámina 4 al 8 ocurre un cambio de instrucción, pues se solicita al niño que halle la figura escondida entre líneas que confunden la respuesta correcta. En el segundo subtest (ítems del 9 al 13) se pide lo mismo, pero se añade una aclaración sobre las variaciones que puede tener la respuesta respecto al modelo presentado.

En el tercer subtest (ítems del 14 al 21) se pide que el niño observe una figura y luego de pasar a otra página, se le solicita que elija entre las opciones la figura que le fue mostrada. En los siguientes subtest el formato de instrucción cambia; en el cuarto subtest (ítems del 22 al 32) se le solicita que imagine qué forma tendría el modelo incompleto si fuese completado, y que ubique la figura completada entre las opciones de respuesta. Finalmente, en el quinto subtest (ítems del 33 al 36) se requiere que el niño halle la figura que es diferente del resto.

Calificación. El puntaje total se obtiene de la suma simple de los aciertos en todos los ítems. El puntaje máximo es 36 y el mínimo posible es 0. Aunque hay 5 tipos diferentes de áreas cubiertas en la prueba total, los potenciales puntajes obtenidos de ellas no forman cinco puntajes independientes; el número reducido de ítems en cada parte restringen su uso como mediciones independientes de las áreas que exploran. En la Tabla 1, se presenta la organización de las tareas dentro de la prueba.

Tabla 1
Lista de áreas, número de ítems y ubicación de éstos en el TPVNM

Áreas	Nro ítems	Ubicación
Figura-Fondo	8	1 al 8
Discriminación visual	5	9 al 13
Memoria visual	8	14 al 21
Cierre visual	11	22 al 32
Relaciones espaciales	4	33 al 36

Normas e interpretación. 883 niños entre 4 y 8 años y 11 meses, normales, extraídos de 22 estados en los Estados Unidos de Norteamérica. El manual no informa más sobre la distribución del género, zona de residencia, tipo de colegio y grupo étnico en la muestra de estandarización, así como tampoco el método de selección muestral.

Un análisis de varianza especificando las variables demográficas de zona geográfica y género de los niños no arrojó diferencias estadísticamente significativas entre los grupos de edad pre-definidos, así que los autores elaboraron una sola norma. Cada grupo de edad desde los 5 años hasta los 8 años y 11 meses tuvo un intervalo de 6 meses; el primer grupo de edad tuvo un rango de 12 meses (4:0 hasta 4:11).

Después de las normas originalmente publicadas en Estados Unidos, un reciente estudio en curso en 183 niños desde 6 años hasta 9 años reportó pequeñas diferencias en términos de magnitud del efecto y diferencias estadísticamente no significativas comparadas con la muestra de estandarización americana (Merino et al., 2007); desde este estudio preliminar, las normas provenientes de la muestra americana son moderadamente comparables con la muestra peruana, y éstas pueden ser consideradas como normas itinerantes (Butcher & García, 1978). Leal y Alarcón (2000) en Chile proporcionaron normas locales debido a las diferencias estadísticamente significativas en los niveles de edad desde los 5 hasta los 6 años. Estos estudios latinoamericanos demostraron diversos patrones de diferencias normativas comparadas con la muestra original.

Los puntajes interpretativos se presentan como coeficientes perceptuales y edad perceptual. Los cocientes perceptuales son puntajes de desviación, centrados en una media de 100 y una desviación estándar de 15. Los autores enfatizan que la interpretación del puntaje se acompañe del error estándar de medición (*e.e.m.*). Por otro lado, la edad perceptual se obtiene por la edad promedio de los niños que obtuvieron el puntaje directo de interés. El manual también informa de intervalos de confianza para cada edad perceptual asociado a cada puntaje directo posible.

Los autores recomiendan que la interpretación individual de los puntajes directos menores de 10 no se interprete como expresión cuantitativa del constructo medido. Aparentemente, la interpretación para un niño en particular se compromete debido que niveles bajos de puntajes provendrían principalmente de desempeños aleatorios. Sin embargo, para investigaciones en que la interpretación del desempeño grupal es el objetivo, esta advertencia podría ser pasada por alto (Collaruso & Hammill, 1980).

La interpretación basada en el número de aciertos tampoco es relevante, pues las tareas elegidas no tienen en común algún contenido curricular que deba ser dominado y representado dentro de un marco de interpretación criterial. Además, el patrón de errores no puede ser homogéneo de acuerdo al contenido, ya que el número de ítems por área es pequeño como para ser interpretativamente útil.

Confiabilidad. La confiabilidad se examinó por los métodos de estabilidad y consistencia interna; esta última desde el KR-20 y mitades. Tomando un intervalo de 20 días, los coeficientes de estabilidad variaron entre 0.77 y 0.83 en una muestra de 162 niños de 4 a 8 años (muestra total = 0.81). Por otro lado, los coeficientes de consistencia interna para el método de mitades variaron entre 0.81 y 0.84 (muestra total = 0.88); y usando el método KR-20, la confiabilidad varió desde 0.71 a 0.82 (muestra total = 0.86). En el estudio latinoamericano peruano, los niveles reportados tienden a ser adecuados (muestra total = 0.75, I.C. 95%: 0.69 – 0.79) y estadísticamente no diferentes (Merino et al., 2007) a los niveles excelentes reportados en el manual; sin embargo, en el manual los índices son más elevados y más similares a los hallados en el estudio chileno (Leal & Alarcón, 2000). En general, los reportes de confiabilidad tienden a ser desde adecuados a excelentes, y en las muestras latinoamericanas, la confiabilidad es comparable estadísticamente pero relativamente baja.

Validez. La validez de contenido tiene como evidencias la clasificación de las habilidades perceptuales más citadas basadas en la revisión de Chalfant y Scheffelin (1969) y el análisis de ítems. Desde esta revisión, se identificaron y seleccionaron cinco áreas de percepción visual. Por otro lado, la dificultad y la validez de los ítems seleccionados en la forma final (36 ítems) respondieron casi totalmente a los criterios de selección.

En la validez de constructo, el manual reporta correlaciones con pruebas de inteligencia, de rendimiento, aptitudes y de visomotricidad en varias muestras de niños que variaron entre 35 y 107 en cada grupo. La varianza entre otras medidas de percepción visual fue de mayor magnitud ($r_{\text{mediana}} = 0.49$) que las demás pruebas ($r_{\text{mediana}} = 0.38$ y 0.31 con pruebas de rendimiento e inteligencia, respectivamente), lo que corroboró la hipótesis del constructo medido. De mayor interés para el psicólogo escolar es saber que las habilidades involucradas con tareas de apareamiento visual, escritura, identificación de números y algunas tareas relacionadas con la lectura tuvieron correlaciones alrededor de 0.41. Con la prueba de Frostig, la correlación fue alta ($r = 0.73$) y se concluyó que pueden usarse de manera equivalente, considerando la ventaja que el TPVNM es más breve y de menor entrenamiento para su administración.

La validez de constructo se obtuvo posteriormente mediante estudios independientes desde las cuales se tienen evidencias desde la diferenciación de grupos conocidos obtenidos en el área de la rehabilitación ocupacional de adultos con lesiones cerebrales (York & Cermak, 1995; Mazer et al., 1998; Su et al., 2000). En estos estudios, los puntajes bajos en el TPVNM diferenciaron tipos de pacientes y lesiones frente a grupos de control y grupos con otras localizaciones de las lesiones. En esta misma área de pacientes, las correlaciones convergentes/divergentes entre el TPVNM y medidas perceptuales y de funcionamiento visual desde baterías complejas estuvieron entre niveles adecuados y excelentes (Su et al., 2000).

La validez concurrente fue corroborada con un cuestionario de detección de déficits en la percepción visual (Cornoldi, Venneri, Marconato, Molin, & Montinari, 2003) reportado por profesores escolares, rendimiento escolar y habilidades visomotoras (Merino et al., 2007); aquí la varianza compartida entre el cuestionario y el TPVNM fue mayor que un test visomotor, lo que corroboró que ambos evalúan el mismo funcionamiento pero desde métodos diferentes.

En general, las evidencias de validez dan un respaldo sólido de la medición del constructo en muestras de niños y adultos con lesión cerebral y adultos mayores.

Comentarios

Ya que es una prueba breve y de exploración amplia del funcionamiento de la percepción visual, no debería usarse independientemente de otras fuentes para determinar el estatus de esta función perceptual. Es muy útil como instrumento que confirme impresiones iniciales del examinador respecto a los déficits que puede presentar el examinado. Otras baterías complejas contienen subtests de percepción visual que pueden ser usados como un primer dato diagnóstico del examinado; de este modo, un bajo puntaje en estos subtests deberían ser motivo para referir a una evaluación usando el TPVNM, y que puede ser determinante para la diferenciación de niños con problemas en el área de la percepción visual.

El tiempo de respuesta total a través de todos los ítems se podría registrar, y sería una información clínica muy útil que mejoraría la interpretación normativa del puntaje total, ya que el tiempo registrado está influenciado por la demora del niño para responder a cada ítem, pero otro factor es el tiempo que al examinador le toma cambiar las láminas. Este cambio está asociado también al registro de la respuesta, pues el examinador debe atender a estas dos acciones casi simultáneamente. Por lo tanto, el examinador puede ser una fuente de error en la interpretación del tiempo de respuesta. Se sugiere que el examinador maneje el material con la menor demora posible, y que se entrene en aplicaciones previas, pues la familiaridad con la hoja de registro y el cambio de lámina tendrá menos impacto en la interpretación del tiempo de respuesta total. Mientras la versión explicada aquí no reporta el tiempo de desempeño, la edición actual incluye el registro del tiempo total de desempeño para obtener una información que puede servir a la interpretación clínica. En la investigación de la adaptación normativa comentada párrafos atrás (Merino et al., 2007), el registro del tiempo se está incluyendo y complementará la interpretación del puntaje estandarizado.

Originalmente, la versión aquí presentada sólo informa de puntajes de tipo CI (media 100 y desviación estándar de 15) y edades perceptuales. Sin embargo, la información para la interpretación también puede ser presentada en otros puntajes estandarizados, como puntajes T, eneatis, percentiles o equivalentes de curva normal. Los puntajes estandarizados se transformaron linealmente, pero el manual no informa si la distribución del desempeño de los niños se desvió de la curva normal. Un examen de este aspecto estadístico en nuevas muestras de estandarización permitirá ajustarlas a la curva normal y obtener inferencias válidas sobre el puntaje de los participantes y su ubicación dentro de una curva normal (Charter & Dobbs, 1998). Un estudio preliminar en Latinoamérica observó este aspecto estadístico y sugirió una transformación no lineal de los puntajes para ajustarlas a la normalidad (Merino et al., 2007).

Aunque los autores (Collaruso & Hammill, 1980) desarrollan un ejemplo usando el e.e.m. promedio a los grupos de edad, la actual práctica recomienda usar el e.e.m. correspondiente a cada subgrupo desde el cual se ha obtenido la confiabilidad. La idea de reportar la confiabilidad condicional al nivel de edad permite construir e.e.m e intervalos de confianza más apropiados.

Actualmente, hay una edición revisada y ampliada hasta personas de 80 años de edad. Aparentemente, la versión de 1972 puede ser innecesaria existiendo una versión actual, pero el uso de la primera versión aún se reporta en artículos empíricos, especialmente en el área de la rehabilitación ocupacional y pacientes con lesiones cerebrales (York & Cermak, 1995; Mazer et al., 1998; Su et al., 2000). Por ejemplo, Mazer et al., (1998) reportaron que los bajos puntajes en el TPVNM tienden a ser más predictivos que otros test que evalúan habilidades asociadas a la conducta de manejar carros, específicamente en pacientes con lesiones traumáticas en el cerebro. Esta primera versión posee por lo tanto un poder discriminativo de desempeños provenientes de poblaciones esperablemente atípicas, y su información psicométrica aún se mantiene robusta dentro de rangos de adecuabilidad para uso de investigación de grupo o individual.

Referencias

- Butcher, J. N. & García, R. E. (1978). Cross-national application of psychological tests. *Personnel and Guidance Journal*, 56(8), 472-475.
- Chalfant, J. C. & Scheffelin, M. A. (1969). *Task force III. Central processing dysfunctions in children: A review of research*. Bethesda, MD: US Department of Health, Education and Welfare. Citado en Colaruso y Hammill (1972; 1980)
- Charter, R. A., & Dobbs, S. M. (1998). Long and short forms of the Speech-Sounds Perception Test: Item analysis and age and education corrections. *The Clinical Neuropsychologist*, 12(2), 213-216.
- Colaruso, R. P. & Hammill, D. D. (1972). *Motor-free visual perception test*. Novato CA: Academic Therapy Publications.
- Colaruso, R. P. & Hammill, D. D. (1980). *Test de Percepción Visual No Motriz: TPVNM*. Buenos Aires: Panamericana.
- Cornoldi, C., Veneri, A., Marconato, F., Molin, A. & Montinari, C. (2003) A rapid screening measure for the identification of visuospatial learning disabilities in schools. *Journal of Learning Disabilities*, 36(4), 298-306.
- Koppitz, E. M. (1994) *El test gestáltico visomotor para niños*. Buenos Aires: Guadalupe.
- Leal, F. A. & Alarcón, A. (2000) Evaluación de las características psicométricas y obtención de normas locales para el TPVNM de Colaruso y Hammill en escolares de la ciudad de Arica (Chile). *Revista Iberoamericana de Diagnóstico y Evaluación Psicológica*, 9, 91 – 103.
- Mazer, B. L., Korner-Bitensky, N. & Sofer, S. (1998). Predicting ability to drive after stroke. *Archive of Physical and Medicine Rehabilitation*, 79, 743-750.
- Merino, C., Sotelo, L. & Ramos, M. (2007). *Propiedades psicométricas del Test de Percepción Visual No Motor*. Documento no publicado.
- Navas, M. (2001). La medición de los psicológicos. In M. Navas (Ed.) *Métodos, diseños y técnicas de investigación psicológica* (pp. 59-88). Madrid: UNED.
- Su, C. Y., Charm, J.J., Chen, H. M., Su, C. J., Chien, T. H. & Huang, M-H. (2000). Perceptual differences between stroke patients with cerebral infarction and intracerebral hemorrhage. *Archive of Physical and Medicine Rehabilitation*, 81, 706-714
- York, C. D. & Cermak, S. A. (1995). Visual perception and praxis in adults after stroke. *American Journal of Occupational Therapy*, 49(6), 543-550.

Reseña

¿EN QUÉ CONSISTE LA QUINTA EDICIÓN DEL D.A.T? UNA APROXIMACIÓN ACTUALIZADA A LA MEDICIÓN DE ACTITUDES¹

Felipe Andrés Velásquez Ospina*

Servicio de Atención Psicológica, Universidad Nacional de Colombia, Colombia

Prueba: Cordero, A. & Corral, S (2006)- *DAT – 5 Tests de Aptitudes Diferenciales: Manual (2ª ed)*. Madrid, TEA Ediciones S. A.

La *Batería de Pruebas de Aptitudes Diferenciales DAT* fue diseñada en 1947 por Bennett, Seashore y Wesman para medir la capacidad que los estudiantes tendrán para aprender o actuar en determinadas áreas de conocimiento (Esquivel, Lozano & Peralta, 2004). Originalmente diseñada para jóvenes en cursos de educación media, es empleada con frecuencia en procesos de orientación vocacional, consejería educativa y en contextos de selección laboral por la utilidad de datos que suministra. Ha sido editada en cinco ocasiones (1947, 1962, 1972, 1980, 1990), y la edición más reciente fue realizada por el Equipo Técnico de The Psychological Corporation con sede en San Antonio, Texas (Esquivel et. al, 2004; Cordero & Corral, 2006). Para la validación de dicha edición se tuvieron en cuenta los mismos parámetros con los que se validaron las ediciones anteriores. Tiene dos versiones reconocidas en el idioma castellano: La española, producida por TEA Ediciones, cuyo nombre es “Tests de Aptitudes Diferenciales, Versión 5 (DAT-5) ®” (Cordero & Corral, 2006), y la mexicana, elaborada por Manual Moderno y denominada “La Prueba Diferencial de Aptitudes DAT 2®” (Esquivel et. al, 2004). Esta última no es la quinta edición mexicana, pues de las cinco ediciones del DAT sólo se han generado dos versiones en dicho país.

Al ser una prueba revisada en cuatro ocasiones, se esperaba encontrar en su quinta edición un concepto de aptitud que dé cuenta de las evoluciones teóricas que se han suscitado a su alrededor y de los cambios en su medición. En esa medida, el propósito del presente texto es mostrar un panorama general sobre lo que esta batería (a la que se denominará en adelante DAT5) puede ofrecer para la medición de las aptitudes. Ello implica en primera instancia presentar la noción de aptitud que es empleada en la prueba, así como conceptos que puedan relacionarse con la misma; en segundo lugar, dar a conocer características técnicas de esta prueba en las versiones en idioma castellano; como tercer punto, exponer las aplicaciones que esta prueba ha tenido, especialmente en la predicción del rendimiento académico puesto que ha resultado muy efectiva en tal dirección; y, finalmente, a partir de este panorama, plantear algunas cuestiones relacionadas con la aplicación de esta prueba en Colombia.

Concepto de Aptitud

Este concepto tiene origen en el marco de referencia de los estudios en inteligencia. El concepto de inteligencia acuñado por Binet y Simon (1916, citado por Ackerman, 2003) era considerado

¹ Este trabajo se enmarca dentro de la línea de Investigación en Orientación Vocacional que se desarrolla en el Servicio de Atención Psicológica de la Universidad Nacional de Colombia. Agradecimientos para Érika Arias, quien proporcionó los espacios para darle vida a esta reseña durante mi paso como practicante en el Servicio de Atención Psicológica.

* Servicio de Atención Psicológica. Universidad Nacional de Colombia. E-mail: favelasquezo@unal.edu.co

independientemente de la instrucción² recibida por una persona, y como tal, se medía a partir de capacidades de juicio, memoria y un conocimiento general. En esa medida, no se consideró en su medición el grado de experiencia y experticia que una persona pudiera mostrar en dominios específicos. En la década de 1940, sin embargo, la necesidad de reconocer precisamente las diferencias individuales en dichos dominios con el fin de orientar a los jóvenes hacia una u otra área de conocimiento en la que pudiesen sentirse satisfechos y prepararse mejor, llevó a que se estudiaran diversos oficios y profesiones, lo que resultaría en la conclusión de que cada tarea requiere unas capacidades distintas, por ejemplo, los mecánicos automotrices necesitan conocer principios físicos y los abogados saber argumentar para desempeñarse adecuadamente en sus oficios (Esquivel et. al, 2004).

Wesman (1948, citado por Esquivel et. al, 2004), parte de la definición que se da en el Diccionario Warren de Psicología, en el que la aptitud se entiende como un conjunto de características relacionadas con la habilidad de una persona para adquirir, como fruto de una capacitación, conocimientos específicos o patrones de respuesta (como son la capacidad para hablar un idioma o producir música). Sobre esta base, Wesman considera que las aptitudes no son logros netamente hereditarios, sino que son posibles por la interacción entre dotes genéticas y elementos ambientales.

Este autor reconoce que tal definición es muy amplia y de ese modo pueden incluirse otros conceptos, tales como inteligencia, intereses y personalidad, aunque manifiesta que limitarla a capacidades especializadas (aprender técnicas útiles en la ingeniería, por ejemplo) resulta “engañoso”. Incluso afirma que una prueba de inteligencia puede medir aptitudes, ya que permite reconocer potencialidades académicas y vocacionales. De este modo, resume el concepto de aptitud como la capacidad para aprender.

Para Snow (1989, citado por Snow & Swanson, 1992), las aptitudes son los estados iniciales que reflejan la historia de aprendizaje en cada estudiante y tendrán influencia en los procesos de aprendizaje actuales. Este concepto, trabajado por este autor en el ámbito escolar, no se limita a habilidades cognitivas, pues incluye aspectos conativos y afectivos en esta historia de aprendizaje, estrategias para aprender de las diferentes situaciones, etc. Snow considera que analizar dichos estados permitiría sugerir formas de instruir a los estudiantes que se adapten a sus fortalezas y puedan responder a sus debilidades, o incluso remover dichas debilidades por medio de un entrenamiento directo. Tal análisis debería centrarse en las diferencias individuales que puedan existir en la adquisición, transferencia y organización de esas habilidades. Así, Snow y Swanson (1992) trabajan tres conceptos de aptitudes: Afectivas, conativas y cognitivas (sobre las que el presente escrito hará mayor énfasis, pues son las más relacionadas con las aptitudes que se miden en el DAT5).

Las aptitudes cognitivas son las que tradicionalmente han sido organizadas dentro de dominios componentes de un factor general (G) de inteligencia. La investigación en torno a las diferencias individuales de aprendizaje se ha enfocado cada vez más en los dominios y cobra especial importancia en el diseño de programas de enseñanza. De este cuerpo de investigación se destacan cinco dominios³:

Razonamiento Analítico Fluido (Gf): Da cuenta de la facilidad para generar, conectar y emplear estrategias en situaciones a las que es necesario adaptarse rápidamente y resolver problemas, especialmente si son novedosas.

Dominio Visual-Espacial (Gv): Es el dominio en el que se encuentra la capacidad para hacer operaciones sobre objetos y espacios, como rotación, inversión, duplicación, análisis y síntesis. Se le ha prestado mucha atención pues empíricamente se ha encontrado que es muy semejante a Gf (Lohman,

² Entendida como proceso de enseñanza de conocimientos, no como una orden específica.

³ Snow y Swanson (1992) presentan seis dominios, pero el sexto habla de creencias personales y resulta confuso para los propósitos de la presente reseña.

1988; en Snow & Swanson, 1992), de modo que su medición se ha refinado y por lo general implica análisis profundos en situaciones que invitan a emplear estrategias para lograr una visualización precisa de los objetos.

Capacidad Verbal Cristalizada (Gc): Este dominio está centrado en el uso de vocabulario y de la facilidad para la comprensión lectora, pero incluye otras habilidades verbales así como conocimientos lingüísticos. Se ha propuesto como un dominio independiente con respecto a otras formas de conocimiento, ya que la interpretación de un texto va ligada no sólo a la capacidad de comprenderlo sino de utilizar conocimiento previo para darle un sentido a esa comprensión.

Dominio Matemático: Este dominio hace su aparición debido a que en diferentes modelos factoriales resulta independiente a otros conocimientos cristalizados y a formas fluidas de razonamiento al especializarse en operaciones numéricas.

Capacidad de Memoria y Rapidez Mental: Aunque las capacidades para almacenar información y para procesarla rápidamente han sido trabajadas como independientes, se ha postulado que pueden ser dos caras de las múltiples subyacentes a una misma aptitud (Snow & Swanson, 1992). El almacenamiento de información es de vital importancia para otras aptitudes puesto que suministra el material para que puedan ponerse en marcha, y la rapidez en el procesamiento da más tiempo para que éste se refine (Kyllonen & Christal, 1990; Kyllonen et. al, 1992; ambos citados en Snow & Swanson, 1992).

Medición de las Aptitudes

Como se mencionó en el apartado anterior, la existencia de diferencias de habilidad exigida entre los oficios obligó a pensar de otra manera la organización de las capacidades humanas. En consecuencia, se piensa en hallar los factores que le resultan subyacentes, para lo que el mecanismo del análisis factorial ha resultado muy útil (Esquivel et. al, 2004). Se han encontrado así factores como la comprensión verbal, el manejo de números, la rapidez de percepción, la representación espacial entre otras, relacionadas en cierto modo entre sí.

Para Snow y Swanson (1992), la medición de las aptitudes aún requiere una amplia investigación. Acotan que la combinación de información sobre aptitudes conativas, cognitivas y afectivas, junto con los hallazgos en la adquisición de conocimientos y habilidades, puede llevar a nuevas formas de medición de las aptitudes, pero que es necesario investigar más estos descubrimientos con el fin de conectarlos y tener un constructo que pueda medirse efectivamente.

La combinación de los hallazgos en la investigación teórica existente sobre las aptitudes junto con la investigación acerca de la validez de las pruebas que pretenden medirlas lleva en la actualidad a la elaboración de pruebas de aptitud que se utilizan principalmente en orientación vocacional para jóvenes, así como en selección de personal y desarrollo humano en las empresas. Como consecuencia, se crean el DAT y otras pruebas de aptitud (Esquivel et. al, 2004). La Quinta Edición del DAT5, a partir de lo que conserva de la primera, se creó con la consideración de que la inteligencia humana está compuesta por aptitudes diferentes que deben medirse de modos distintos.

Estructura de la Batería

La Batería DAT5, construida con dos niveles para abarcar dos grupos de escolaridad distintos (Nivel 1, que se aplica hasta el 9° Grado, y el Nivel 2, empleado con población que por lo menos curse 10° en el momento de la aplicación), tiene 470 reactivos en total, repartidos en siete pruebas en el caso de la adaptación española, y 490 reactivos distribuidos en ocho pruebas en la adaptación mexicana. En cada una de estas pruebas sus condiciones de aplicación así como el número y la naturaleza de los reactivos son

diferentes con respecto a las demás. Pero la diferencia fundamental entre cada prueba es la aptitud que cada una pretende medir. A continuación se presenta cada una de estas pruebas en su orden de aplicación:

Razonamiento Verbal (VR en el DAT5 original): Esta prueba fue diseñada para medir la habilidad de descubrir relaciones entre palabras a través de analogías incompletas.

Razonamiento Numérico (NR en el DAT5 original): Esta prueba mide la habilidad para responder a tareas que exijan un razonamiento matemático a través de problemas con operaciones relativamente adecuadas a la escolaridad de los participantes.

Razonamiento Abstracto (AR en el DAT5 original): Esta prueba mide la habilidad para razonar sin recursos verbales, mediante la organización y continuación de series con dibujos geométricos que se le presentan a los respondientes.

Razonamiento Mecánico (MR): Mide la capacidad para comprender principios básicos y mecánicos en torno al movimiento, las propiedades de las herramientas y el funcionamiento de las máquinas por medio de ilustraciones que componen esta prueba.

Relaciones Espaciales (SR): Esta prueba mide la habilidad con la que se puede visualizar la transformación de un objeto en otro (En este caso, un objeto bidimensional en un objeto tridimensional).

Ortografía (SP): Esta prueba mide el grado en el que los estudiantes pueden deletrear correctamente palabras de uso común en el idioma (en el original, inglés; en las adaptaciones, castellano) mediante la detección de palabras incorrectas.

Velocidad y Exactitud Perceptual (PSA): Esta prueba está diseñada para medir la habilidad de comparación y comprobación de documentos escritos a gran velocidad mediante la búsqueda de sílabas iguales en documentos diferentes.

Uso del Lenguaje (LU): Con esta prueba se aprecia el manejo de reglas sintácticas y de puntuación mediante la detección de errores en la construcción de frases.

Adicional a estas ocho aptitudes, hay una aptitud más que se deriva de la suma de los puntajes directos de dos pruebas, que es la *Aptitud Académica* (VR + NR), o capacidad para aprender con base en información proveniente de bibliografía consultada y de enseñanzas de los docentes.

Estrategias de Construcción

El DAT5 fue construido a partir de los principios de la Teoría de Respuesta al Ítem (TRI), pues resultaban mucho más ventajosos para la clase de medición que se espera de esta prueba. Una virtud de la construcción del DAT5 desde la TRI, es que permite predecir la forma en que los estudiantes responderán a la prueba, pues tiene en cuenta la relación entre la habilidad de la persona que responde con la dificultad de cada uno de los reactivos. Además, en la TRI se busca que el resultado de la medición de una habilidad dada no dependa de la naturaleza del reactivo con el que se mida, ni que la dificultad del reactivo dependa de las tendencias demostradas por una población para responderlo. Dentro de estos principios, se basan en el modelo de Rasch, que evalúa la probabilidad de que una persona responda acertadamente a un reactivo a partir de la relación existente entre la habilidad de esa persona y la dificultad que presenta el reactivo (Esquivel et. al, 2004).

El DAT5 en sus versiones en castellano difiere notoriamente de la versión original debido a que tres de sus pruebas (*Razonamiento Verbal*, *Ortografía* y *Uso del Lenguaje*) debieron ser adaptadas para mantener su validez, ya que la traducción de los reactivos no la garantizaba. En la versión mexicana, se

construyeron nuevos reactivos para las pruebas de Razonamiento Verbal, Ortografía y Uso del Lenguaje, ya que el equipo técnico de IEGE encargado de su construcción consideró impertinente adaptar los reactivos al idioma castellano, pues perderían sus características gramaticales y pragmáticas. Se diseñaron reactivos que intentaban cumplir las especificaciones de cada una de las pruebas, con ayuda de los manuales técnicos de The Psychological Corporation para construir analogías (empleadas en Razonamiento Verbal), y con especial consideración de las reglas gramaticales del castellano y de la idiosincrasia latinoamericana. En la versión española también se reconstruyeron las pruebas de Razonamiento Verbal y de Ortografía, pero bajo condiciones distintas, y la prueba de Uso del Lenguaje se suprimió porque no se pudo garantizar una exigencia gramatical semejante a la de la prueba original en su rediseño.

La muestra con la que se estandarizó la prueba en México fue de 3027 personas: 1208 hombres y 1819 mujeres, pertenecientes a 47 planteles educativos que provenían de la Escuela Nacional Preparatoria, el Colegio de Ciencias y Humanidades y el Sistema incorporado de la UNAM (Esquivel et. al, 2004). La muestra española se hizo principalmente con población de entre 12 y 17 años de edad, que cursan entre 1° de ESO (equivalente a 6° grado) y 2° Bachillerato (equivalente a 11° grado), en la que se buscó representación para los dos géneros, las regiones del país (sectorizadas por puntos cardinales) y los tipos de institución (pública o privada). Por cada grado escolar se tomó una muestra de aproximadamente 350 estudiantes, y adicionalmente, se tomaron muestras de estudiantes de cursos de formación superior, así como profesionales, pero no se estandarizaron para el total de la batería (Cordero & Corral, 2006).

Características de Validez y Confiabilidad

Para cada una de las pruebas se calculó la consistencia interna de los reactivos mediante el algoritmo de Kuder- Richardson con poblaciones femenina, masculina y mixta. Se encontraron coeficientes en un rango de 0.75 a 0.92 para la versión española y en un rango de 0.71 a 0.9 para la mexicana.

En las dos versiones en castellano se hizo una correlación entre pruebas. En la versión española del DAT5, aunque la tendencia de coeficientes oscilaba entre 0.11 y 0.59, se encontraron coeficientes muy altos (entre 0.8 y 0.89) entre Razonamiento Verbal y Numérico, atribuido a la asociación de estas aptitudes, que se mide con el puntaje en Aptitud Académica (Cordero & Corral, 2006). En la versión mexicana, los coeficientes oscilaron entre 0.17 y 0.55, y de 0.6 a 0.62 para Uso de Lenguaje y Ortografía y Razonamiento Abstracto y Relaciones Espaciales, atribuidos en ambos casos a semejanzas entre los atributos que se miden (Esquivel et. al, 2004).

El DAT en sus cinco versiones también se ha correlacionado con otras pruebas de aptitud, como el Armed Services Vocational Battery, y el Preliminary Scholastic Aptitude Test. Los coeficientes obtenidos son superiores a 0.6, por lo que en principio se presume que el DAT5 está diseñado para medir aptitudes.

Conclusiones

El DAT5 puede considerarse una prueba muy efectiva para la medición de las aptitudes que Snow y Swanson (1992) llaman cognitivas, pues se acerca mucho a las necesidades teóricas para la medición de éstas, ya que las hace distintivas entre sí y permite evaluar qué tanta capacidad demuestra una persona en dominios concretos pero dentro de un todo que hable de la historia personal de aprendizaje, además la prueba presenta altos índices de confiabilidad y validez, y sus reactivos resultan muy pertinentes, ya que se construyeron con base en los principios de la TRI. También es de destacar que el DAT5 es un instrumento que en líneas generales está escrito en un lenguaje comprensible y su diseño es óptimo pues es más legible, ya que la letra de los reactivos es más grande con respecto a las ediciones anteriores y su diagramación facilita la lectura de los mismos; asimismo su presentación es muy práctica, puesto que

engloba a todas las pruebas en un sólo cuadernillo (en las anteriores, cada prueba tenía cuadernillo y hojas de respuesta aparte) por lo que son más fáciles de manejar y almacenar.

Dentro de los hallazgos empíricos sobre aptitudes, cabe señalar el empleo de una medida para el Razonamiento Analítico Fluido (Razonamiento Abstracto), otra para el Dominio Visual-Espacial (Relaciones Espaciales), para el Dominio Matemático (Razonamiento Numérico), tres que podrían emplearse para la Capacidad Verbal Cristalizada (Ortografía, Razonamiento Verbal y Uso del Lenguaje) y una que podría emplearse para la Rapidez Mental (Rapidez y Exactitud Perceptiva). Asimismo, confirma la relación existente entre el Razonamiento Analítico Fluido y el Dominio Visual-Espacial (Snow & Swanson, 1992), mediante la interrelación entre Razonamiento Abstracto y Relaciones Espaciales, y las múltiples facetas de la Capacidad Verbal Cristalizada con la interrelación entre Ortografía y Uso del Lenguaje.

Con respecto a Colombia, en el DAT5 es necesario adaptar las pruebas de Ortografía y Razonamiento Verbal, pues nuestro castellano es distinto al de España o México, debido a que las expresiones más frecuentes son distintas no sólo con respecto a estos países sino incluso entre regiones. Otra es intentar adaptar la prueba de Uso del Lenguaje, pues ésta mide un aspecto diferente al uso de reglas ortográficas y de relaciones semánticas y puede resultar importante para tener un perfil aptitudinal más amplio de las personas a las que se les aplica el DAT5.

Adicionalmente, puede considerarse que esta prueba, dadas sus características, aporta información muy útil dentro de un proceso de orientación vocacional sobre las capacidades personales de los estudiantes o bachilleres que tienen inquietudes al respecto, especialmente en lo que se refiere a procesos de razonamiento en diferentes campos. Sin embargo, las ocho aptitudes que mide esta prueba no son las únicas capacidades que pesan en las diferentes vocaciones, por lo que el DAT5 no debe ser la única medida de capacidades presente en un proceso de orientación vocacional. Aun más, tal como acotan Snow y Swanson (1992), esta medida debe acompañarse de otros aspectos personales como los intereses y los afectos para tener una mejor idea de lo que cada estudiante puede revelar de sí en su futuro académico y profesional.

En líneas generales, es una prueba recomendable para medir aptitudes ya que es confiable, válida, y fácil de manejar. Esto la hace muy efectiva en un proceso de orientación vocacional siempre y cuando se acompañe de otras pruebas durante dicho proceso, ya que las aptitudes, al menos como las presenta Wesman (1948; citado en Esquivel, Lozano & Peralta, 2004), son atributos muy complejos que requerirían diferentes formas para medirla, y el DAT5 sería solamente una de éstas.

Referencias

- Ackerman, P. (2003). Cognitive Ability and Non – Ability Trait Determinants of Expertise. *Educational Researcher*, 32 (8), 15 – 20.
- Cordero, A. & Corral, S. (2006). *DAT – 5 Tests de Aptitudes Diferenciales: Manual (2ª ed)*. Madrid, TEA Ediciones S. A.
- Esquivel, F, Lozano, R & Peralta, R. (2004). *DAT2 Prueba de Aptitudes Diferenciales: Manual Técnico*. Instituto de Evaluación a Gran Escala (De IEGE, 2004, publicación electrónica).
- Snow, R. & Swanson, J. (1992). Instructional psychology: Aptitude, adaptation and assessment. *Annual Reviews of Psychology*, 43, 583 – 626.

Obituario

JOHN WILDER TUKEY (1915 -2000)

Camilo Ernesto Monroy Muñoz*
Universidad Nacional de Colombia, Colombia

John Wilder Tukey fue uno de los estadísticos más influyentes de gran parte del siglo XX no solamente por ser el pionero del llamado *Análisis Exploratorio de Datos* (EDA por sus iniciales en inglés) y sus *técnicas robustas* sino por su “filosofía”. Estaba convencido de que la estadística se había desviado de su propósito original -la búsqueda de métodos de análisis de datos que describieran patrones, tendencias y relaciones y que detectaran anomalías- y que era necesario cerrar la brecha entre matemáticas y ciencias.



John W. Tukey. Tomado de la página web Memories of John W. Tukey <http://cm.belllabs.com/cm/ms/departments/sia/tukey/images.html>.

Inspirándose en su experiencia con los problemas reales y con los datos reales, John Tukey, a través del EDA, cambió el enfoque del análisis de datos, defendiendo un análisis de datos mucho más flexible, cuyo primer propósito es contestar a la pregunta ¿Qué dicen los datos?; dentro de tales análisis propuso procedimientos que permiten la representación gráfica de grupos de datos de manera clara y entendible para personas poco expertas en estadística, los más conocidos son el diagrama de cajas y el diagrama de tallos y hojas. En esta misma línea, desarrolló lo que se denominó el *Análisis Robusto* para trabajar con bases de datos “defectuosas” que no cumplen los supuestos de las técnicas de la estadística clásica;

No menos importantes fueron sus contribuciones en el análisis de series de tiempo y comparaciones múltiples, dentro de lo que llamó *Análisis Confirmatorio de Datos* (CDA por las iniciales en inglés) para diferenciarlo de sus técnicas exploratorias no probabilísticas; pa él, “una respuesta aproximada a una pregunta correcta es bastante más valiosa que una respuesta precisa a una pregunta errónea”, en este sentido defendió siempre la idea de que en ciencia hay que buscar la pregunta correcta antes que la respuesta. En el desarrollo de sus técnicas de análisis percibió la necesidad del uso de la computación, tanto en el cálculo de las fórmulas del análisis exploratorio como en el manejo de las masivas bases de datos y se le atribuye el uso pionero de los términos “bit” y “software” en la literatura científica.

Todos estos trabajos y algunos otros en los campos del Análisis de Varianza, la Regresión, y el Análisis Factorial, han tenido un profundo impacto en campos tan diversos como la ingeniería eléctrica, la investigación de operaciones, la gestión de riesgos comerciales y financieros y las ciencias sociales. En el campo de la ingeniería es conocido además, por el desarrollo, junto con James Cooley, del algoritmo de la Transformada rápida de Fourier (*Fast Fourier Transform*), aplicable en ciencias físicas e ingeniería.

* E-mail: cemonroy@unal.edu.co

Tukey nació en 1915 en New Bedford, Massachusetts, se graduó como Master en Química en 1937 en la Universidad de Brown. Posteriormente, con el propósito de obtener el doctorado, viajó a la Universidad de Princeton, pero debido a que no le permitían ser asistente de laboratorio, optó por hacer un Ph. D. en matemáticas. Su tesis doctoral, llamada "*Denumerability in topology*" le permitió graduarse en 1939. Inmediatamente después fue nombrado instructor en matemáticas y una década más tarde fue ascendido a la categoría de profesor de tiempo completo. Fue el director fundador del *Princeton's Statistical Techniques Research Group* en 1956, vicepresidente de la Sociedad Filosófica Americana entre 1975 y 1977, y presidente fundador del Departamento de Estadística de Princeton, cargo que ocupó hasta 1970. Por otro lado, trabajó en los Laboratorios AT & T Bell desde 1945 como miembro del personal técnico, para posteriormente avanzar al cargo de Director Asistente de Investigación en 1958 y al cargo de Director Ejecutivo Asociado en 1961, cargo que mantuvo hasta su retiro en 1985, año de su también retiro como profesor de la Universidad de Princeton.

Colaboró igualmente con el gobierno estadounidense durante la Segunda Guerra Mundial en el *Fire Control Research Office* (FCRO) encargado de la investigación sobre el control de fuego de artillería, colaborando también en el descifrado de códigos secretos. La experiencia en el FCRO y su formación anterior en química lo impulsaron a seguir el camino de la estadística, ya que el estar en contacto con problemas y datos del mundo real le hizo tomar conciencia de la importancia del análisis de los mismos, campo por el que es reconocido mundialmente. Sirvió asimismo al gobierno de su país en la conferencia sobre la discontinuación de pruebas de armas nucleares y en las primeras investigaciones realizadas sobre el daño a la capa de ozono y sobre la contaminación del aire. En el campo educativo fue asesor del *Educational Testing Service* (ETS) y participó en 1950 en la revisión de los problemas estadísticos de los estudios sobre la conducta sexual del hombre realizados por el famoso sexólogo Alfred Kinsey.

Entre los principales reconocimientos que recibió están el Premio "Estadístico del Año" de la *American Statistical Association* (ASA) en 1956, la Medalla S. S. Wilks de la ASA en 1965, la Medalla Nacional de Ciencia, el "Premio al Servicio" del ETS, la Membresía de la *National Academy of Sciences*, de la *American Philosophical Society*, de la *American Academy of Arts and Sciences*, y de la *Royal Society of London*. También obtuvo doctorados honoríficos del *Case Institute of Technology*, de la Universidad de Brown, de la Universidad de Yale, de la Universidad de Chicago, de la Universidad de Princeton y de la Universidad de Waterloo.

John Wilder Tukey, one of the most influential statisticians of the last 50 years and a wide-ranging thinker credited with inventing the word "software," died on Wednesday in New Brunswick, N.J. He was 85. The cause was a heart attack after a short illness, said Phyllis Anscombe, his sister-in-law.

By DAVID LEONHARDT, July 28, 2000 © The New York Times Company

INSTRUCCIONES PARA LOS AUTORES

La revista Avances en Medición publica artículos que representen contribuciones en los campos de la investigación, la teoría o la aplicación práctica de la medición y evaluación en las Ciencias del Comportamiento. Esta publicación recibe artículos de autores de cualquier nacionalidad, en idioma castellano, inglés o francés; no obstante, el idioma oficial de la revista es el castellano. El material debe enviarse en formato Microsoft-WORD a las direcciones electrónicas: ravances_fchbog@unal.edu.co, ó en impreso a: Laboratorio de Psicometría. Edificio 212. Oficina 229. Universidad Nacional de Colombia. Bogotá. Colombia. Se deben cumplir los siguientes criterios:

1. Para los artículos teóricos y empíricos se recomienda una extensión máxima de 40 páginas (incluidas tablas y figuras), escritas a doble espacio. Se deben evitar las notas a pie de página. Para el caso de los comentarios, reseñas bibliográficas y Noticias de Actualidad, la extensión máxima recomendada es de 5 páginas. En todo caso deben seguir normas APA.
2. Las tablas y figuras deberán aparecer en páginas separadas al final del documento. Deben estar diseñadas en blanco y negro a doble espacio.
3. Las citas bibliográficas deben estar intercaladas en el texto. Se debe colocar el apellido del autor y el año de publicación. La lista completa de las referencias aparecerá en orden alfabético al final del manuscrito.
4. En la primera página del manuscrito aparecerá el título del artículo (en inglés y en castellano), el(los) nombre(s) del(los) autor(es), así como la institución a la que están afiliados. Del mismo modo, se debe señalar el autor al cual contactar para correspondencia, indicando su dirección, apartado postal, teléfono, correo electrónico, y número de fax (si está disponible). En la segunda página se incluirá un resumen en inglés y en castellano, con una extensión de entre 100 y 150 palabras. En la tercera página aparecerá el título del artículo en idiomas inglés y castellano.
5. Para el caso de los artículos empíricos, se debe adjuntar toda la información que se considere crítica para la evaluación del documento. Se debe garantizar, ya sea por escrito o en la descripción de los procedimientos, que se han cumplido las normas éticas en investigación.

Los artículos que se envíen a la Revista Avances en Medición deben ser inéditos y no pueden considerarse para revisión de forma simultánea en más de una publicación. Una vez recibido el trabajo propuesto se somete a proceso de revisión por pares académicos a quienes se les solicita un concepto evaluativo sobre la calidad del trabajo, su presentación formal y su pertinencia y relevancia para los objetivos de la revista. La decisión final sobre la publicación del trabajo toma en consideración tales conceptos y se comunica al autor con un resumen de los comentarios recibidos. Al ser aceptado un artículo, los autores transfieren los derechos de éste a los editores de la revista Avances en Medición. No obstante, los autores podrán emplear el material en cualquier momento sin previa autorización. En todo caso los autores son responsables de los contenidos del artículo, así como de la obtención de los permisos correspondientes para el uso de material publicado en cualquier otra fuente.

INFORMACION PARA SUSCRIPTORES

Procedimiento para suscripción

Los precios para el año 2007 y 2008 sobre suscripciones a la Revista Avances en Medición, una publicación anual del Laboratorio de Psicometría y el Grupo Métodos e Instrumentos de Investigación en Salud, son los siguientes:

Suscripciones en Colombia

Instituciones: \$35.000

Individuos en Bogotá: \$12.000

Individuos fuera de Bogotá: \$24.000

Los pagos pueden hacerse:

Por medio de *Publicaciones Psicológicas Alejandro Pineda* Nit. 7.334.393-9, distribuidor autorizado de la Coordinación de Publicaciones de la Facultad de Ciencias Humanas de la Universidad Nacional de Colombia, consignando a la Cuenta de Ahorros de Bancolombia No. 18617689827 a nombre de Carlos Alejandro Pineda Roa.

Directamente con la Facultad de Ciencias Humanas consignando a la Cuenta No 01272006-6 del Banco Popular, a nombre del Fondo Especial de la Facultad de Ciencias Humanas.

Copia de la consignación junto con los datos completos del suscriptor deben enviarse a: Laboratorio de Psicometría Oficina 229 Edificio de Aulas de Ciencias Humanas (212) de la Universidad Nacional de Colombia sede Bogotá, Carrera 30 calle 45; o al telefax 3165000 Ext. 16311.

Igualmente lo invitamos a que visite nuestro sitio en Internet: www.humanas.unal.edu.co/psicometria o a que se dirija a las instalaciones del Laboratorio donde también puede adquirir la suscripción así como números previos.

Suscripciones fuera de Colombia

La información sobre suscripciones desde el extranjero puede solicitarse a través de los correos alejandropineda@rlpsi.org, capinedar@unal.edu.co.

Esta publicación se terminó de imprimir en el mes de Septiembre en UNIBIBLOS,
la editorial de la Universidad Nacional de Colombia con un tiraje de 500 ejemplares