

Ampliación de la Global Scale of English (GSE) a la Global Scale of Languages (GSL)

Parte 1: Alineación de los objetivos de
aprendizaje del español con la GSL

Septiembre de 2023

Ying Zheng, University of Southampton

Catherine Doyle, Pearson

David Booth, Pearson

Mike Mayor, Pearson

Contenidos

Resumen ejecutivo.....	4
1. Presentación de la GSE y de sus objetivos de aprendizaje.....	5
2. Objetivo del estudio.....	5
3. Metodología	6
3.1 El juicio comparativo y sus aplicaciones.....	6
3.2 Diseño del estudio.....	7
3.2.1 Traducciones de los objetivos de aprendizaje: del inglés al español	8
3.2.2 Selección de evaluadores.....	8
3.2.3 Descripción del conjunto de datos	9
4. Resultados	9
4.1 Estadísticas de los jueces ponderadas según información	9
4.2 Estadísticas de los objetivos de aprendizaje ponderadas según información .	11
4.2.1 Comprensión oral	11
4.2.2 Comprensión lectora	13
4.2.3 Expresión e interacción oral.....	15
4.2.4 Expresión e interacción escrita	17
4.3 Correlaciones entre la GSE y las estimaciones de los juicios comparativos	19
4.4 Ecuación de transformación para cada competencia.....	20
5. Debate y conclusiones	21
Glosario	22
Referencias	22
Apéndice: Estadísticas demográficas de los evaluadores	24

Resumen ejecutivo

La Global Scale of English (GSE) ofrece un medio más detallado para describir y evaluar el progreso y el rendimiento de los estudiantes de inglés. Pearson ha llevado a cabo una amplia investigación (**Pearson**) sobre el uso de los objetivos de aprendizaje de la GSE como escala de referencia para ampliar el conjunto de las definiciones de las capacidades lingüísticas expresadas en el Marco Común Europeo de Referencia para las Lenguas (MCER) de 2001 con el fin de abordar las necesidades de más estudiantes.

Este estudio comparó el orden de clasificación de los objetivos de aprendizaje de la GSE traducidos al español para establecer si los valores de la GSE existentes se pueden aplicar a los estudiantes adultos de español como lengua extranjera (ELE). Se tradujeron al español 320 objetivos de aprendizaje. Se invitó a un panel de 20 evaluadores cualificados de un grupo de profesores de ELE a realizar 25 comparaciones de juicios comparativos por objetivo de aprendizaje, por lo que se obtuvieron 16 000 puntos de datos. Se realizó una serie de análisis, incluidas las estadísticas ajustadas de los evaluadores y los ítems, para evaluar la dificultad de los objetivos de aprendizaje del inglés existentes en los logits del modelo de Rasch y compararlos con las estimaciones de juicio comparativo de las versiones en inglés y español de cuatro competencias lingüísticas. Las ecuaciones de transformación se derivaron de estas comparaciones para alinear los resultados de los objetivos de aprendizaje del español con la GSE existente, lo que en última instancia dio lugar a la creación de una nueva Global Scale of Spanish.

Para obtener más información sobre la Global Scale of Languages, visita pearson.com/languages.

1. Presentación de la GSE y de sus objetivos de aprendizaje

La GSE es una escala estandarizada de competencia del inglés que va del 10 al 90 y está psicométricamente alineada con el Marco Común Europeo de Referencia para las lenguas (MCER, Consejo de Europa, 2001). Se ha desarrollado un conjunto de objetivos de aprendizaje de la GSE para describir la competencia del estudiante en cada punto de la escala, y se ha incorporado y ampliado el conjunto de descriptores del MCER. Estos objetivos de aprendizaje han sido evaluados por profesores de inglés como lengua extranjera (EFL) y calibrados con la Global Scale of English (De Jong, Mayor y Hayes, 2016). A diferencia del MCER y de otras escalas que describen la adquisición de una lengua utilizando niveles amplios, la Global Scale of English identifica qué sabe hacer el estudiante en cada punto de la escala en lo que respecta a las competencias de expresión e interacción oral, comprensión oral, comprensión lectora y expresión e interacción escrita, con el fin de ofrecer una descripción más detallada del aumento del dominio del idioma. El trabajo para desarrollar los objetivos de aprendizaje de la GSE se basa en la investigación llevada a cabo por Brian North y el Consejo de Europa al crear el MCER (North, 2000) y la amplía. Pearson English ha desarrollado los objetivos de aprendizaje de la GSE durante varios años en colaboración con más de 6 000 docentes, autores de ELT y expertos lingüísticos de todo el mundo.

2. Objetivo del estudio

El objetivo de este estudio es comparar el orden de clasificación de los objetivos de aprendizaje de la GSE que han sido traducidos al español para ver si los valores existentes de la GSE se pueden aplicar a los estudiantes adultos de español como lengua extranjera, es decir, si se pueden poner en la misma escala. La hipótesis de trabajo es la siguiente: dado que la GSE se basa en el MCER, que en sí mismo es independiente del idioma, se cree que el orden general estará altamente correlacionado tanto con la GSE como con el MCER; y este proyecto se propone verificar esta hipótesis utilizando el enfoque del juicio comparativo.

3. Metodología

3.1 El juicio comparativo y sus aplicaciones

El juicio comparativo (j. c.) implica juicios holísticos de pares de trabajos de estudiantes llevados a cabo por un grupo de jueces independientes que determinan qué trabajo tiene el mayor constructo global especificado. El resultado es una matriz de decisión binaria del «ganador» y del «perdedor» de cada par, que luego se ajusta al modelo de Bradley-Terry (Bradley y Terry, 1952) para producir valores de parámetros (puntuaciones) y errores estándares para el trabajo de cada estudiante. El valor del parámetro permite crear un orden de clasificación escalado del trabajo del estudiante de «mejor» a «peor», que se puede utilizar con fines de evaluación, como la calificación.

Además de usarse en las juntas de exámenes británicas para analizar la comparabilidad entre tribunales (por ejemplo, Fearnley, 2000; Gray, 2000) y la comparabilidad de los estándares a lo largo del tiempo, y para mantener los estándares (por ejemplo, Chambers y Cunningham, 2022), los juicios comparativos también se han aplicado a una gran variedad de contextos educativos. Esto incluye la evaluación por pares de los informes de proyectos de pensamiento de diseño de estudiantes de grado (Mentzer et al., 2021), pruebas escritas sobre la comprensión conceptual de un curso de matemáticas (Jones y Alcock, 2014), la evaluación docente de estadísticas sumativas y evaluaciones de inglés (Marshall et al., 2020), ensayos (Steedle y Ferrara, 2016) y textos argumentativos (Lesterhuis et al., 2022). Pearson empleó juicios comparativos para alinear los objetivos de aprendizaje de la Global Scale of English (GSE) de los estudiantes jóvenes con la Chinese Scale of English proficiency (CSE) comparando la dificultad de los descriptores en cada estándar (Pearson, 2020).

La base psicológica de los juicios comparativos es que las personas son competentes al comparar un objeto con otro, pero poco fiables al evaluar objetos de forma aislada (Gill y Bramley, 2013; Thurstone, 1927). Los enfoques analíticos tradicionales implican que los docentes califiquen el trabajo de los estudiantes individualmente de manera absoluta utilizando rúbricas, lo que puede dar lugar a diferentes interpretaciones y aplicaciones de los descriptores de las rúbricas, así como a la posibilidad de hacer uso de su percepción del trabajo de otros estudiantes. Por el contrario, los juicios comparativos minimizan esta influencia comparativa de las rúbricas detalladas y específicas (Pollitt, 2004), y aprovechan el aspecto comparativo de la evaluación directamente, prescindiendo de las rúbricas y de la calificación. Otros estudios anteriores han establecido cómo los juicios comparativos cumplen los altos estándares de validez, fiabilidad y eficiencia.

3.2 Diseño del estudio

Para llevar a cabo este estudio se utilizó *NoMoreMarking* (Wheadon, 2019), una herramienta de juicios comparativos. El número de veces que se juzga un objeto determinado en comparación con otro es un elemento importante en un estudio de juicios comparativos. Verhavert *et al.* (2019) recomiendan tener de 10 a 30 comparaciones por objeto para garantizar una fiabilidad aceptable. De acuerdo con esta recomendación, se recopilaron 25 comparaciones por objetivo de aprendizaje para asegurar un buen diseño.

En este estudio, seleccionamos 320 objetivos de aprendizaje de la GSE para adultos, lo que supone el 30% del número total disponible. En lo que se refiere al tamaño y a la selección de la muestra, un 20% es, por lo general, la coincidencia mínima necesaria para alinear las escalas (Kolen y Brennan, 2004). La muestra está estratificada para representar tanto el número de objetivos de aprendizaje de cada una de las cuatro competencias como la cantidad en cada nivel del MCER (consulta la Tabla 1 a continuación).

Tabla 1: Distribución de los objetivos de aprendizaje

MCER/GSE	Comprensión oral	Comprensión lectora	Expresión e interacción oral	Expresión e interacción escrita	TOTAL	% de la base de datos
Por debajo del A1 (10-21)	3	3	10	4	20	34%
A1 (22-29)	5	5	14	8	32	27%
A2 (30-35)	6	6	17	10	39	30%
A2+ (36-42)	6	6	16	10	38	27%
B1 (43-50)	7	7	18	11	43	35%
B1+ (51-58)	7	7	18	11	43	33%
B2 (59-66)	7	7	19	11	44	28%
B2+ (67-75)	5	5	14	8	32	27%
C1 (76-84)	3	3	10	4	20	28%
C2 (85-90)	1	1	4	3	9	47%
TOTAL	50	50	140	80	320	30%
% de la base de datos	26%	35%	28%	33%	30%	

También se tuvo en cuenta la diversidad y la amplitud de las funciones lingüísticas, además de evitar seleccionar objetivos de aprendizaje bastante similares. También se seleccionaron algunos de los descriptores originales del MCER (que también se incluyen en la GSE) para proporcionar enlaces estadísticos al modelo del MCER/de North.

3.2.1 Traducciones de los objetivos de aprendizaje: del inglés al español

Una agencia de traducción tradujo al español los 320 objetivos de aprendizaje de la GSE. A la agencia se le proporcionaron las traducciones oficiales del MCER del Consejo de Europa del inglés al español como guía. Tras las traducciones iniciales, se pidió a dos miembros del equipo editorial de Pearson en España que revisaran el trabajo. Se identificaron y corrigieron problemas menores en la traducción antes de utilizar la versión final en español de los objetivos de aprendizaje.

Para garantizar que las versiones en español e inglés se evaluaran con el mismo marco de referencia, las versiones en español e inglés de cada objetivo de aprendizaje se pusieron en el mismo grupo de evaluación (dividido por competencias), por lo que los evaluadores veían dos en español, dos en inglés o uno de cada.

3.2.2 Selección de evaluadores

La junta de exámenes de Pearson Edexcel seleccionó a los evaluadores de entre un grupo de profesores de español como lengua extranjera que trabajaban o habían trabajado calificando exámenes de español de GCSE o de A-level (títulos de la enseñanza secundaria/preuniversitaria en el Reino Unido). Un total de 138 personas expresaron interés en participar en la investigación y proporcionaron información sobre su trayectoria. En función de experiencia en la enseñanza a estudiantes adultos y de su grado de conocimiento del MCER, se seleccionaron 20 evaluadores para el proyecto. También se tuvo en cuenta la creación de un grupo de evaluadores lo más diverso posible en términos de género, nacionalidad y experiencia. Además, todos los evaluadores tenían experiencia en la enseñanza de al menos otro idioma aparte del español (consulta el apéndice para ver las estadísticas demográficas de los evaluadores).

A los evaluadores se les proporcionaron instrucciones escritas sobre la tarea y la plataforma antes de pedirles que realizaran el juicio comparativo basado en esta pregunta: «¿Cuál de estos objetivos de aprendizaje describe la competencia más difícil para un estudiante de idiomas?».

3.2.3 Descripción del conjunto de datos

Tabla 2: Número de objetivos de aprendizaje y comparaciones para cada competencia

Competencia	Objetivos de aprendizaje del inglés	Objetivos de aprendizaje del español	Número total de juicios
Comprensión oral	50	50	2.500
Comprensión lectora	50	50	2.500
Expresión e interacción oral	140	140	7.000
Expresión e interacción escrita	80	80	4.000
TOTAL	320	320	16.000

4. Resultados

En el juicio comparativo, la escala de fiabilidad de separación (EFS) se utiliza para indicar la fiabilidad, en este caso, la del orden de clasificación de los objetivos de aprendizaje producidos por la actividad de juicios comparativos. La EFS consiste en una escala de 0 a 1, en la que los valores superiores a 0,90 indican una escala de juicios comparativos muy fiable. La Tabla 3 a continuación muestra la EFS para las cuatro competencias.

Tabla 3: Escala de fiabilidad de separación

Comprensión oral	Comprensión lectora	Expresión e interacción oral	Expresión e interacción escrita
0,939	0,938	0,938	0,943

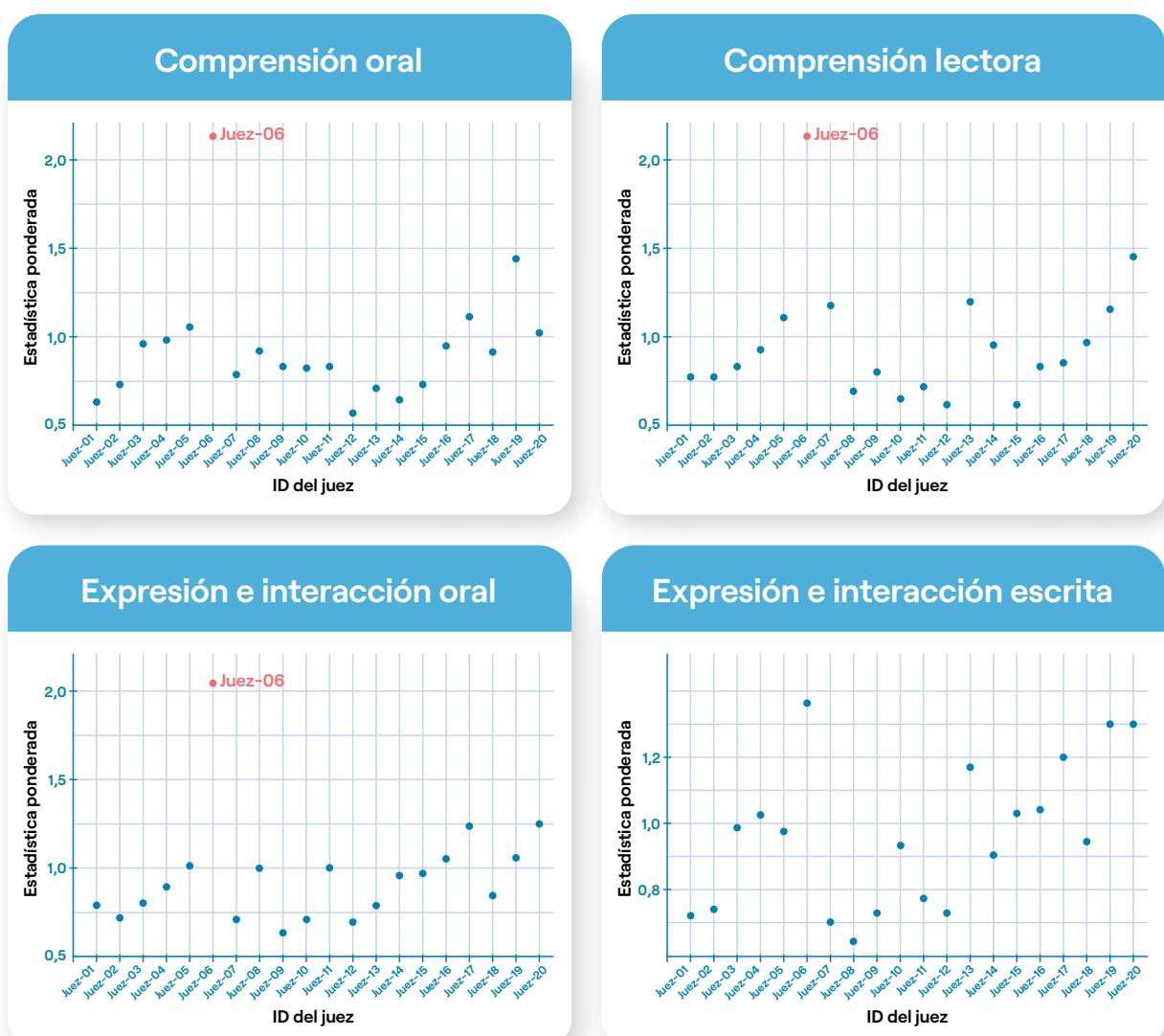
4.1 Estadísticas de los jueces ponderadas según información

Se calcularon estadísticas de ajuste tanto para los evaluadores como para los ítems (es decir, los objetivos de aprendizaje) utilizados en este ejercicio de juicio comparativo. Se excluyeron a los evaluadores cuyos datos ponderados según información estaban dos desviaciones estándares por encima de los datos promedio ponderados según información, ya que esto indicaba que podían haber realizado los juicios de manera irregular o que no se alineaban con el consenso de otros evaluadores. De los evaluadores, 19 de 20 (el 95 %) tenían

estadísticas ponderadas según información aceptables en los cuatro ejercicios. Un evaluador mostró datos no ajustados en comprensión oral, de lectura y expresión e interacción oral (posiblemente debido a una mala interpretación de la tarea). Aunque las estadísticas ponderadas según información de este evaluador en particular para expresión escrita no estaban fuera del rango aceptable, sus estadísticas no ajustadas de evaluación en esta tarea seguían siendo las más altas de todos los evaluadores.

Por lo tanto, los datos de su evaluación en las cuatro tareas se eliminaron de los análisis posteriores. Al eliminar los datos de este evaluador, la correlación general entre los valores de la GSE existentes y las puntuaciones de los juicios comparativos pasó de 0,79 a 0,93. Consulta la Figura 1 para ver una representación visual de las estadísticas del evaluador ponderadas según información.

Figura 1: Estadísticas del evaluador ponderadas según información (cuatro competencias)



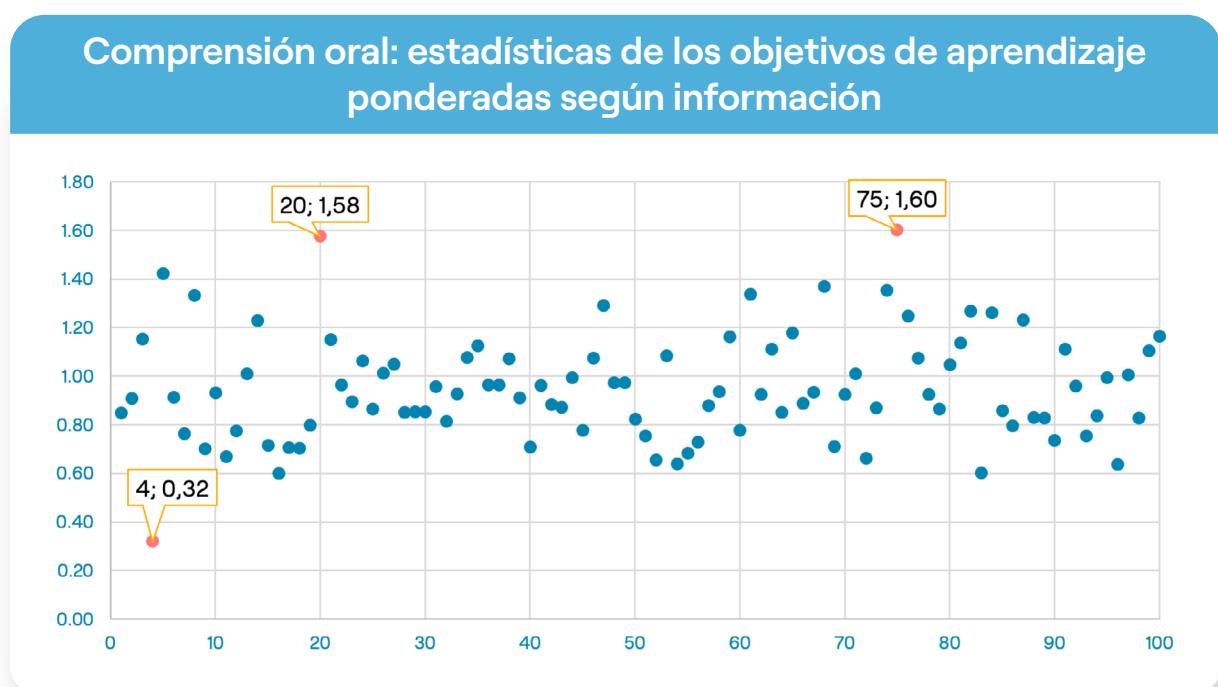
4.2 Estadísticas de los objetivos de aprendizaje ponderadas según información

Las siguientes secciones reflejan las estadísticas ponderadas según información de los objetivos de aprendizaje de las cuatro competencias. Las Figuras 2, 4, 6 y 8 muestran los diagramas de dispersión para cada competencia. En el eje y se indican las estadísticas ponderadas según información del ítem y en el eje x, el número de ítems. Hay 100 ítems de comprensión oral, 100 ítems de comprensión lectora, 160 ítems de expresión e interacción escrita y 260 ítems de expresión e interacción oral.

4.2.1 Comprensión oral

La correlación general entre los valores existentes de la GSE de los objetivos de aprendizaje de la comprensión oral (tanto en inglés como en español) y los valores de los juicios comparativos es de **0,931**.

Figura 2: Comprensión oral: estadísticas de los objetivos de aprendizaje ponderadas según información

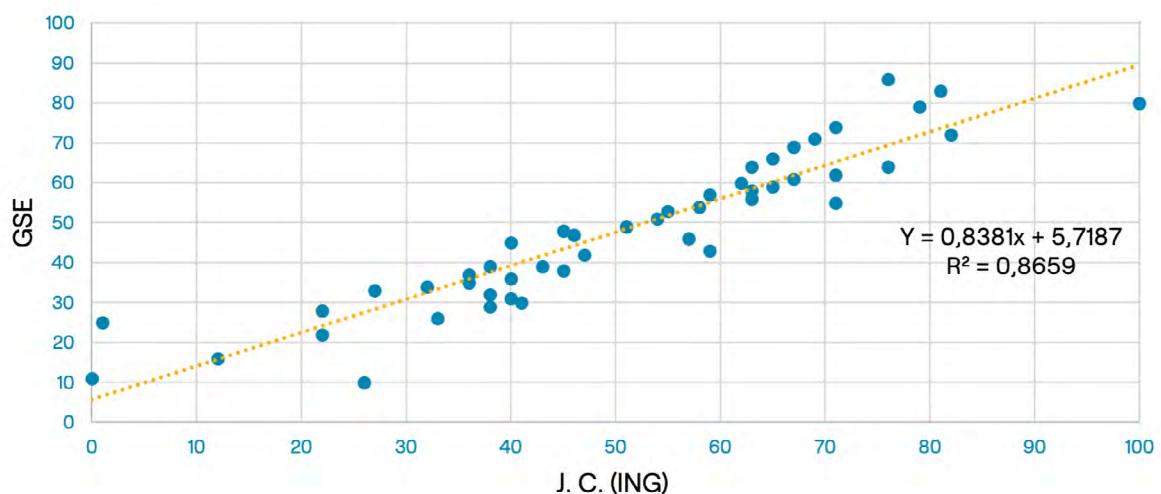


Los tres objetivos de aprendizaje destacados en la Figura 2 se marcaron para realizar más comprobaciones cualitativas. Como estaban dentro del rango de ponderación según información aceptable, se mantuvieron en el estudio. [Nota: En las dos figuras anteriores, los puntos rojos indican el ID del ítem seguido de su estadística ponderada según información]

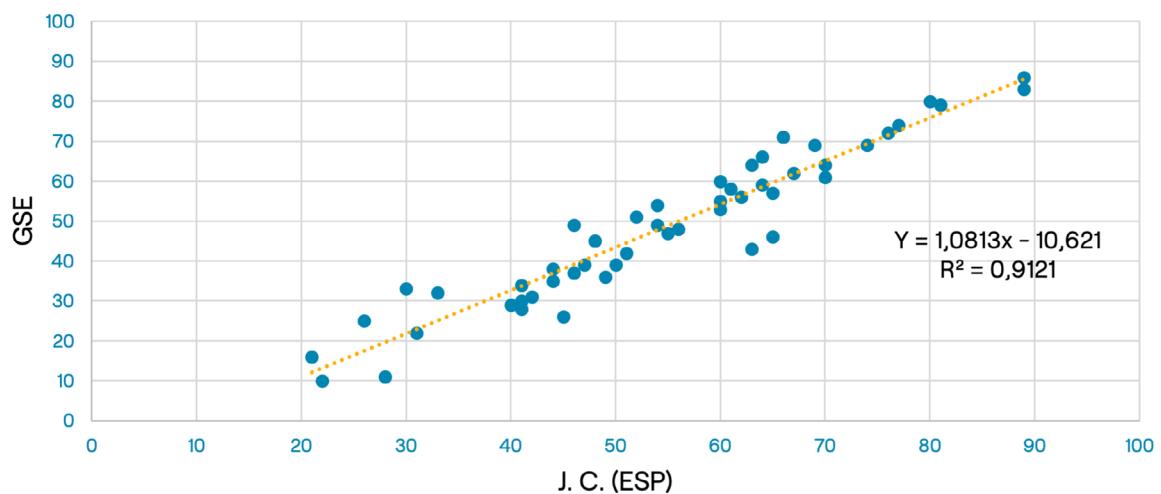
Los tres gráficos siguientes (Figura 3) muestran las comparaciones de los valores existentes de la GSE en la comprensión oral con las estimaciones de los juicios comparativos de los objetivos de aprendizaje del inglés y con las estimaciones de los juicios comparativos de los objetivos de aprendizaje del español, así como las dos estimaciones de los juicios comparativos de cada objetivo de aprendizaje en los dos idiomas. Para la comparación, se utilizaron puntuaciones estandarizadas de los juicios comparativos proporcionadas por *NoMoreMarking*. Los resultados indican un gran nivel de acuerdo entre estas comparaciones.

Figura 3: Comprensión oral: comparación de la dificultad de los objetivos de aprendizaje existentes con las estimaciones de los juicios comparativos

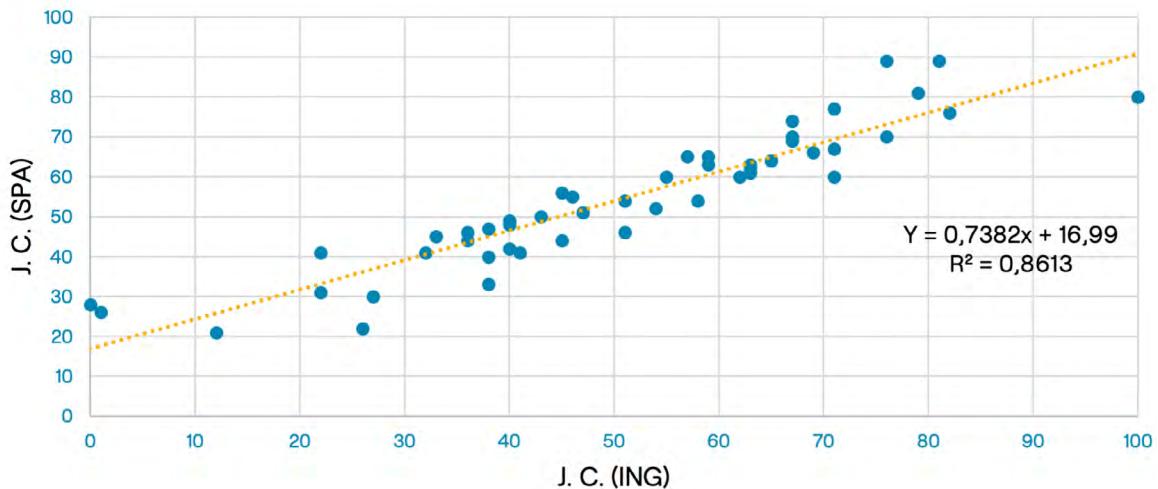
Comprensión oral: juicio comparativo (inglés) vs. GSE



Comprensión oral: juicio comparativo (español) vs. GSE



Comprensión oral: juicio comparativo (inglés) vs. juicio comparativo (español)

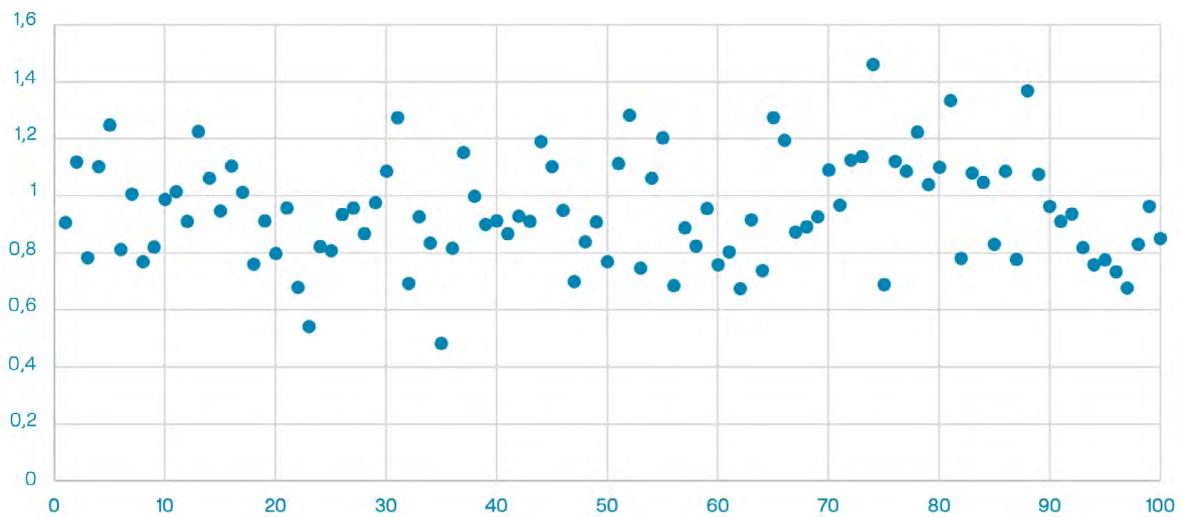


4.2.2 Comprensión lectora

La correlación general entre los valores existentes de la GSE de los objetivos de aprendizaje de la comprensión lectora (tanto en inglés como en español) y los valores de los juicios comparativos es de **0,923**. No se marcaron objetivos de aprendizaje para realizar más investigaciones (Figura 4).

Figura 4: Comprensión lectora: estadísticas de los objetivos de aprendizaje ponderadas según información

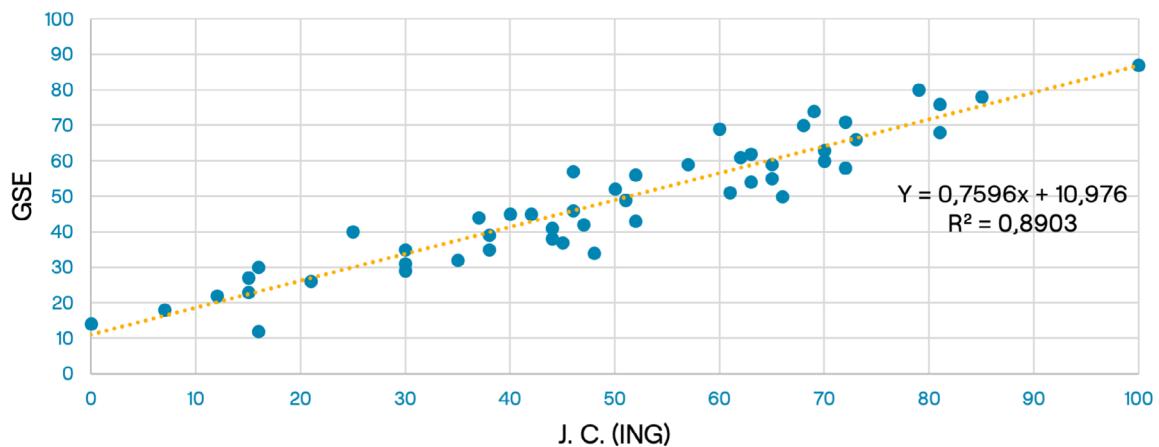
Comprensión lectora: estadísticas de los objetivos de aprendizaje ponderadas según información



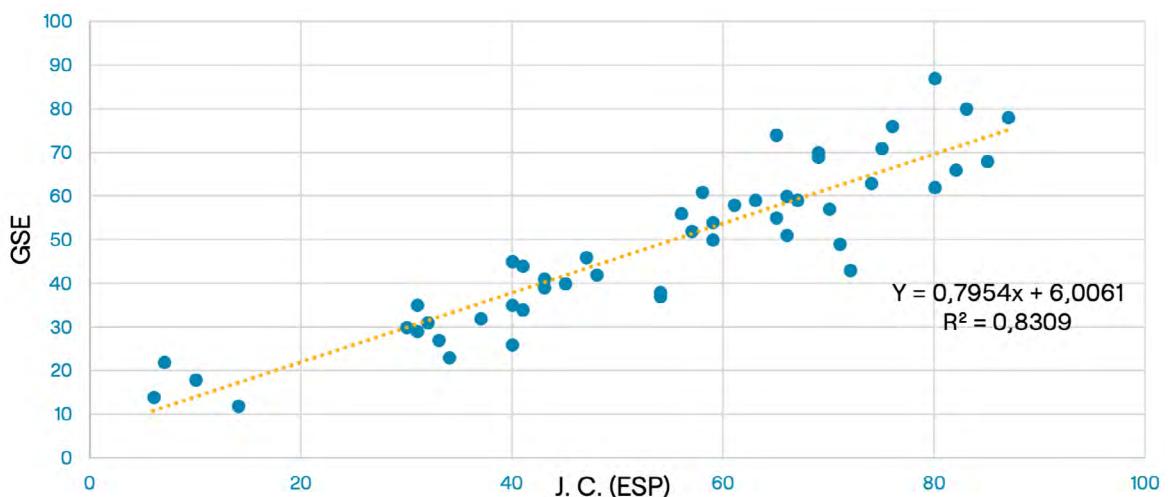
Los tres gráficos siguientes (Figura 5) muestran las comparaciones de los valores existentes de la GSE en la comprensión lectora con las estimaciones de los juicios comparativos de los objetivos de aprendizaje en inglés y con las estimaciones de los juicios comparativos de los objetivos de aprendizaje en español, así como las dos estimaciones de juicios comparativos de cada objetivo de aprendizaje en los dos idiomas. Los resultados indican un gran nivel de acuerdo entre estas comparaciones.

Figura 5: Comprensión lectora: comparación de la dificultad de los objetivos de aprendizaje existentes con las estimaciones de los juicios comparativos

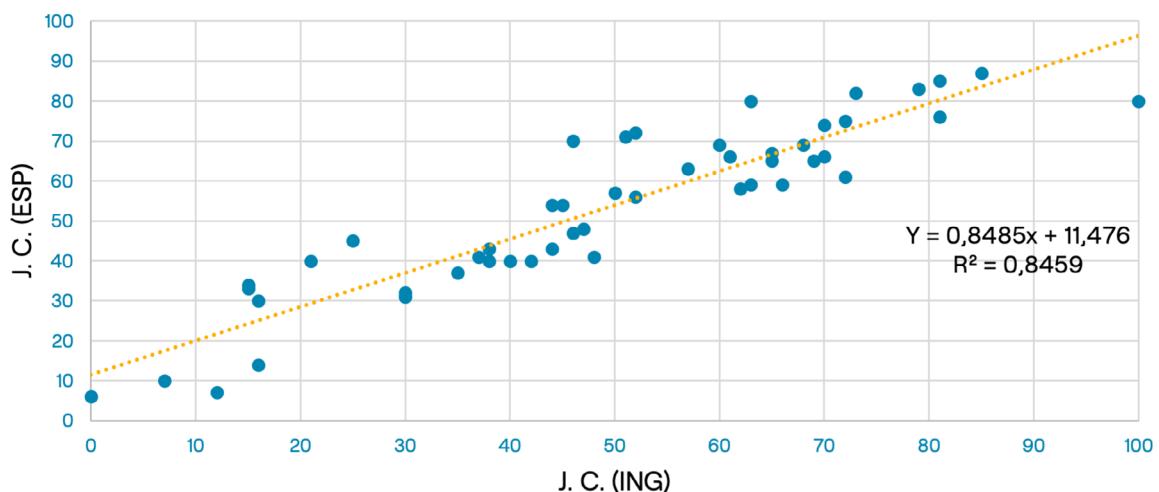
Comprensión lectora: juicio comparativo (inglés) vs. GSE



Comprensión lectora: juicio comparativo (español) vs. GSE



Comprensión lectora: juicio comparativo (inglés) vs. juicio comparativo (español)

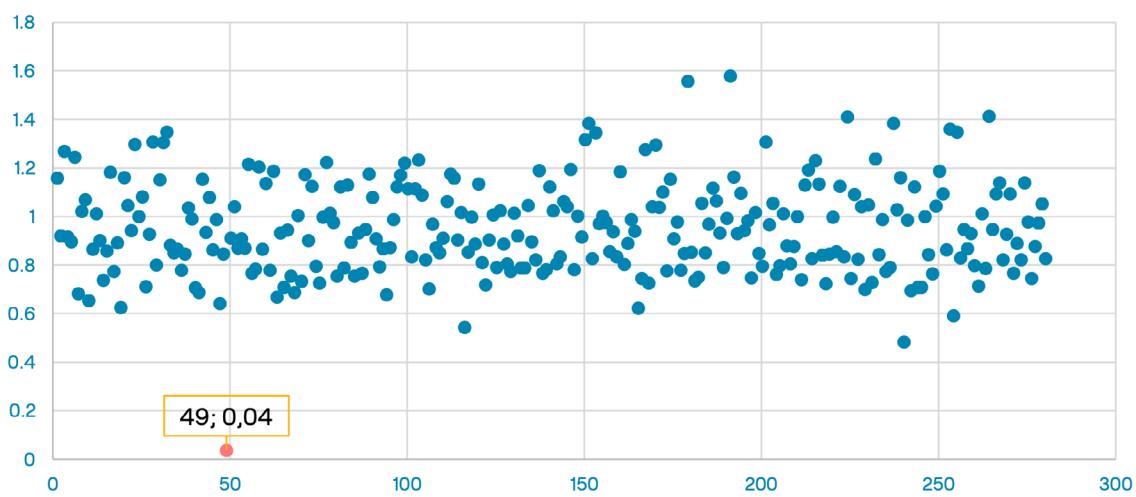


4.2.3 Expresión e interacción oral

La correlación general entre los valores existentes de la GSE de los objetivos de aprendizaje de la expresión e interacción oral (tanto en inglés como en español) y los valores de los juicios comparativos es de **0,917**. Se marcó un objetivo de aprendizaje para realizar más investigaciones (Figura 6). [Nota: Los dos números sobre el punto rojo indican el ID del ítem seguido de su estadística ponderada según información].

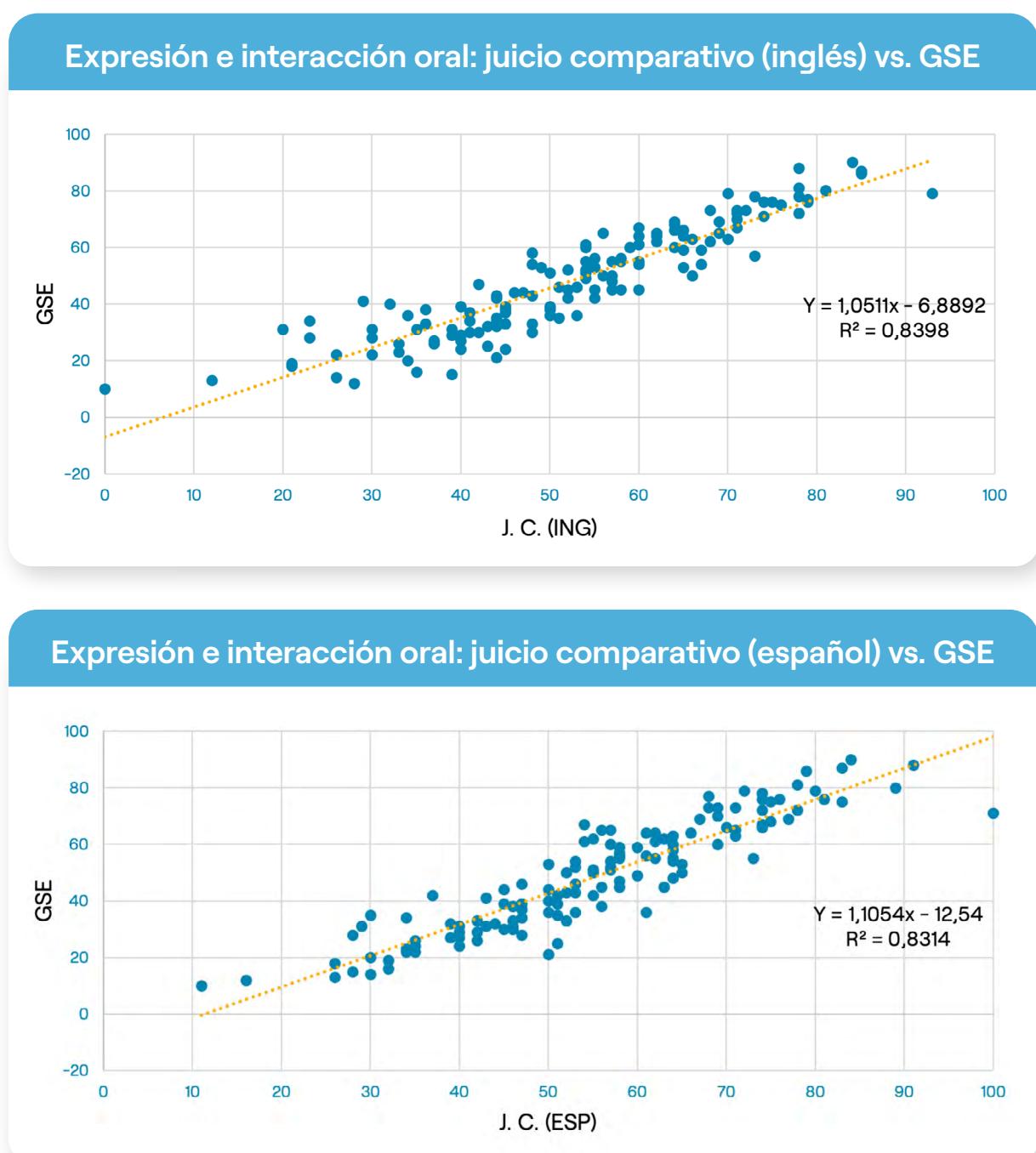
Figura 6: Expresión e interacción oral: estadísticas de los objetivos de aprendizaje ponderadas según información

Expresión e interacción oral: estadísticas de los objetivos de aprendizaje ponderadas según información

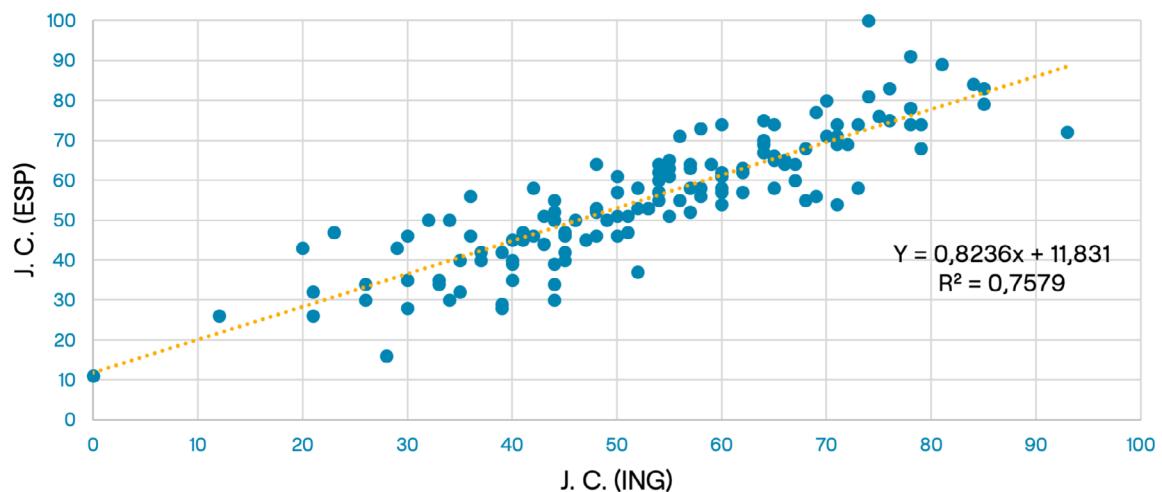


Los tres gráficos siguientes (Figura 7) muestran las comparaciones de los valores existentes de la GSE en la expresión e interacción oral con las estimaciones de los juicios comparativos de los objetivos de aprendizaje en inglés y con las estimaciones de los juicios comparativos de los objetivos de aprendizaje en español, así como las dos estimaciones de juicios comparativos de cada objetivo de aprendizaje en los dos idiomas. Los resultados indican un gran acuerdo entre estas comparaciones.

Figura 7: Expresión e interacción oral: comparación de la dificultad de los objetivos de aprendizaje existentes con las estimaciones de los juicios comparativos



Expresión e interacción oral: juicio comparativo (inglés) vs. juicio comparativo (español)

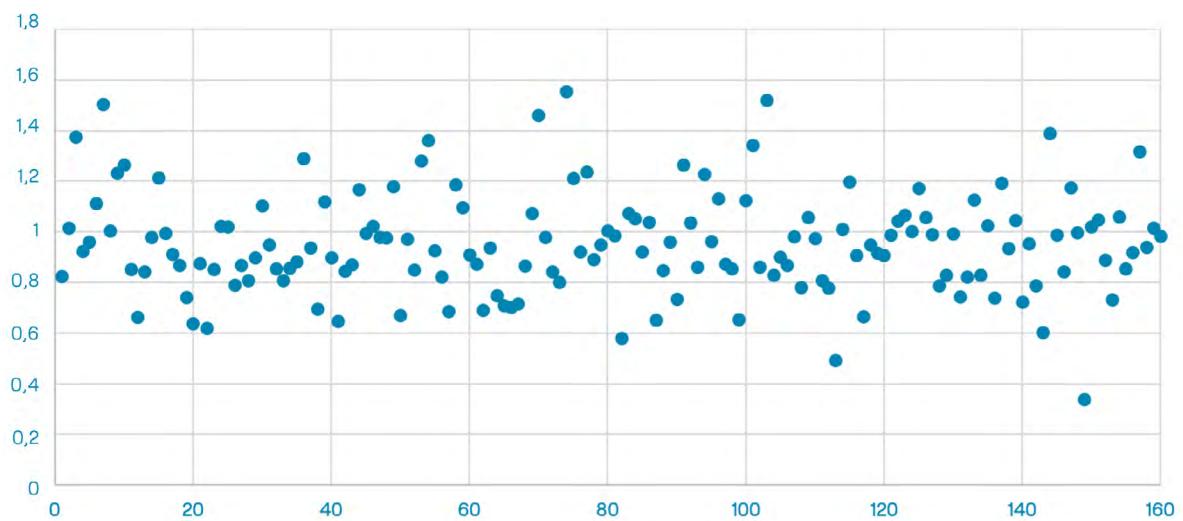


4.2.4 Expresión e interacción escrita

La correlación general entre los valores existentes de la GSE de los objetivos de aprendizaje de la expresión e interacción escrita (tanto en inglés como en español) y los valores de los juicios comparativos es de **0,925**. No se marcaron objetivos de aprendizaje para realizar más investigaciones (Figura 8).

Figura 8: Expresión escrita: estadísticas de los objetivos de aprendizaje ponderadas según información

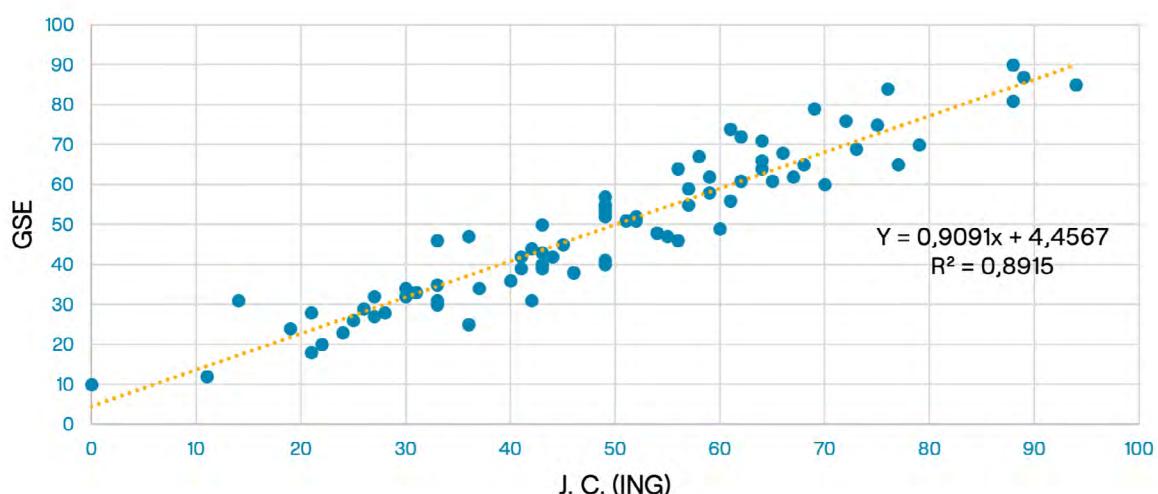
Expresión escrita: estadísticas de los objetivos de aprendizaje ponderadas según información



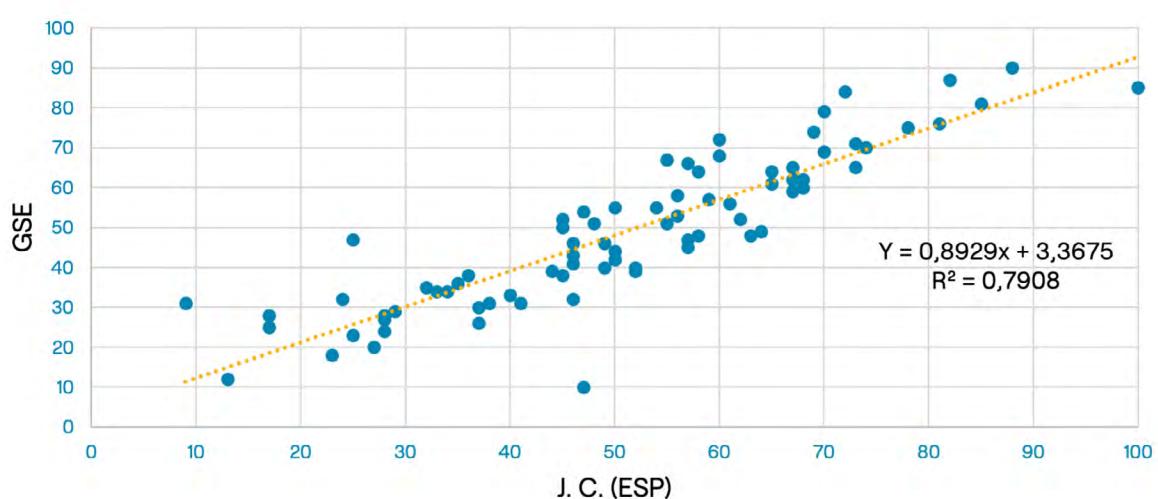
Los tres gráficos siguientes (Figura 9) muestran las comparaciones de los valores existentes de la GSE en la expresión e interacción escrita con las estimaciones de los juicios comparativos de los objetivos de aprendizaje en inglés y con las estimaciones de los juicios comparativos de los objetivos de aprendizaje en español, así como las dos estimaciones de juicios comparativos de cada objetivo de aprendizaje en los dos idiomas. Los resultados indican un gran nivel de acuerdo entre estas comparaciones.

Figura 9: Expresión e interacción escrita: comparación de la dificultad de los objetivos de aprendizaje existentes con las estimaciones de los juicios comparativos

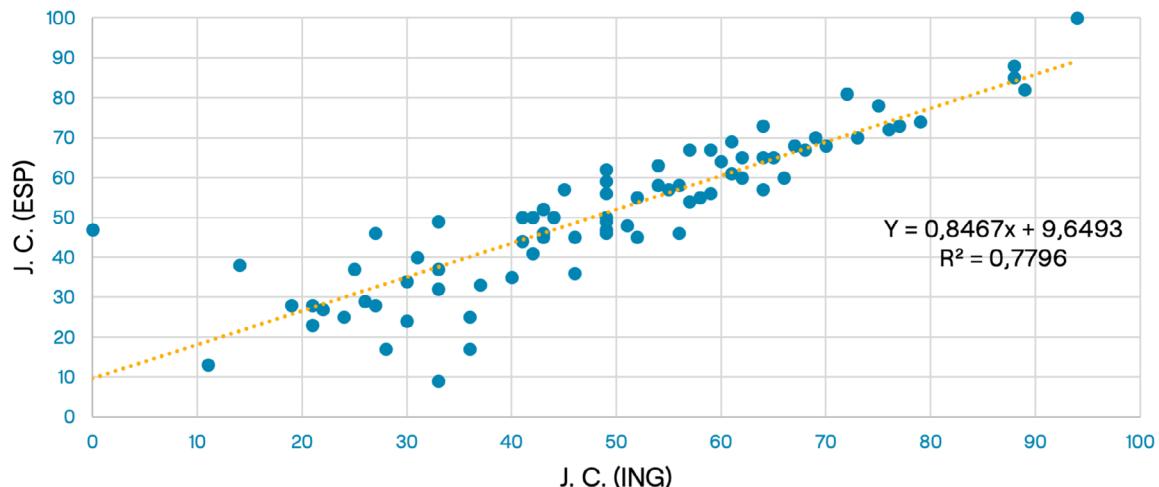
Expresión e interacción escrita: juicio comparativo (inglés) vs. GSE



Expresión e interacción escrita: juicio comparativo (español) vs. GSE



Expresión e interacción escrita: juicio comparativo (inglés) vs. juicio comparativo (español)



4.3 Correlaciones entre la GSE y las estimaciones de los juicios comparativos

Las correlaciones entre los valores existentes de los objetivos de aprendizaje de la GSE y los valores producidos en el estudio de los juicios comparativos se presentan en la tabla siguiente. Se realizaron comparaciones entre los valores de los juicios comparativos de los objetivos de aprendizaje de la GSE existentes y los de las traducciones al español. Salvo por una excepción (la de la expresión e interacción escrita), todas las correlaciones lograron más de un 0,9 en las comparaciones.

Tabla 4: Correlaciones entre los valores existentes de la GSE de los objetivos de aprendizaje y los valores producidos en el estudio de juicios comparativos por idioma y por competencia

	Valores existentes de la GSE vs. valores de los juicios comparativos (inglés)	Valores existentes de la GSE vs. valores de los juicios comparativos (español)
Comprensión oral	0,931	0,955
Comprensión lectora	0,944	0,912
Expresión e interacción oral	0,916	0,912
Expresión e interacción escrita	0,944	0,889

4.4 Ecuación de transformación para cada competencia

Según los resultados satisfactorios obtenidos hasta el momento, se generaron ecuaciones de transformación para cada competencia lingüística.

A continuación, la Tabla 5 demuestra las ecuaciones de transformación de las estimaciones de los juicios comparativos para los valores de la GSE.

Con fines demostrativos, se muestran dos ecuaciones para cada competencia: la primera, para transformar los valores de los juicios comparativos en inglés a la GSE y la segunda, para transformar los valores de los juicios comparativos en español a la GSL. Para usarse, solo se necesita la segunda ecuación para alinear los objetivos de aprendizaje del español con la escala global que va del 10 al 90.

Tabla 5: Ecuación de transformación de las estimaciones de los juicios comparativos para los valores de la GSE

x = puntuación escalada del juicio comparativo; y = GSE

Comprensión oral: juicio comparativo (inglés) vs. GSE

$$y = 0,8381x + 5,7187$$

Comprensión oral: juicio comparativo (español) vs. GSL

$$y = 1,0813x - 10,621$$

Comprensión lectora: juicio comparativo (inglés) vs. GSE

$$y = 0,7596x + 10,976$$

Comprensión lectora: juicio comparativo (español) vs. GSL

$$y = 0,7954x + 6,0061$$

Expresión e interacción oral: juicio comparativo (inglés) vs. GSE

$$y = 1,0511x - 6,8892$$

Expresión e interacción oral: juicio comparativo (español) vs. GSL

$$y = 1,1054x - 12,54$$

Expresión e interacción escrita: juicio comparativo (inglés) vs. GSE

$$y = 0,9091x + 4,4567$$

Expresión e interacción escrita: juicio comparativo (español) vs. GSL

$$y = 0,8929x + 3,3675$$

5. Debate y conclusiones

Los resultados de este estudio de juicios comparativos muestran una alta correlación entre los niveles de las competencias de cada objetivo de aprendizaje en inglés y español. Como validación adicional, se proporcionan ecuaciones de transformación específicas para cada competencia, lo que finalmente supuso el establecimiento de la concordancia entre las estimaciones de la dificultad de los objetivos de aprendizaje del español en la GSE y en el MCER, respectivamente.

El MCER en sí es un marco lingüístico neutral que «se puede adaptar y utilizar en varios contextos y aplicarse a todos los idiomas» (**Consejo de Europa**) y, desde su elaboración en 2001, se ha traducido a 40 idiomas (*ibid.*). El trabajo de Pearson para ampliar el MCER y crear la GSE se concibió originalmente dentro del contexto del inglés como lengua extranjera (EFL, por sus siglas en inglés). Sin embargo, se creía que esta extensión también podría ser adecuada y útil para docentes y estudiantes de otros idiomas. El estudio de juicios comparativos descrito en este documento ofrece pruebas para respaldar la consideración de que los actos comunicativos y funcionales del lenguaje expresados en las definiciones de las capacidades lingüísticas, tanto en inglés como en español, tienen un valor comparable en términos de competencia, es decir, ambos se pueden colocar en la misma escala.

Dada la similitud entre el español y otras lenguas romances como el francés y el italiano, confiamos en que estos resultados puedan extrapolarse y aplicarse en esos contextos, de la misma manera que el MCER se puede aplicar a estas lenguas europeas. Se están realizando más estudios para las lenguas no romances que se sumarán al cuerpo de pruebas de validación de la Global Scale of Languages.

Glosario

MCER: Marco Común Europeo de Referencia para las Lenguas

j. c.: juicio comparativo

CSE: China Scale of English

GSE: Global Scale of English

GSL: Global Scale of Languages

Referencias

Bradley, R. A. and Terry, M. E. (1952). *Rank analysis of incomplete block designs. I. The method of paired comparisons*. Biometrika 39, 324–345.

Chambers, L., & Cunningham, E. (2022). *Exploring the Validity of Comparative Judgement: Do Judges Attend to Construct-Irrelevant Features?* Frontiers in Education (7).

Council of Europe (2001). *Common European framework of reference for languages: learning, teaching, assessment*. Cambridge: Cambridge University Press.

De Jong, J., Mayor, M., & Hayes, C. (2016). *Developing Global Scale of English Learning Objectives aligned to the Common European Framework*. Disponible en <https://www.pearson.com/languages/why-pearson/the-global-scale-of-english/resources.html>

Fearnley, A. (2000). *A comparability study in GCSE mathematics. A study based on the summer 1998 examination*. En Assessment and Qualifications Alliance (Northern Examinations and Assessment Board). Manchester: Joint Forum for the GCSE and GCE.

Gill, T., & Bramley, T. (2013). How accurate are examiners' holistic judgements of script quality?. *Assessment in Education: Principles, Policy & Practice*, 20(3), 308–324.

Gray, E. (2000). *A comparability study in GCSE science 1998. A study based on the 1998 summer examination*. Organizado por Oxford, Cambridge and RSA Examinations (Midland Examining Group) en nombre del foro conjunto para GCSE y GCE.

-
- Jones, I., & Alcock, L. (2014). Peer assessment without assessment criteria. *Studies in Higher Education*, 39(10), 1774–1787.
- Kolen, M. J., & Brennan R. L. (2004). *Test equating, scaling, and linking: Methods and practices*. 2nd. New York: Springer.
- Lesterhuis, M., Verhavert, S., Coertjens, L., Donche, V., & De Maeyer, S. (2017). Comparative judgement as a promising alternative to score competences. In *Innovative practices for higher education assessment and measurement* (pp. 119–138). IGI Global.
- Marshall, N., Shaw, K., Hunter, J., & Jones, I. (2020). Assessment by comparative judgement: An application to secondary statistics and English in New Zealand. *New Zealand Journal of Educational Studies*, 55, 49–71.
- Mentzer, N., Lee, W., & Bartholomew, S. R. (2021). Examining the Validity of Adaptive Comparative Judgment for Peer Evaluation in a Design Thinking Course. En *Frontiers in Education* (p. 492). Frontiers.
- North, B. (2000). *The development of a common framework scale of language proficiency*. New York: Peter Lang.
- Pollitt, A. (2004). *Let's stop marking exams*, International Association for Educational Assessment Conference. Philadelphia PA.
- Steedle, J. T., & Ferrara, S. (2016). Evaluating comparative judgment as an approach to essay scoring. *Applied Measurement in Education*, 29(3), 211–223.
- Pearson technical report (2020): Aligning Global Scale of English-Young Learner to the CSE. Disponible en <https://m.i21st.cn/elt/15934.html>
- Thurstone, L. L. (1927). A law of comparative judgment. *Psychological review*, 34(4), 273.
- Verhavert, S., Bouwer, R., Donche, V., & Maeyer, S. D. (2019). A meta-analysis on the reliability of comparative judgement. *Assessment in Education: Principles, Policy & Practice*, 26(5), 541–562.
- Wheadon, C. (2019). *No More Marking [Computer Software]*. Recuperado de <https://www.nomoremarking.com/>

Apéndice: Estadísticas demográficas de los evaluadores

Nacionalidad	Recuento
Británica	8
Francesa	1
Italiana	1
Mexicana	2
Española	8
TOTAL	20

Género	Recuento
Hombre	7
Prefiero no responder	1
Mujer	12
TOTAL	20

Años en la enseñanza del español	Recuento
Más de 10 años	16
De 5 a 10 años	4
TOTAL	20

Conocimiento del MCER	Recuento
Conocimiento en profundidad	5
Conocimiento general	8
Lo conoce	7
TOTAL	20

Otras lenguas impartidas	Recuento*
Francés	16
Inglés	6
Alemán	2
Italiano	2
Latín	2
Griego antiguo	1
Japonés	1
Catalán	1

* Todos habían enseñado al menos otro idioma

Grupo(s) de edad de los estudiantes (de español)	Recuento
Adultos (más de 18 años)	20
Educación secundaria (último curso)/ preuniversitaria/bachillerato/formación profesional (15-19 años)	18
Educación secundaria (primeros cursos) (12-15 años)	18
Educación primaria (ciclo medio y superior) (9-12 años)	12
Educación primaria (ciclo inicial y medio) (6-9 años)	3
Educación infantil (3-5 años)	2

