

Evaluación de

Escala de depresión geriátrica

GDS

RESUMEN DE LA VALORACIÓN DEL TEST

Descripción general

Característica	Descripción
Nombre del test	Escala de depresión geriátrica de Yevesage abreviada (Versión española): GDS-VE (o GDS-15)
Autor	J. A. Yesavage y J.I. Sheikh
Autor de la adaptación española	J. Martínez de la Iglesia, M. Onís Vilches, R. Dueñas Herrero, C. Albert Colomer, C. Aguado Taberné, y R. Luque Luque
Variable(s)	Depresión
Áreas de aplicación	Clínica y salud Gerontopsicología
Soporte	Lápiz y Papel. Oral. Corrección manual.

Valoración general

Característica	Valoración	Puntuación
Materiales y documentación	No evaluado	-
Fundamentación teórica	Buena-Excelente	4.5
Adaptación	Buena-Excelente	4.5
Análisis de ítems	Adecuado	3
Validez: contenido	Inadecuado-Adecuado con carencias	1.5
Validez: relación con otras variables	Buena-Excelente	4.3
Validez: estructura interna	Adecuada, pero con algunas carencias	2
Validez: análisis del DIF	Buena	4
Fiabilidad: equivalencia	-	-
Fiabilidad: consistencia interna	Buena-Excelente	4.3
Fiabilidad: estabilidad	Adecuado con carencias-Adecuada	2.5
Fiabilidad: TRI	Buena-Excelente	4.5
Fiabilidad inter-jueces	Adecuada	3
Baremos e interpretación de puntuaciones	Buena	4

Comentarios generales

El GDS, y en concreto la versión abreviada que nos ocupa (GDS-15), es la prueba recomendada para el cribado de la depresión en personas mayores por sociedades como la Sociedad Española de Medicina Familiar y la Sociedad Española de Medicina interna. La versión corta de la prueba se construyó con la finalidad de reducir la fatiga y mantener niveles de atención correctos durante su aplicación. Algunas de las ventajas de la prueba a nivel práctico son: a) que se administra fácilmente (puede ser administrada por un médico u otra persona o autoadministrada), b) que las instrucciones de respuesta son muy sencillas (se debe responder sí o no) y c) que es breve y apenas requiere unos 5 minutos.

Teniendo en cuenta la finalidad de la prueba, los estudios apoyan la recomendación de las sociedades anteriores en términos generales. En primer lugar, los estudios existentes apoyan la fiabilidad de las puntuaciones, cuando esta es evaluada como consistencia interna o fiabilidad inter-jueces. También existe evidencia de fiabilidad como estabilidad temporal, aunque en este último caso los estudios revisados no incluyen muestras locales, por lo que la consistencia temporal debe reconfirmarse. Lo mismo puede decirse de los estudios basados en TRI. Las funciones de información del test son elevadas, especialmente para detectar depresión moderada, pero los estudios se han realizado con muestras no españolas y es conveniente realizar estudios de TRI que permitan evaluar la fiabilidad de la prueba para personas con distinta severidad de depresión y distintos puntos de corte. Esto es fundamental para realizar una interpretación criterial que maximice los diagnósticos correctos.

En cuanto a las evidencias de validez, teniendo en cuenta la finalidad de una prueba de cribado, los estudios apoyan el uso del GDS-15, ya que presenta niveles satisfactorios de sensibilidad y especificidad para discriminar entre quienes probablemente tienen un problema de depresión y quienes no (recomendándose puntos de corte de 5 o más de 5). Esta es una evidencia fundamental de validez. También se obtiene apoyo a la validez de las inferencias basada en relaciones con otras variables (convergente y discriminante), con correlaciones altas y positivas con otras escalas de depresión o medidas de bienestar, y bajas con constructos menos relacionados. Existen evidencias que apoyan su uso incluso cuando las personas presentan cierto grado deterioro cognitivo o demencia, siempre que no sea severa. También existen evidencias basadas en el Funcionamiento Diferencial de los Ítems (en función de sexo y edad, o el grado de deterioro cognitivo -excluyéndose personas con demencia severa o estados avanzados de Alzheimer). Sin embargo, los estudios de DIF analizados no incluyen muestras locales, que deberían considerarse en futuros estudios.

Aunque no hay estudios tradicionales de validez de contenido, teniendo en cuenta el proceso de creación, la finalidad de la prueba y los resultados empíricos que avalan su uso, no se considera un aspecto central. Puede decirse lo mismo respecto de la estructura factorial que es muy inestable a través de los distintos estudios analizados. Aunque la estructura factorial no sea un aspecto crítico, sí es conveniente realizar investigaciones que ayuden a entender de qué depende la estructura factorial (ya hay algunos estudios que muestran que en pacientes psiquiátricos la estructura factorial es diferente de la población general o con cierto deterioro cognitivo) y si las distintas dimensiones que puedan ser estables para distintos grupos de distintas características contribuyen a hacer predicciones diferentes de variables relevantes para la calidad de vida y el bienestar de las personas mayores. También conviene centrarse en un análisis de ítems más pormenorizado, ya que los pocos estudios que realizan análisis de ítems desde la aproximación más clásica no son concluyentes sobre la adecuación de todos los ítems.

Finalmente, aunque la prueba tiene una finalidad con una clara interpretación criterial, también existe un baremo normativo elaborado a partir de una muestra local de gran tamaño, que puede complementar la información. Esta muestra, sin embargo, está constituida por personas que no presentan deterioro cognitivo ni problemas de salud que pudieran aumentar la probabilidad de sufrir una depresión. Esta instancia debe tenerse en cuenta al usar los baremos para interpretar las puntuaciones de la prueba.

En la valoración global de la escala sobresale su carácter internacional. Se trata de una escala utilizada en varios países, lo que permite obtener una imagen general de sus características psicométricas, si bien son necesarios más estudios locales sobre sus propiedades psicométricas (estructura interna, análisis de TRI). La mayor deficiencia de la escala se relaciona con su estructura interna, pero con relación al uso propuesto, se puede concluir que en general, los trabajos publicados sobre GDS-VE avalan la utilidad de la escala en la práctica clínica.

ANÁLISIS DETALLADO DE LA PRUEBA

1. DESCRIPCIÓN GENERAL DEL TEST

1.1. Nombre del test:

GDS. Escala de depresión geriátrica de Yesavage abreviada (Versión española): GDS-VE (o GDS-15)

1.2. Nombre del test en su versión original:

Geriatric Depression Scale (GDS) - short version

1.3. Autor/es del test original:

J. A. Yesavage y J.I. Sheikh

1.4. Autor/es de la adaptación española:

J. Martínez de la Iglesia, M. Onís Vilches, R. Dueñas Herrero, C. Albert Colomer, C. Aguado Taberné, y R. Luque Luque

1.5. Editor del test en su versión original:

No procede, puesto que no es un test comercial. El cuestionario reducido se publicó en Yesavage, J. A., & Sheikh, J. I. (1986). 9/Geriatric depression scale (GDS) recent evidence and development of a shorter version. *Clinical gerontologist*, 5(1-2), 165-173.

NOTA: La versión original de 30 ítems a partir de la cuál se creó la versión reducida que nos ocupa se publicó en: Brink, T. L., Yesavage, J. A., Lum, O., Heersema, P. H., Adey, M., & Rose, T. L. (1982). Screening tests for geriatric depression. *Clinical gerontologist*, 1(1), 37-43.

1.6. Editor de la adaptación española:

Se trata de un test no comercializado. Se publicó en: Martínez de la Iglesia, J., Onís Vilches, M., Dueñas Herrero, R., Albert Colomer, C., Aguado Taberné, C., & Luque Luque, R. (2002). Versión española del cuestionario de Yesavage abreviado (GDS) para el despistaje de depresión en mayores de 65 años: adaptación y validación. *Medifam*, 12(10), 26-40.

1.7. Fecha de publicación del test original:

1986 (la versión reducida sometida a revisión)

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

2002 (la versión reducida sometida a revisión).

1.9. Fecha de la última revisión del test:

2021 (último estudio, que propone datos normativos): Delgado Losada, M. L., López Higes, R., Rubio Valdehita, S., Facal Mayo, D., Lojo Seoane, C., Montenegro Peña, M., ... & Fernández Blázquez, M. Á. (2021). Spanish consortium for ageing normative data (scand): screening tests (mmse, gds-15 and mfe). *Psicothema*, 33, 70-76.

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

Escalas clínicas. Depresión gerontología.

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

GDS-VE es una escala breve de screening o de cribado rápido para la evaluación de cuadros depresivos en personas mayores de 65 años. El uso de escalas adaptadas a esta población es necesario, dadas las particularidades propias del envejecimiento. Por un lado, algunos factores pueden ser más frecuentes que en población más joven y les hacen más vulnerables para presentar sintomatología depresiva (p. ej., síntomas cognitivos como la pérdida de memoria, presencia de problemas médicos, factores psicosociales como la pérdida de seres queridos). Por otro lado, algunos síntomas que resultan importantes para el diagnóstico de depresión son algo natural en el proceso de envejecimiento (p. ej., síntomas somáticos), lo que hace que el uso de escalas no adaptadas dé lugar a muchos falsos positivos.

La GDS-VE fue desarrollada a partir de la versión original de la escala GDS que está compuesta por 30 ítems. La versión analizada aquí se compone de 15 ítems (10 directos y 5 inversos), que se responden en un formato dicotómico SI/NO. El tiempo de aplicación es de 5-7 minutos. Aunque inicialmente se propone como una prueba de autoinforme de lápiz y papel, es frecuente que se administre por personal sanitario de forma oral, especialmente en personas mayores que presentan cierto grado de deterioro cognitivo.

Los ítems representan características de la depresión relacionadas con la ausencia de sentimientos positivos (p. ej., sentirse satisfecho, feliz, de buen humor, con energía), la presencia de sentimientos negativos (sentirse aburrido, vacío, desprotegido, desesperanzado, temeroso, inútil), síntomas conductuales (p. ej., inactividad). El GDS no contiene preocupaciones somáticas que pueden ser más comunes en los ancianos (p. ej., alteraciones del sueño).

Se recomienda el uso de la escala como método de cribado rápido y ayuda a la clasificación de pacientes susceptibles de ser evaluados con criterios más fiables. Se ofrece una puntuación única en una escala 0-15, especificando un punto de corte de 5 o más para la clasificación.

Nótese que para la GDS hay una versión larga (30 ítems; Yesavage et al., 1983) y una versión corta (15 ítems; Yesavage y Sheikh, 1986). Por otro lado, existen distintas adaptaciones al español, que varían en número de ítems y traducción de los enunciados, por lo que ha sido necesario seleccionar una de ellas para la presente revisión. Para localizar las distintas versiones se partió del trabajo de Cabañero-Martínez et al. (2007), que identifican como adaptaciones de mayor calidad la de 30 ítems, de Izal y Montorio (1993), y la de 15 ítems, de Martínez de la Iglesia et al. (2002). Para la presente evaluación se ha seleccionado la adaptación de 15 ítems de Martínez de la Iglesia et al. (2002), puesto que: (a) en poblaciones mayores es especialmente deseable la reducción del tiempo; (b) esta adaptación es una de las más utilizadas (citadas) de acuerdo con las bases de datos bibliográficas (p. ej., Google Scholar, Web of Science y/o Scopus); (c) es una de las versiones en español que aparece en la página web del autor original de la escala (<https://web.stanford.edu/~yesavage/GDS.html>); (d) se dispone de algunos datos de baremación, por aplicación en muestras amplias.

Referencias:

Cabañero-Martínez, M. J., Richart-Martínez, M., Muñoz-Mendoza, C. L., & Reig-Ferrer, A. (2007). Revisión estructurada de las escalas de depresión en personas mayores. *International Journal of Clinical and Health Psychology*, 7(3), 823-846.

Izal, M. y Montorio, I. (1993). Adaptation of the Geriatric Depression Scale: A preliminary study. *Clinical Gerontologist*, 13, 83-91.

Martínez de la Iglesia, J., Onís, M.C., Dueñas, H.R., Albert, C.C., Aguado, T.C. y Luque, L.R. (2002). Versión española del cuestionario de Yesavage abreviado (GDS) para el despistaje de depresión en mayores de 65 años: adaptación y validación. *Revista de Medicina Familiar y Comunitaria*, 12, 620-630.

Yesavage, J. A., Brink, T.L., Rose, T.L., Lum, O., Huang, V., Adey, M., & Leirer, V.O. (1983). Development and validation of a geriatric depression scale: a preliminary report. *Journal of Psychiatric Research*, 17 (1): 37-49.

Yesavage, J. A., y Sheikh, J. I. (1986). Geriatric Depression Scale (GDS): Recent evidence and development of a shorter version. *Clinical Gerontologist*, 5, 165-173.

1.12. Áreas de aplicación:

Psicología clínica, Salud general y bienestar. Gerontopsicología.

1.13. Formato de los ítems:

Respuesta dicotómica (sí/no)

1.14. Número de ítems:

15

1.15. Soporte:

Administración oral.

Papel y lápiz.

Smarr y Keefer (2011) sugieren que si el cuestionario se administra repetidamente debe aplicarse la prueba con el mismo procedimiento de administración, ya que el modelo de administración puede producir resultados variables.

Referencias:

Smarr, K. L., & Keefer, A. L. (2011). Measures of depression and depressive symptoms: Beck depression Inventory-II (BDI-II), center for epidemiologic studies depression scale (CES-D), geriatric depression scale (GDS), hospital anxiety and depression scale (HADS), and patient health Questionnaire-9 (PHQ-9). *Arthritis care & research*, 63(S11), S454-S466.

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

Nivel C.

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

La prueba se dirige a personas mayores, independientemente de su nivel educativo. La propuesta inicial en la versión española abreviada de 15 ítems (Martínez de la Iglesia et al., 2002), así en como la original de 30 ítems propuesta por Brink et al. (1982) se dirige a población mayor de 65 años, aunque el rango de edad difiere en los distintos estudios. Por ejemplo, Delgado-Losada et al. (2021), construye baremos de la versión española abreviada a partir de una muestra de personas de 50 años o más, y en la propuesta original abreviada de 15 ítems previa a la adaptación española, así como el estudio de Lucas-Carrasco (2012) se emplean muestras de personas mayores de 55 años.

Según los autores de la prueba original, el GDS es aplicable tanto a población general de personas mayores, como a poblaciones que tengan algún tipo de problema físico o cierto grado de deterioro cognitivo que no sea muy elevado, existiendo algunos estudios que muestran su aplicabilidad en ancianos con demencia leve o moderada (Lucas-Carrasco, 2012) según el MMSE (Mini-Mental State Examination). Aunque otros estudios recomiendan no usarlo cuando hay problemas cognitivos moderados. No se aconseja su uso con personas que presentan deterioro cognitivo severo.

En los baremos españoles se excluyen pacientes con alteración psiquiátrica grave no depresiva o con problemas sensoriales o motores severos.

Referencias:

Brink, T. L., Yesavage, J. A., Lum, O., Heersema, P. H., Adey, M., & Rose, T. L. (1982). Screening Tests for Geriatric Depression. *Clinical Gerontologist*, 1(1), 37-43. https://doi.org/10.1300/J018v01n01_06

Delgado-Losada, M. L., López-Higes, R., Rubio-Valdehita, S., Facal, D., Lojo-Seoane, C., Montenegro-Peña, M., Frades-Payo, B., & Fernández-Blázquez, M. Á. (2021). Spanish consortium for ageing normative data (SCAND): Screening tests (MMSE, GDS-15 and MFE). *Psicothema*, 33(1), 70-76.

Lucas-Carrasco, R. (2012). Spanish version of the geriatric depression scale: Reliability and validity in persons with mild-moderate dementia. *International Psychogeriatrics*, 24(8), 1284-1290.

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

Además de la versión abreviada que se analiza, compuesta por 15 ítems, la versión original del GDS (Brink et al, 1982) adaptada al contexto español por Izal y Montorio (1993), cuenta con 30 ítems. También, a partir de la versión de 15 ítems, se han propuesto versiones aún más reducidas de cinco y cuatro ítems y de un ítem. La comparación de las distintas versiones puede consultarse en Martínez de la Iglesia et al. (2005). Finalmente, a partir de la versión de 30 ítems también se han propuesto otras

versiones breves, como la de Izal et al. (2010) de 10 ítems que busca optimizar la predicción de la depresión mayor y la versión de ocho ítems de Buz (1996) para ancianos institucionalizados.

Se han propuesto distintas traducciones/adaptaciones de la escala (ver revisión y comparación de las distintas propuestas en Cabañero et al., 2007). Entre ellas, la propuesta de Martínez de la Iglesia et al. (2002) objeto de revisión en este informe es la que sigue un proceso de traducción y adaptación más cuidado.

Para la versión original en inglés de 15 ítems existen versiones informatizadas en web (<http://www.medafile.com/GDS15.htm>) y apps móviles (para iOS y ANDROID). También existe una app denominada GDS que permite presentar el test en varias lenguas, incluido el español. Sin embargo, se trata de una versión en español-mexicano, en la que los ítems son algo diferentes a la versión que está siendo evaluada en este informe.

Referencias:

Brink, T. L., Yesavage, J. A., Lum, O., Heersema, P. H., Adey, M., & Rose, T. L. (1982). Screening Tests for Geriatric Depression. *Clinical Gerontologist*, 1(1), 37-43. https://doi.org/10.1300/J018v01n01_06

Buz, J. (1996). Mini-GDS 8: Una nueva versión breve para ancianos institucionalizados. *Geríatrika*, 12, 41-45.

Cabañero-Martínez, M. J., Richart-Martínez, M., Muñoz-Mendoza, C. L., & Reig-Ferrer, A. (2007). Revisión estructurada de las escalas de depresión en personas mayores. *International Journal of Clinical and Health Psychology*, 7(3), 823-846.

Izal, M. y Montorio, I. (1993). Adaptation of the Geriatric Depression Scale: A preliminary study. *Clinical Gerontologist*, 13, 83-91

Izal, M., Montorio, I., Nuevo, R., Pérez-Rojo, G., & Cabrera, I. (2010). Optimising the diagnostic performance of the Geriatric Depression Scale. *Psychiatry Research*, 178(1), 142-146.

Martínez de la Iglesia, J., Onís, M.C., Dueñas, H.R., Albert, C.C., Aguado, T.C., Colomer, A., Arias, C., Blanco, MC. (2005). Abbreviating the brief. Approach to ultra-short versions of the Yesavage questionnaire for the diagnosis of depression. *Atención Primaria*. 2005;35(1):14-21.

1.19. Procedimiento de corrección:

Manual.

1.20. Puntuaciones:

En la escala se obtiene una puntuación directa. Los ítems son puntuados de forma dicotómica (1/0). Para 10 de los ítems la respuesta "sí" es indicativa de depresión (por ejemplo "¿Siente que su vida está vacía?") y se puntúa con un 1. Los cinco ítems restantes están formulados en sentido inverso, de manera que la respuesta "no" es indicativa de depresión (por ejemplo, "¿En general, está satisfecho/a con su vida?") y es ésta la que se puntúa como un 1. La puntuación directa en la escala se obtiene sumando las puntuaciones asignadas a los ítems (oscilando de 0 a 15).

En la página web (<https://web.stanford.edu/~yesavage/GDS.html>) el autor recomienda sustituir las omisiones por la media de la persona en los ítems contestados y, si se obtiene una puntuación decimal, redondear esta hacia arriba.

1.21. Escalas utilizadas:

Se interpretan principalmente las puntuaciones directas. Puntuaciones de 5 a 8, indicaría la presencia de síntomas de depresión leve, de 9 a 11, de depresión moderada y de 12 a 15, de depresión severa.

En el estudio de Delgado-Losada et al. (2021) se indican puntuaciones percentiles asociadas a las puntuaciones directas. También en este estudio se presentan unas Puntuaciones Escalares (Scalar Scores), que son puntuaciones típicas normalizadas y transformadas para tener media de 10 y desviación típica de 3.

Referencias:

Delgado Losada, M. L., López Higes, R., Rubio Valdehita, S., Facal Mayo, D., Lojo Seoane, C., Montenegro Peña, M., ... & Fernández Blázquez, M. Á. (2021). Spanish consortium for ageing normative

data (scand): screening tests (mmse, gds-15 and mfe). Psicothema, 33, 70-76.

1.22. Posibilidad de obtener informes automatizados:

No

1.23. Tiempo estimado para la aplicación del test:

En aplicación individual: 5-7 minutos (según información de Martínez de la Iglesia et al., 2002; Yesavage y Sheikh, 1986) y las instrucciones del cuestionario según el Instituto Nacional de Geriátrica, de Mexico). Las instrucciones indican que las preguntas deben ser leídas por los profesionales de la salud a las personas a evaluar.

En aplicación autoadministrada, el tiempo será menor. Smarr y Keefer (2011) informan de un tiempo de aplicación de entre 2 y 5 minutos.

En aplicación colectiva: No procede

Referencias:

Martínez de la Iglesia, J., Onís, M.C., Dueñas, H.R., Albert, C.C., Aguado, T.C. y Luque, L.R. (2002). Versión española del cuestionario de Yesavage abreviado (GDS) para el despistaje de depresión en mayores de 65 años: adaptación y validación. Revista de Medicina Familiar y Comunitaria, 12, 620-630.

Smarr, K. L., & Keefer, A. L. (2011). Measures of depression and depressive symptoms: BDI-II, CES-D, GDS, HADS, and PHQ-9. Arthritis Care and Research, 63(S11), 454-466.

Yesavage, J. A. & Sheikh, J. I., (1986). Geriatric Depression Scale (GDS): Recent evidence and development of a shorter version. Clinical Gerontologist, 5, 165-173.

1.24. Documentación aportada por el editor:

No se trata de un test comercializado. En este caso, la documentación ha sido aportada por el coordinador de la evaluación de test (i.e., artículos seleccionados sobre la escala original y la adaptada).

1.25. Precio de un juego completo de la prueba:

No procede. No se trata de un test comercializado.

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

No procede. No se trata de un test comercializado.

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

No procede. No se trata de un test comercializado.

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

No procede. No se trata de un test comercializado.

2. VALORACIÓN DE LAS CARACTERÍSTICAS DEL TEST

2.1. Aspectos generales:

Contenido	Valoración	Puntuación
2.1. Calidad de los materiales del test	No evaluado	--
2.2. Calidad de la documentación aportada	No evaluado	--
2.3. Fundamentación teórica	Buena-Excelente	4.5
2.4. Adaptación del test	Buena-Excelente	4.5
2.5. Desarrollo de los ítems del test	Buena-Excelente	4.5
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	Buena-Excelente	4.5
2.9. Bibliografía del manual	No evaluado	-
2.10. Datos sobre el análisis de los ítems	Adecuada	3

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	Adecuada, pero con algunas carencias	2
2.11.1.2. Consultas a expertos	No se ha consultado a expertos sobre la representación del contenido	1

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 entre las puntuaciones del test y otras variables:

Contenido	Valoración	Puntuación
2.11.2.1.1. Diseños empleados	Correlaciones con otros test Diferencias entre grupos *Diseños experimentales y cuasi-experimentales	
2.11.2.1.2. Tamaño de las muestras	Varios estudios con muestras de tamaño moderado ($200 \leq N \leq 500$) o con alguna muestra grande ($N > 500$) y otras pequeñas ($N < 200$)	4
2.11.2.1.3. Procedimiento de selección de las muestras	Incidental. Depende de los estudios concretos, pero lo más frecuente es que se invite a participar a las personas mayores que acceden a centros de salud concretos, o que están en residencias de mayores.	
2.11.2.1.4. Calidad de los tests empleados como criterio o marcador	Buena-Excelente	4.5
2.11.2.1.5. Promedio de las correlaciones con otros tests que miden constructos similares	Buena ($0.60 \leq r < 0.70$)	4
2.11.2.1.6. Promedio de las correlaciones con otros tests que miden constructos no relacionados	Buena	4
2.11.2.1.7. Resultados de la matriz multirrasgo-multimétodo	-	-

2.11.2.1.8. Resultados de las diferencias intergrupo	Buena	4
--	-------	---

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	<p>Estudios españoles:</p> <p>Martínez de la Iglesia et al (2002): Entrevistas clínicas y neuropsicológicas protocolizadas. Se emplea una muestra no representativa de 249 personas mayores de 65 años. Se excluyen pacientes con alteración psiquiátrica grave no depresiva o con déficit motor y/o sensorial severo</p> <p>Franco et al. (2017) Entrevistas semiestructuradas ajustadas a los parámetros DSM-IV. Se emplea una muestra no representativa de un centro de atención primaria formada por 301 pacientes mayores de 64 años. Se excluyen pacientes con deterioro cognitivo o problemas de comunicación severos.</p> <p>Otros estudios: Se consideran también distintos estudios de revisión y meta-análisis que incluyen muestras de diferentes países.</p> <p>Se utiliza como gold-standard entrevistas semiestructuradas.</p>	
2.11.2.2.2. Calidad de los criterios empleados	Buena-Excelente	4.5
2.11.2.2.3. Relación temporal entre test y criterio	Concurrente, predictivo	
2.11.2.2.4. Tamaño de las muestras	Varios estudios con muestras grandes ($N \geq 200$) o estudios meta-analíticos.	5
2.11.2.2.5. Procedimiento de selección de las muestras	Incidental. Se emplean distintos procedimientos en los distintos estudios, aunque suelen ser personas que acuden a un determinado centro de salud (entre los cuáles, en ocasiones, se hace una selección aleatoria) o personas mayores residentes o usuarios de centros para mayores concretos.	
2.11.2.2.6. Promedio de las correlaciones del test con los criterios	Buena (valores de AUC entre 0.8 y 0.9)	4

2.11.3. Evidencias de validez basadas en la estructura interna:

Contenido	Valoración	Puntuación
-----------	------------	------------

2.11.3.1. Resultados del análisis factorial	Adecuada, pero con algunas carencias	2
2.11.3.2. Funcionamiento diferencial de los ítems	Excelente	5

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	Sí. Se pueden dar instrucciones orales en vez de responder mediante lápiz y papel, que es el procedimiento originalmente propuesto.	

2.11.5. Comentarios generales sobre evidencias de validez:

A continuación, se resumen las principales fuentes de evidencia (de más a menos importante):

En cuanto a la validez basada en la predicción de un criterio, el aspecto clave, dada la finalidad del test, es su capacidad para diferenciar entre personas mayores con problemas de depresión y aquellas que no presentan dichos problemas. Tanto los estudios con muestras españolas como los estudios meta-analíticos realizados con muestras diversas apuntan a que la prueba es eficaz y eficiente como prueba de cribado de depresión en personas mayores, alcanzando niveles adecuados de sensibilidad y especificidad: en las muestras españolas alcanzan valores en torno a 0.811 y 0.767, respectivamente, para el punto de corte recomendado de 5 puntos o más. Convendría realizar estudios adicionales que deberán reconfirmar el punto de corte y su utilidad para distintos grupos muestrales, por ejemplo, con distinto grado de deterioro cognitivo u otros problemas.

Evidencias de validez de relaciones con otras variables. Existen evidencias sobre las relaciones con otras variables (convergente y discriminante). Generalmente, la validez convergente se ha valorado en relación con otras escalas de depresión. Por ejemplo, Martínez de la Iglesia et al. (2002) encuentran correlaciones de Spearman de 0.618 con el cuestionario de Montgomery-Asberg de detección de depresión. En una muestra de personas mayores con demencia, Lucas-Carrasco (2012) encuentra correlaciones con la escala de depresión de Cornell (0.450). La validez discriminante se ha evaluado con respecto a escalas de cribado para alteraciones cognitivas. Por ejemplo, Martínez de la Iglesia et al. (2002) encuentran correlaciones de 0.235 con el SPMSQ-VE para despistaje de alteración cognitiva y Lucas-Carrasco (2012) encuentra correlaciones bajas con medidas más relacionadas con las alteraciones cognitivas, como el Mini-Mental State Examination (MMSE, $r = 0.073$) o el Índice de Barthel, para medir el nivel funcional en actividades de la vida diaria ($r = -0.058$). También se han encontrado las correlaciones esperadas con otros constructos relacionados. Por ejemplo, Lucas-Carrasco (2012) encuentra correlativas negativas esperadas con las dimensiones medidas por un cuestionario de calidad de vida, el WHOQOL-BREF: -0.465 (salud física), -0.593 (salud psicológica), -0.232 (relaciones sociales) y -0.423 (ambiente).

Evidencias de validez de estructura interna. Existe un meta-análisis de 26 estudios en distintas lenguas (Kim et al., 2013), que sugiere que, globalmente, hay 4 factores con significado, encontrándose diferencias a través de distintas culturas (con un número de factores que oscila entre 4 y 6). Tomando los resultados en conjunto, hay 3 factores que aparecen de forma sistemática: disforia, apatía-aislamiento social-deterioro cognitivo, y afecto positivo (siendo este último el de configuración más estable). También existe un estudio (Brown et al., 2007) que sugiere que la estructura factorial depende de las muestras (si tienen problemas de demencia o no). Mientras que en personas sin problemas o con cierto grado deterioro cognitivo se obtiene una solución de dos factores (Satisfacción con la vida y Afecto depresivo general), en una muestra psiquiátrica con problemas de depresión se obtiene un único factor.

Solo hay un estudio factorial, de la adaptación analizada aquí, con muestra española (población de personas mayores con demencia) (Lucas-Carrasco, 2012). Los resultados no son concluyentes puesto que no emplean los métodos recomendados a la hora de realizar un Análisis

Factorial Exploratorio y están basados en una muestra pequeña (N=91). Se encuentran dos factores, uno afectivo y uno mixto con dos ítems con saturaciones por debajo de .40 que no parece tener una interpretación clara. Dadas las diferencias observadas en los análisis factoriales a través culturas y tipos de muestras (clínicas o no), es necesario ampliar el número de estudios que tengan en cuenta estas variables y que empleen muestras locales.

Estudios de DIF. Existen algunas evidencias basadas en el Funcionamiento Diferencial de los Ítems (en función de sexo y edad, o el grado de deterioro cognitivo -excluyéndose personas con demencia severa o estados avanzados de Alzheimer), aunque los estudios se han realizado en otros países (p. ej., Italia, Estados Unidos; Chiesi et al., 2018a y b; Midden et al., 2018). Los resultados muestran que, aunque existe DIF en algún ítem, éste es de escasa magnitud y no provoca diferencias en el funcionamiento del test. Estas últimas evidencias deberían replicarse empleando muestras locales.

Evidencias de validez de contenido. No se aporta información sobre validez de contenido. En el proceso de creación del test participaron expertos en psiquiatría geriátrica, pero no hay estudios sistemáticos de validación basada en el contenido de la prueba. Para el desarrollo de la escala original se partió de un grupo de 100 ítems expresando 7 manifestaciones comunes de la depresión en personas mayores: preocupaciones somáticas, disminución del afecto, deterioro cognitivo, sentimientos de discriminación, reducción de la motivación, falta de orientación hacia el futuro y falta de autoestima. Posteriormente, se seleccionaron los ítems con mayor correlación ítem-total en un estudio piloto. Estos ítems incluían disminución del afecto, inactividad, irritabilidad, retraimiento, pensamientos angustiosos y juicios negativos sobre el pasado, el presente y el futuro. Los ítems relativos a preocupaciones somáticas quedaron descartados (probablemente por su mayor frecuencia en esta población). Para la formación de la escala corta de 15 ítems se siguió una aproximación también empírica, tratando de discriminar entre personas con y sin problemas de depresión.

Referencias:

Brown, P. J., Woods, C. M., & Storandt, M. (2007). Model stability of the 15-item Geriatric Depression Scale across cognitive impairment and severe depression. *Psychology and aging*, 22(2), 372-379.

Chiesi, F., Primi, C., Pigliatile, M., Baroni, M., Ercolani, S., Boccardi, V., ... & Mecocci, P. (2018a). Is the 15-item Geriatric Depression Scale a fair screening tool? A differential item functioning analysis across gender and age. *Psychological reports*, 121(6), 1167-1182.

Chiesi, F., Primi, C., Pigliatile, M., Baroni, M., Ercolani, S., Paolacci, L., ... & Mecocci, P. (2018b). Does the 15-item Geriatric Depression Scale function differently in old people with different levels of cognitive functioning?. *Journal of affective disorders*, 227, 471-476.

Kim, G., DeCoster, J., Huang, C. H., & Bryant, A. N. (2013). A meta-analysis of the factor structure of the Geriatric Depression Scale (GDS): the effects of language. *International psychogeriatrics*, 25(1), 71-81.

Lucas-Carrasco, R. (2012). Spanish version of the geriatric depression scale: Reliability and validity in persons with mild-moderate dementia. *International Psychogeriatrics*, 24(8), 1284-1290.

Martínez de la Iglesia, J., Onís, M.C., Dueñas, H.R., Albert, C.C., Aguado, T.C. y Luque, L.R. (2002). Versión española del cuestionario de Yesavage abreviado (GDS) para el despistaje de depresión en mayores de 65 años: adaptación y validación. *Revista de Medicina Familiar y Comunitaria*, 12, 620-630.

Midden, A. J., & Mast, B. T. (2018). Differential item functioning analysis of items on the Geriatric Depression Scale-15 based on the presence or absence of cognitive impairment. *Aging & mental health*, 22(9), 1142-1148.

2.12. Fiabilidad:

Contenido	Valoración	Puntuación
2.12.1. Datos aportados sobre fiabilidad	Varios coeficientes de fiabilidad (para cada escala o subescala) Coeficientes de fiabilidad para diferentes grupos de personas Cuantificación del error mediante TRI (Función de información u otros)	

2.12.2. Equivalencia formas paralelas:

Contenido	Valoración	Puntuación
2.12.2.1. Tamaño de las muestras	-	-
2.12.2.2. Puesta a prueba de los supuestos de paralelismo	-	-
2.12.2.3. Promedio de coeficientes de equivalencia	-	-

2.12.3. Consistencia interna:

Contenido	Valoración	Puntuación
2.12.3.1. Tamaño de las muestras	Varios estudios con muestras de tamaño moderado ($200 \leq N < 500$) o con alguna muestra grande ($N \geq 500$) y otras pequeñas ($N < 200$)	4
2.12.3.2. Coeficientes de consistencia interna presentados	Coeficiente Alfa o KR-20	
2.12.3.3. Promedio de coeficientes de consistencia	Buena ($0.80 \leq r < 0.85$)	4

2.12.4. Estabilidad (test-retest):

Contenido	Valoración	Puntuación
2.12.4.1. Tamaño de las muestras	Varios estudios con muestras de tamaño moderado ($100 \leq N < 200$) o con alguna muestra grande ($N \geq 200$) y otras pequeñas	4

	(N < 100).	
2.12.4.2. Coeficientes de estabilidad	Excelente ($r \geq 0.80$), pero se requieren estudios con muestras locales.	4.5

2.12.5. Cuantificación de la precisión mediante TRI:

Contenido	Valoración	Puntuación
2.12.5.1. Tamaño de las muestras	Varios estudios con muestras grandes	5
2.12.5.2. Coeficientes proporcionados	Función de información	
2.12.5.3. Tamaño de los coeficientes	Buena o Excelente (información ≥ 10)	4.5

2.12.6. Fiabilidad inter-jueces:

Contenido	Valoración	Puntuación
2.12.6.1. Tipos de coeficientes presentados	Índice de Kappa ponderado, tanto interobservador como intraobservador -para punto de corte de 5.	
2.12.6.2. Promedio de los coeficientes	Adecuado ($0.60 \leq r < 0.70$)	3

2.12.7. Comentarios generales sobre evidencias de fiabilidad:

Los distintos estudios evaluados muestran en general buenos resultados sobre la fiabilidad de la prueba, con distintas muestras de distintos tamaños. El estudio de la fiabilidad de GDS-VE se ha llevado a cabo desde distintas aproximaciones y en varios países. En general, los valores de fiabilidad (consistencia, inter-jueces, TRI) son muy buenos, tanto en muestras clínicas como muestras no-clínicas. En un estudio sobre generalización de la fiabilidad, Kieffer y Reese (2002) encuentran valores promedio de 0.82 para la escala de 15 ítems en su versión original. A continuación, se resumen los resultados según cada tipo de aproximación al estudio de la fiabilidad.

Consistencia interna. Existen distintos estudios con muestras locales que avalan la consistencia interna de las puntuaciones obtenidas con el GDS-15. Por ejemplo, Lucas-Carrasco (2012) informa de una consistencia interna de 0.81, para personas con demencia suave o moderada (N = 91). Martínez de la Iglesia et al. (2005) informan de un coeficiente de consistencia interna de 0.8. En su revisión sobre la escala GDS original, Kieffer y Reese (2011) encuentran valores de consistencia interna entre 0.74 y 0.86.

Fiabilidad test-retest. En el estudio de Martínez de la Iglesia et al. (2002) se obtiene la fiabilidad intra-observador (50 sujetos fueron evaluados por el mismo profesional dos veces, con un intervalo de 7-10 días), obteniéndose un coeficiente kappa ponderado de 0.951 y correlación intraclase de 0.96. Aparentemente, estos valores se calculan para el criterio de clasificación diagnóstica, puesto que se indica que se obtienen para un punto de corte concreto (5 o más). Con respecto a la versión original en inglés, los resultados de fiabilidad test-retest para las puntuaciones (en plazos de una y dos semanas) avalan la fiabilidad de las puntuaciones de la prueba, estimada como estabilidad temporal. En la revisión de Smarr y Keefer (2011) se informa de valores entre 0.84 y 0.85, para intervalos entre 1 y 2 semanas. En su estudio de generalización de la fiabilidad, Kieffer y Reese (2002) encuentran valores promedio de 0.84 (sin distinguir si estos valores se obtienen con la escala de 15 ítems o con la de 30). En cualquier caso, se deberían realizar estudios con muestras locales que estimen estos indicadores de fiabilidad para esta versión adaptada.

Fiabilidad inter-observadores. En el estudio original de Martínez de la Iglesia et al. (2002) se obtiene la fiabilidad inter-jueces (30 sujetos fueron evaluados por dos profesionales, de forma independiente y en el mismo momento), obteniéndose un coeficiente Kappa de 0.655, y correlación intraclase de 0.89. Aparentemente, estos valores se calculan para el criterio de clasificación diagnóstica, puesto que se indica que se obtienen para un punto de corte concreto (5 o más).

Fiabilidad basada en la TRI. En cuanto a los estudios de TRI, existen algunos estudios desarrollados con muestras no locales (p.ej., Chiesi et al., 2017, en muestra italiana; Zhang et al., 2020, en muestra china). Las funciones de información toman valores que pueden ser clasificados como buenos (en la muestra italiana) o excelentes (en la muestra china). Los dos estudios revisados muestran que la función de información más alta, correspondiente con niveles medios-altos en el rasgo latente. En concreto, en el estudio con muestra italiana se observa que, al hacer la equiparación entre niveles de rasgo latentes estimados mediante el modelo de TRI y la puntuación observada, los resultados sugieren que los niveles de información son adecuados alrededor del punto de corte recomendado de 5 (en concreto entre 5 y 6), aunque alcanzan el pico informativo con un nivel de rasgo (i.e. síntomas de depresión) algo superior. En el estudio con muestra china, el punto de corte para diagnosticar síntomas depresivos es 7, mientras que se considera depresión moderada y alta a partir de 10 y 13 puntos, respectivamente. Es para el punto de corte de depresión moderada para el que la prueba muestra funciones de información más elevadas, aunque para los 3 puntos de corte ofrece resultados satisfactorios. Este tipo de estudios deben ser replicados con muestra española, donde no se han encontrado estudios de TRI para el GDS-15.

Referencias:

- Chiesi, F., Primi, C., Pigliatile, M., Ercolani, S., Della Staffa, M. C., Longo, A., ... & Mecocci, P. (2017). The local reliability of the 15-item version of the Geriatric Depression Scale: An item response theory (IRT) study. *Journal of psychosomatic research*, 96, 84-88.
- Kieffer, K. M., & Reese, R. J. (2002). A reliability generalization study of the Geriatric Depression Scale. *Educational and Psychological Measurement*, 62(6), 969-994.
- Lucas-Carrasco, R. (2012). Spanish version of the geriatric depression scale: Reliability and validity in persons with mild-moderate dementia. *International Psychogeriatrics*, 24(8), 1284-1290.
- Martínez de la Iglesia, J., Onís, M.C., Dueñas, H.R., Albert, C.C., Aguado, T.C. y Luque, L.R. (2002). Versión española del cuestionario de Yesavage abreviado (GDS) para el despistaje de depresión en mayores de 65 años: adaptación y validación. *Revista de Medicina Familiar y Comunitaria*, 12, 620-630.
- Martínez de la Iglesia, J., Onís, M.C., Dueñas, H.R., Albert, C.C., Aguado, T.C., Colomer, A., Arias, C., Blanco, MC. (2005). Abbreviating the brief. Approach to ultra-short versions of the Yesavage questionnaire for the diagnosis of depression. *Atención Primaria*. 2005;35(1):14-21.
- Smarr, K. L., & Keefer, A. L. (2011). Measures of depression and depressive symptoms: BDI-II, CES-D, GDS, HADS, and PHQ-9. *Arthritis Care and Research*, 63(S11), 454-466.
- Zhang, C., Zhang, H., Zhao, M., Liu, D., Zhao, Y., & Yao, Y. (2020). Assessment of Geriatric Depression Scale's applicability in longevous persons based on classical test and item response theory. *Journal of Affective Disorders*, 274, 610-616.

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	Un baremo adecuado para la población objetivo	3
2.13.1.2. Tamaño de las muestras	Muy grande ($N \geq 1000$)	5
2.13.1.3. Aplicación de estrategia de tipificación continua	No	
2.12.1.4. Procedimiento de selección de las muestras	Incidental: Parece que es incidental pero no se dice explícitamente. Solo se indica que los datos provienen de tres estudios de cohorte llevados donde se recogieron datos entre 2008 y 2019 (Aging Brain Projects of the Complutense University of Madrid, the Vallecas Project, and the Compostela Aging Study)	
2.12.1.5. Actualización de baremos	Buena (entre 10 y 14 años)	4

2.13.2. Interpretación referida a criterio:

Contenido	Valoración	Puntuación
2.13.2.1. Adecuación del establecimiento de los puntos de corte	Excelente	5
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	-	-

2.13.3. Comentarios generales sobre baremación e interpretación de las puntuaciones:

El test se emplea principalmente con fines de cribado de la depresión en personas mayores, haciéndose principalmente una interpretación criterial. Como se ha indicado anteriormente en el apartado de validez referida a un criterio, tanto los estudios con muestras españolas como los estudios meta-analíticos realizados con muestras diversas y de distintos tamaños indican que la prueba es adecuada. En muestra española el punto de corte recomendado de 5 puntos o más, alcanza niveles satisfactorios de sensibilidad y especificidad. Algunos estudios internacionales sugieren distintos puntos de corte en función de la naturaleza de las muestras. Por ello, como se ha indicado anteriormente, conviene continuar investigando en el punto de corte óptimo para distintos grupos muestrales, por ejemplo, con distinto grado de deterioro cognitivo u otros problemas.

También desde 2021 se cuenta con baremos locales (datos recogidos entre 2008 y 2019) que

permiten hacer una interpretación normativa (Delgado Losada et al., 2021). La muestra es amplia, pero la conforman personas de menor edad de la considerada inicialmente por los autores de la prueba y su adaptación (la muestra de los baremos la integran personas entre 50 y 88 años. Sin embargo, la edad (igual que sucede con las variables género y nivel de estudios), no parece afectar a los resultados, por lo que pueden considerarse adecuados para hacer una interpretación normativa (siempre que se trate de personas sin deterioro cognitivo ni otros problemas que dieron lugar a la exclusión, como por ejemplo problemas psiquiátricos severos y enfermedades severas o adicciones). Los usuarios del GDS-15 que deseen hacer una interpretación normativa deberían tener en cuenta esta cuestión.