

EVALUACIÓN DEL TEST DE EVALUACIÓN MATEMÁTICA TEMPRANA (TEMT)

RESUMEN DE LA VALORACIÓN DEL TEST

Descripción general

Características	Descripción
Nombre del test	Test de evaluación matemática temprana
Autor	J.E.H. van Luit, B.A.M. van de Rijt & A.H.Pennings
Autor de la adaptación española	J.E.H. van Luit, B.A.M. van de Rijt, José. I. Navarro, Manuel Aguilar, Concepción Alcalde, Esperanza Marchena, Gonzalo Ruiz, Inmaculada Menacho y Manuel G. Sedeño.
Variable(s)	Conocimiento numérico
Área de aplicación	Psicología educativa
Soporte	Administración oral y manipulativa.

Valoración general

Características	Valoración	Puntuación
Materiales y documentación	Excelente	5
Fundamentación teórica	Excelente	5
Adaptación	Excelente	5
Análisis de los ítems	Buena-excelente	4,5
Validez: contenido	Buena	4
Validez: relación con otras variables	Adecuada-buena	3,5
Validez: estructura interna	Adecuada-buena	3,5
Validez: análisis del DIF	-	-
Fiabilidad: equivalencia	Adecuada	3
Fiabilidad: consistencia interna	Buena-excelente	4,5
Fiabilidad: estabilidad	Adecuada	3
Fiabilidad: TRI	-	-
Fiabilidad: inter-jueces	-	-
Baremos e interpretación de puntuaciones	Buena-excelente	4,5

Comentarios generales

El TEMT evalúa el nivel de competencia matemática temprana en niños de 4 a 7 años, requisito imprescindible para poder seguir en los años siguientes una educación matemática formal. Disponer de un instrumento de estas características adaptado al contexto español es sin duda de gran utilidad para la adaptación escolar de alumnado con dificultades de aprendizaje. El test dispone de tres versiones paralelas, de 40 ítems cada una, que se aplican aproximadamente en unos 30 minutos. Los ítems se agrupan en 8 tareas diferentes, por lo que cada componente es evaluado con 5 ítems. Permite analizar el conocimiento numérico temprano en sus ocho componentes: Conceptos de comparación; Clasificación; Correspondencia uno a uno; Seriación; Conteo (verbal, estructurado y resultante) y Conocimiento general de los números. Los cuatro primeros subtests evalúan habilidades de tipo piagetiano, mientras que los cuatro últimos estiman las habilidades numéricas de naturaleza más cognitiva.

Todos los ítems son presentados oralmente y los niños responden señalando en un material con dibujos o, en el caso de las tareas de contar y de numeración, manipulando pequeños cubos. En otros tres ítems el alumno debe usar el lápiz para unir los objetos del dibujo presentado.

Es de destacar el diseño de los materiales empleados para realizar las distintas tareas incluidas en los ítems, pues han sido especialmente elaborados para el test y son atractivos para el alumnado de estas edades. Además, las pruebas son muy fáciles de aplicar, siendo posible incluso para un maestro/a que dedique un tiempo a leer el material y el manual aplicar la prueba de forma acertada sin ningún problema. La corrección de la prueba y el análisis de los resultados están también explicados de forma muy amena y pedagógica.

El test se aplica de forma individual y, comparando el resultado de cada niño con un grupo normativo, nos ofrece un nivel de competencia matemática temprana. Además, proporciona una información (cuantitativa y cualitativa) que puede ser usada para realizar una detección precoz del alumnado con dificultades de aprendizaje relacionadas con los números. Estas características hacen de este test una prueba especialmente adecuada para analizar las competencias matemáticas y poder tomar medidas preventivas en los casos en los que el alumnado comience a presentar problemas ya en esta área. Esta detección temprana, que requeriría únicamente 30 minutos, permitiría reducir las altas tasas de fracaso escolar y bajo rendimiento académico en esta disciplina, uno de los problemas más acuciantes de nuestro sistema escolar.

Otro de los puntos fuertes del test es que permite recoger datos de tipo cuantitativo, pero también de tipo cualitativo. El propio manual “entrena” al aplicador para que anote las diferentes estrategias de resolución de los ítems llevadas a cabo por los niños/as evaluados. Esta información puede ser de gran utilidad para analizar los procesos de resolución de problemas puestos en funcionamiento por los sujetos evaluados, pues a menudo los niños dan una respuesta (correcta o incorrecta) a una pregunta matemática, pero sólo cuando observamos los errores, el procedimiento que ha aplicado o las explicaciones que nos da somos capaces de ver distintos niveles de desarrollo en las competencias subyacentes.

Hay que destacar también que el test, en su formato original, viene avalado por un gran número de investigaciones internacionales que muestran que se trata de un instrumento con evidencias de validez y fiabilidad. Sin embargo, existen una serie de áreas de mejoras que los autores deberían considerar en futuras revisiones del test, se comentan algunas:

Primero, al tratarse de una adaptación sería recomendable incluir un apartado explicando el procedimiento llevado a cabo para su adaptación y traducción al contexto español. Por ejemplo, si procede, describir el juicio de expertos, cambios con respecto a la versión original, etc. Asimismo, sería recomendable incluir más información sobre la validez de contenido de la prueba tanto en su versión original como en la adaptación al contexto español.

Segundo, sería interesante obtener otros niveles de fiabilidad con otros estimadores como Omega o utilizando paradigmas psicométricos como la TRI. Por ejemplo, la Función de Información se podría presentar como medida complementaria al coeficiente Alfa de Conbrach permitiendo al usuario identificar el nivel de error de medida en función del nivel de competencia. Si el objetivo es la detección de casos con dificultades de aprendizaje sería deseable que el error típico de medida fuera más reducido para niveles medios y bajos de competencia, lo que permitiría una identificación mucho más precisa de esos casos.

Tercero, sería interesante realizar estudios de DIF y sesgo en función de alguna variable interesante. Por otro lado, se tienen que mejorar los estudios factoriales para comprender mejor la estructura interna de las puntuaciones del cuestionario. Para ello se podrían replicar los factores encontrados, utilizar análisis factoriales confirmatorios e invarianza de medición. Por último, sería interesante acumular más evidencias de validez del instrumento para la detección de casos con dificultades de aprendizaje. Por ejemplo, comprobar la sensibilidad y especificidad del instrumento (curvas ROC) para detectar alumnado con dificultades de aprendizaje. Utilizar diferentes muestras y aumentar la representatividad de las mismas. En definitiva, ampliar la utilidad del instrumento a diversas áreas como la educativa y la clínica.

En suma, los distintos componentes del test nos permiten observar y comparar el rendimiento de cada alumno/a en diferentes competencias matemáticas relacionadas pero diferentes, más asociadas a aspectos relacionales o a aspectos numéricos, por lo que podemos observar los puntos fuertes y débiles de cada una de las personas a las que se aplique el test. Además, el test permite comparar los resultados de cada alumno/a con un grupo normativo de referencia obtenido a través de la aplicación de la prueba a la amplia muestra de sujetos ya comentada.

ANÁLISIS DETALLADO DE LA PRUEBA

1. DESCRIPCIÓN GENERAL DEL TEST

1.1. Nombre del test:

- Test de evaluación matemática temprana.

1.2. Nombre del test en su versión original:

- The utrecht early mathematical competence test.

1.3. Autor del test original:

- J.E.H. van Luit, B.A.M. van de Rijt & A.H.Pennings.

1.4. Autor de la adaptación española:

- J.E.H. van Luit, B.A.M. van de Rijt, José. I. Navarro, Manuel Aguilar, Concepción Alcalde, Esperanza Marchena, Gonzalo Ruiz, Inmaculada Menacho y Manuel G. Sedeño.

1.5. Editor del test en su versión original:

- Graviant Doetinchem.

1.6. Editor de la adaptación española:

- Editorial EOS.

1.7. Fecha de publicación del test original:

- 1998.

1.8. Fecha de la publicación del test en su adaptación española:

- 2011.

1.9. Fecha de la última revisión del test en su adaptación española:

- 2011.

1.10. Área general de la(s) variable(s) que pretende medir el test:

- Aptitudes, rendimiento académico.

1.11. Breve descripción de la(s) variable(s) que pretende medir el test:

- El test evalúa el nivel de competencia matemática temprana en niños de 4 a 7 años, requisito imprescindible para poder seguir en los años siguientes una educación matemática formal.
- El test dispone de tres versiones paralelas, de 40 ítems cada una, que se aplican aproximadamente en unos 30 minutos. Los ítems se agrupan en 8 tareas diferentes, por lo que cada componente es evaluado con 5 ítems. Permite analizar el conocimiento numérico temprano en sus ocho componentes: Conceptos de comparación; Clasificación; Correspondencia uno a uno; Seriación; Conteo (verbal, estructurado y resultante) y Conocimiento general de los números. Los cuatro primeros subtests evalúan habilidades de tipo piagetiano, mientras que los cuatro últimos estiman las habilidades numéricas de naturaleza más cognitiva.

- Todos los ítems son presentados oralmente y los niños responden señalando en un material con dibujos o, en el caso de las tareas de contar y de numeración, manipulando pequeños cubos. En otros tres ítems el alumno debe usar el lápiz para unir los objetos del dibujo presentado.
- Se aplica de forma individual y, comparando el resultado de cada niño con un grupo normativo, nos ofrece un nivel de competencia matemática temprana. Además, proporciona una información (cuantitativa y cualitativa) que puede ser usada para realizar una detección precoz del alumnado con dificultades de aprendizaje relacionadas con los números.

1.12. Áreas de aplicación:

- Psicología educativa.

1.13. Formato de los ítems:

- Elección múltiple, respuestas manipulativas y respuestas verbales abiertas.

1.14. Número de ítems:

La escala consta de un total de 40 ítems tanto para la forma A como para la B y se distribuyen de la siguiente forma:

Subtest relacionales:

- Comparación: 5 ítems
- Clasificación: 5 ítems
- Correspondencia: 5 ítems
- Seriación: 5 ítems

Subtests Numéricos:

- Conteo verbal: 5 ítems
- Conteo estructurado: 5 ítems
- Conteo resultante: 5 ítems
- Conocimiento general de los números: 5 ítems

1.15. Soporte:

- Administración oral, papel y lápiz y manipulativo.

1.16. Cualificación requerida para el uso del test de acuerdo con la documentación aportada:

- Nivel A.

1.17. Descripción de las poblaciones a las que el test es aplicable:

- El test se ha desarrollado para niños/as de 4-7 años, es decir para ser aplicado en alumnado de 2º y 3º de educación infantil, y 1º y 2º de educación primaria. Aunque no se especifica, el test puede ser aplicado a cualquier tipo de población, siempre y cuando exista una suficiente comprensión de las instrucciones proporcionadas a nivel verbal por el aplicador y, por lo tanto, un nivel elevado de competencia lingüística en castellano.

- No obstante, existen versiones del mismo test en inglés, alemán y finés. La aplicación de la prueba para obtener datos normativos se ha realizado en colegios españoles, con alumnado de clase media / media-baja, aunque no se especifica las nacionalidades ni orígenes étnicos.

1.18. Existencia o no de diferentes formas del test y sus características:

- Existen tres versiones del test (A, B y C), aunque todas miden lo mismo y no implican niveles de dificultad diferente. La versión A y la versión B incluyen ítems totalmente diferentes, aunque se mantiene constante la estructura en ocho subescalas y los aspectos de la competencia matemática que miden cada una de ellas. La versión C incluye una mezcla de las otras dos versiones previas.
- En el cuadernillo se explica que las distintas versiones están disponibles por si el evaluador quiere comprobar el resultado no esperado de un alumno en el test, repitiendo la prueba unos días más tarde para descartar que los resultados estén sesgados por alguna condición de la administración del test.

1.19. Procedimiento de corrección:

- Manual.

1.20. Puntuaciones:

- Se generan tres puntuaciones directas a partir del sumatorio del número de ítems correctamente resueltos. Por tanto una puntuación para los ítems de los subtests relacionales (máximo 20 puntos), otra puntuación para los subtests numéricos (máximo 20 puntos) y la puntuación sumatoria total a partir de las dos anteriores (máximo 40 puntos). No existe ninguna estrategia de inversión o corrección de ítems.
- Con la puntuación total del test se obtiene una escala de percentiles a la que se denomina Nivel de Competencia Matemática y que se puede interpretar en cinco grupos distintos de competencia de acuerdo con los distintos grupos de edad.
- Se facilita la tabla de equivalencia de la puntuación directa a centil (NCM) y los criterios de clasificación del nivel (desde A hasta E) para cada grupo de edad.

1.21. Escalas utilizadas:

- Centiles.

1.22. Posibilidad de obtener informes automatizados:

- No.

1.23. Tiempo estimado para la aplicación del test:

- En aplicación individual: 20-30 minutos.

1.24. Documentación aportada por el editor:

- Manual, láminas, cubos, 10 hojas de registro y 10 hojas de datos.

1.25. Precio de un juego completo de la prueba:

- 50 euros.

1.26. Precio y número de ejemplares del paquete de cuadernillos:

- 2 euros por cada hoja de registro.

1.27. Precio y número de ejemplares del paquete de hojas de respuesta:

- 2 Euros por cada hoja de datos.

1.28. Precio de la corrección y/o elaboración de informes por parte del editor:

- No consta.

2. VALORACIÓN DE LAS CARACTERÍSTICAS DEL TEST

2.1. Aspectos generales

Contenido	Valoración	Puntuación
2.1. Calidad de los materiales del test	Excelente	5
2.2. Calidad de la documentación aportada	Excelente	5
2.3. Fundamentación teórica	Excelente	5
2.4. Adaptación del test	Excelente	5
2.5. Desarrollo de los ítems del test	Buena	4
2.6. Calidad de las instrucciones para el participante	Excelente	5
2.7. Calidad de las instrucciones (administración, puntuación, interpretación)	Excelente	5
2.8. Facilidad para registrar las respuestas	Excelente	5
2.9. Bibliografía del manual	Excelente	5
2.10. Datos sobre el análisis de los ítems	Excelente	5

2.11. Validez

2.11.1. Evidencias de validez de contenido

Contenido	Valoración	Puntuación
2.11.1.1. Calidad de la representación del contenido o dominio	Buena	4
2.11.1.2. Consultas a expertos	De forma informal un pequeño nº de expertos	2

2.11.2. Evidencias de validez basadas en la relación entre las puntuaciones del test y otras variables

2.11.2.1. Evidencias de validez basadas en la relación con otras variables

Contenido	Valoración	Puntuación
2.11.2.1. Diseños empleados	Correlaciones con otros tests	
2.11.2.1.1. Tamaño de las muestras	Un estudio con muestra pequeña	1
2.11.2.1.2. Procedimiento de selección de las muestras	Incidental	
2.11.2.1.3. Calidad de los tests empleados como criterio o marcador	Adecuada	3
2.11.2.1.4. Promedio de las correlaciones con otros tests que miden constructos similares	Buena	4
2.11.2.1.5. Promedio de las correlaciones con otros tests que miden constructos no relacionados	-	-
2.11.2.1.6. Resultados de la matriz multirasgo-multimétodo	-	-
2.11.2.1.7. Resultados de las diferencias intergrupo	-	-

2.11.2.2. Evidencias de validez basadas en la relación entre las puntuaciones del test y un criterio

Contenido	Valoración	Puntuación
2.11.2.2.1. Criterios empleados	Se han utilizado diferentes criterios	
2.11.2.2.2. Calidad de los criterios empleados	Buena	4
2.11.2.2.3. Relación temporal entre test y criterio	Concurrente y predictivo	
2.11.2.2.4. Tamaño de las muestras	Un estudio con muestra pequeña	1
2.11.2.2.5. Procedimiento de selección de las muestras	Incidental	
2.11.2.2.6. Promedio de las correlaciones del test con los criterios	Excelente	5

2.11.3. Evidencias de validez basadas en la estructura interna

Contenido	Valoración	Puntuación
2.11.3.1. Resultados del análisis factorial	Adecuada	3
2.11.3.2. Funcionamiento diferencial de los ítems	-	-

2.11.4. Acomodaciones en la administración del test

Contenido	Valoración	Puntuación
2.11.4. El manual del test informa sobre las acomodaciones en la administración del test	No	

Comentarios generales sobre evidencias de validez

El manual describe sucintamente las evidencias de validez basadas en la estructura interna del test y basadas en la relación con otras variables. En cuanto a la estructura interna del test se realiza un análisis factorial exploratorio pero no se aportan índices de ajuste. A pesar de tratarse de una adaptación no se realiza un análisis factorial confirmatorio ni se compara la estructura factorial encontrada con la original.

Concretamente, el análisis de validez de constructo nos permite ver que el factor resultante del análisis factorial explica el 69% de la varianza encontrada. La peculiaridad de las distintas capacidades incluidas en la prueba asume la variabilidad restante. Por otro lado, en lo que se refiere a la validez divergente, se observan correlaciones medias-altas y significativas entre las diferentes subescalas del test, lo que indica que los resultados en algunas pruebas pueden predecirse a partir de los resultados de otras. Respecto a la validez concurrente, se observa que todas las comparaciones de interés entre la prueba evaluada y otra similar resultan significativas, especialmente las dos puntuaciones totales ($r=0,689$; $p < 0,01$).

Finalmente, la validez predictiva del TEMT es también adecuada, pues todos los subtest presentan coeficientes de correlación estadísticamente significativos con el rendimiento en matemáticas 2 cursos después. Las puntuaciones obtenidas en el test (especialmente, las dos obtenidas en los subtest relacionales por un lado y numéricos por otro) predicen el rendimiento matemático del alumnado dos años después ($r=0,689^{**}$ y $r=0,633^{**}$, respectivamente).

Las evidencias de validez en relación con otras variables resultan adecuadas, se muestra tanto la correlación con otros test que miden lo mismo como con un criterio. Si bien

los resultados presentados ser podrían considerar aceptables, sería altamente recomendable al menos presentar evidencias de validez de contenido, describir el proceso de traducción y adaptación, y estudiar el funcionamiento diferencial de los ítems en función de alguna variable como sexo a través de regresión logística u otro procedimiento.

2.12. Fiabilidad

Contenido	Valoración	Puntuación
2.12.1. Datos aportados sobre fiabilidad	Un coeficiente de fiabilidad y un error de medida por escala o subescala	

2.12.2. Equivalencia formas paralelas

Contenido	Valoración	Puntuación
2.12.2.1. Tamaño de las muestras	Un estudio con una muestra pequeña	1
2.12.2.2. Puesta a prueba de los supuestos de paralelismo	Buenos	4
2.12.2.3. Promedio de coeficientes de equivalencia	Adecuada	3

2.12.3. Consistencia interna

Contenido	Valoración	Puntuación
2.12.3.1. Tamaño de las muestras	Un estudio con muestra grande	3
2.12.3.2. Coeficientes de consistencia interna presentados	Coeficiente alfa o KR-20	
2.12.3.3. Promedio de los coeficientes de consistencia	Excelente	5

2.12.4. Estabilidad (test-retest)

Contenido	Valoración	Puntuación
2.12.4.1. Tamaño de las muestras	Un estudio con una muestra moderada	2
2.12.4.2. Coeficientes de estabilidad	Adecuada	2

2.12.5. Cuantificación de la precisión mediante TRI

Contenido	Valoración	Puntuación
2.12.5.1. Tamaño de las muestras	-	-
2.12.5.2. Coeficientes proporcionados	-	-
2.12.5.3. Tamaño de los coeficientes	-	-

2.12.6. Fiabilidad inter-jueces

Contenido	Valoración	Puntuación
2.12.6.1. Tipos de coeficientes presentados	-	-
2.12.6.2. Promedio de los coeficientes	-	-

Comentarios generales sobre fiabilidad

El test presenta una fiabilidad de las puntuaciones adecuada, aportando datos sobre la consistencia interna del instrumento, las correlaciones entre los diferentes subtest y la fiabilidad test-retest. Por lo tanto, el estudio de la fiabilidad se centra fundamentalmente en la consistencia interna y la estabilidad temporal desde una perspectiva clásica

Concretamente, los resultados obtenidos al aplicar el estadístico alfa de Cronbach superar siempre el valor del coeficiente de fiabilidad de 0,8, siendo el valor en el caso de los Subtest relacionales de 0,94. El dato del coeficiente global que arroja un alfa de Cronbach = 0,95. Así mismo, se calculó la correlación entre los diferentes componentes del test, observándose valores correlacionales significativos y elevados en todos los casos.

Finalmente, se calculó la fiabilidad test-retest de la prueba empleando las dos versiones de la misma, pues en este test la maduración y el aprendizaje pueden afectar a los resultados. Con esta metodología, se observan correlaciones significativas y medias entre las puntuaciones obtenidas en las formas A y B del instrumento, lo que corrobora una fiabilidad test-retest adecuada.

2.13. Baremación e interpretación de las puntuaciones

2.13.1. Interpretación normativa de las puntuaciones

Contenido	Valoración	Puntuación
2.13.1.1. Calidad de las normas	Amplio rango de baremos	5
2.13.1.2. Tamaño de las muestras	Muy grande	5
2.13.1.3. Aplicación de estrategia de tipificación continua	Sí	
2.12.1.4. Procedimiento de selección de las muestras	Incidental	
2.12.1.5. Actualización de baremos	Excelente	5

2.13.2. Interpretación referida a criterio

Contenido	Valoración	Puntuación
2.13.2.1. Adecuación del establecimiento de los puntos de corte	-	-
2.13.2.2. Procedimiento empleado para fijar los puntos de corte		-
2.13.2.3. Procedimiento de obtención del acuerdo inter-jueces	-	
2.13.2.4. Valor del coeficiente de acuerdo inter-jueces	-	-

Comentarios generales sobre baremación e interpretación de las puntuaciones

Los baremos son muy adecuados considerando el muestreo de tipo incidental.