

ANÁLISIS DEL LENGUAJE DE LAS ESCALAS
DE *EXPRESIÓN E INTERACCIÓN ESCRITAS*
EN LA CERTIFICACIÓN DE ELE/L2

*Language analysis of the scales of Written Expression
and Interaction in Spanish as second language certification*

MARTA GARCÍA

Universidad de Salamanca

martagg@usal.es

0000-0003-2609-6074

MARIAN VILLORIA

Cursos Internacionales. Universidad de Salamanca

marianvilloria@usal.es

RESUMEN

Los calificadores de *Expresión e interacción escritas* de pruebas de certificación utilizan las escalas como herramientas de medición, pero el uso de estas escalas no siempre resulta ser lo útil que se desearía para que el calificador sitúe a los candidatos en un nivel u otro porque, en su redacción, en muchas ocasiones, se abusa de elementos lingüísticos que crean ambigüedad y confusión. En este artículo, llevaremos a cabo un análisis de los textos de las escalas de tres exámenes de español, con el fin de llegar a conclusiones provechosas para un calificador.

Palabras clave: escalas; banda; calificadores; candidato; certificación lingüística.

ABSTRACT

Scales are used as measurement tools by raters of Written Expression and Interaction in certification tests, but the use of these scales is not always as useful as desired for the rater to place candidates in one level or another because, in their wording, linguistic elements

are often abused, creating ambiguity and confusion. In this article, we will carry out an analysis of the texts of the scales of three Spanish exams in order to reach useful conclusions for an assessor.

Key words: scales; band; rater; candidate; language assessment.

1. INTRODUCCIÓN

LA HABILIDAD DE ESCRIBIR Y REDACTAR en otra lengua ha acaparado un gran espacio en el mundo de la educación, pero no solo; en casi cualquier contexto se hace necesario comunicarse de forma escrita y hacerlo de forma efectiva. En el caso de la enseñanza-aprendizaje de lenguas, certificar este nivel de dominio exige responder, entre otras cosas, a una tarea de expresión escrita siguiendo unas directrices marcadas por unas instrucciones determinadas que ayudarán a completar la tarea de forma adecuada para reconocer el nivel alcanzado.

El proceso de crear una tarea para evaluar el dominio de la lengua puede resultar a simple vista sencillo; sin embargo, si queremos seguir los principios básicos de evaluación, fiabilidad y validez, sabemos que hay que llevar a cabo un trabajo complejo para que esas tareas nos permitan evaluar lo que realmente queremos evaluar.

Un creador de tareas debe estar formado en creación de ítems. Para ello, debe conocer perfectamente las especificaciones de la tarea, es decir, las características que debe presentar, además de dominar la información de los niveles que recoge el *Marco común europeo de referencia para las lenguas* (MCER 2001), con el fin de poder crear material del nivel o niveles deseados de forma adecuada. Para calificar esas tareas, una vez validadas e insertadas en el examen, necesitamos un elemento de medición, esto es, escalas que, creadas por expertos en evaluación específicamente para esas tareas, nos permitan otorgar una puntuación, situar en un nivel, o ambas cosas. El proceso de confección de escalas tampoco es fácil.

Así las cosas, nos proponemos reflexionar sobre la importancia de una redacción clara de las escalas y de la información que recogen para la calificación de las pruebas escritas, con el fin de extraer conclusiones que nos permitan evaluar mejor a candidatos que realizan exámenes de LE/L2.

2. LAS ESCALAS

La finalidad de las escalas es poner puntuaciones lo más objetivas posible. Además, suponen el referente que necesita todo evaluador y es el estándar común para interpretar los resultados (Cruz Trapero 2025). Con las escalas no se compara la actuación de candidatos entre sí, sino la de cada candidato con los criterios de la escala (Martínez Baztán 2011).

Atendiendo a la clasificación de las escalas por su contenido, distinguiríamos las siguientes:

A) Escalas analíticas, que se basan en puntos discretos, describen y definen la actuación del candidato fijándose en términos de competencia gramatical, léxica o pragmática, es decir, que discriminan entre rasgos de lengua. Así, por ejemplo, evalúan en torno a las categorías de coherencia, cohesión, corrección o alcance, como es el caso de los exámenes DELE. Dentro de las escalas analíticas hemos de distinguir entre escalas analíticas intuitivas y escalas analíticas diseñadas con datos empíricos e intuitivos (Knoch 2009) porque estas últimas se acercan más al análisis que pretendemos realizar.

Las escalas analíticas intuitivas provocan un halo que tiene efecto en la puntuación que les concedemos a otros criterios. Creemos que esto es así porque, en muchos casos, al confeccionar las escalas, se dedica más atención y detalle a un criterio que a otro. Las escalas son imperfectas, bien porque no contienen suficiente información (los descriptores deben ser redactados de forma breve y es imposible ser prolijo), o bien porque no es fácil describir todos los rasgos de lengua de los candidatos en cada nivel.

Las escalas analíticas diseñadas con datos empíricos, por ejemplo, con estudios de adquisición, acotan más y mejor la actuación de los calificadores (Knoch 2009).

B) Escalas holísticas, que interpretan el desempeño de una forma general sin puntuar un elemento en concreto. Podría decirse que se aplican usando la visión general del calificador y de lo que piensa que debe hacer un candidato en un determinado nivel, de ahí que puedan parecer más subjetivas y que suelen usarse para evaluar el uso general de la lengua y no un aspecto en concreto, como puede ser el caso de los exámenes con uso profesional (Martínez Baztán 2011). Este tipo de escala presenta también como ventaja una aceleración del proceso de corrección. Además, es menos costoso porque los calificadores califican más pruebas en menos tiempo (Weigle 2002).

La confección de las escalas debe tener en cuenta la finalidad del examen. Existe una teoría lingüística que la respalda y que aparece en el germen del examen, desde las especificaciones a la prueba en vivo, y se refleja en el constructo (Alderson *et al.* 1998).

La interpretación de la escala es muy diferente dependiendo de si lo hacemos de forma ascendente o descendente, es decir, si hay un ordenamiento de las puntuaciones de mayor a menor o al revés, de si hay puntos para pasar de un nivel a otro o de si realizamos una corrección analítica u holística. Dependiendo de la que apliquemos, llegaremos a puertos diferentes, si bien nuestro objetivo no es calificar pruebas de diferentes exámenes con las mismas escalas, sino analizar el lenguaje que se utiliza en unas y otras.

Sin embargo, unas y otras necesitan de entrenamiento y formación por parte de los calificadoros. Sin un buen conocimiento de los niveles y descriptores del *MCER*, se acentúan más las imprecisiones e imperfecciones de las escalas, pues como hemos dicho, el desempeño del candidato en una producción escrita no tiene por qué corresponderse con su uso en la vida real.

3. METODOLOGÍA

Nos detendremos en el análisis del lenguaje de tres escalas que se utilizan para evaluar la expresión escrita de exámenes de dominio de español: SIELE, DELE y USAL esPro, con el fin de comprobar si la redacción de los descriptores es lo suficientemente clara como para que el calificador pueda evaluar los textos de los candidatos. Estos tres exámenes representan diferentes tipos de exámenes y ofrecen escalas de calificación diferentes.

En primer lugar, analizaremos las escalas del diploma DELE, un examen de dominio en el que un candidato se presenta al nivel que desea certificar. Nos hemos decantado por el nivel B1 porque es el primero de los seis diplomas DELE que presenta en las escalas un desglose en cuatro categorías de la parte analítica de la escala: adecuación, coherencia, corrección y alcance. Tanto es así que la creación de la escala del nivel B1 constituye el andamiaje sobre el que se han ido creando las escalas de los exámenes DELE en los niveles superiores: B2, C1 y C2.

En segundo lugar, hemos decidido incluir también las escalas de calificación del examen progresivo SIELE (Servicio Internacional de Evaluación de la Lengua Española), en el que el candidato comienza respondiendo tareas de nivel A1 y, de forma progresiva, termina respondiendo tareas de nivel C1 y, finalmente, consigue situarse en uno de los cinco niveles.

Por último, estudiaremos la escala del examen adaptativo USAL esPro, examen con fines profesionales, que cuenta con un algoritmo que proporciona al candidato una tarea de nivel B1 al inicio del examen y, dependiendo de la respuesta, le proporciona una de mayor o menor dificultad hasta situarlo en el nivel que le corresponde.

Al tratarse de tres certificaciones diferentes, los criterios de las bandas y los descriptores de cada examen están orientados a evaluar lo que se recoge en las especificaciones de la tarea, es decir, las características que presenta cada una y que debe responder el candidato.

TABLA 1. Descripción de las tareas (<https://exámenes.cervantes.es/es/dele/exámenes/b1> y <https://exámenes.cervantes.es/es/siele/preparar-prueba> y <https://cursosinternacionales.usal.es/USALesPro/es/>)

DELE (Nivel B1)	SIELE	USAL esPro
Tarea 1 Correo electrónico. (100-120 palabras) Ámbito personal y público	Tarea 1 Correo electrónico. (100-120 palabras) Ámbito personal y/o público	Tarea 1 Correo electrónico. (50-60 palabras) Ámbito profesional y educativo
Tarea 2 Texto narrativo-descriptivo (130-150 palabras) Ámbito personal y público	Tarea 2 Artículo de opinión/carta a la redacción de un periódico (250-300 palabras) Ámbito público, educativo y profesional	Tarea 2 Informe/carta (180-200 palabras) Ámbito profesional y educativo

Como hemos mencionado anteriormente, las escalas pueden ser analíticas y holísticas. Así, en estos tres exámenes, tenemos de los dos tipos. En el caso del examen DELE y SIELE, podemos observar que hay dos escalas: la analítica, que profundiza en cuestiones lingüísticas, y, la holística, para la actuación general del candidato. Sin embargo, para el examen USAL esPro, se utiliza una escala holística porque, más que el uso de la lengua, interesa el desempeño del candidato en el ámbito profesional.

TABLA 2. Escalas analíticas DELE y SIELE (<https://exámenes.cervantes.es/es/dele/exámenes/b1> y [guia_examen_dele_b1_1.pdf](https://exámenes.cervantes.es/es/guia_examen_dele_b1_1.pdf) y <https://exámenes.cervantes.es/es/siele/preparar-prueba>)

EXAMEN	DELE (Nivel B1)	SIELE
Criterios de las bandas	Adecuación Coherencia Corrección Alcance	Cohesión Corrección Alcance
Número de bandas	4 bandas	5 bandas (una por cada nivel)

TABLA 3. Escalas holísticas DELE y SIELE (<https://exámenes.cervantes.es/es/dele/exámenes/b1> y [guía_examen_dele_b1_1.pdf](https://exámenes.cervantes.es/es/guia_examen_dele_b1_1.pdf) y <https://exámenes.cervantes.es/es/siele/preparar-prueba>)

EXAMEN	DELE (Nivel B1)	SIELE
Criterios de las bandas	Eficacia: comunicativa, ejecutiva y lingüística	Cumplimiento de la tarea
Número de bandas	4 bandas	5 bandas (una para cada nivel)

TABLA 4. Escala holística Usal esPro
(<https://cursosinternacionales.usal.es/USALesPro/es/>)

EXAMEN	Usal esPro
Criterios	Consecución de la tarea Recursos lingüísticos Organización del texto
Bandas	6 (una para cada nivel)

Con el propósito de analizar el lenguaje de las escalas, hemos realizado un estudio comparativo para observar elementos comunes o rasgos peculiares que pueden merecer un comentario.

En primer lugar, observamos que el empleo de la modalidad en las escalas provoca que el lenguaje pueda ser ambiguo en un contexto como el de la evaluación. En muchos casos, la subjetividad que apreciamos en la definición de las escalas viene determinada por este tipo de lenguaje. El amplio espectro de matices semánticos de los adjetivos y adverbios que aparecen en los descriptores de las bandas orientan al calificador (receptor) sobre cómo interpretar lo que en ellos se dice. Si Bally (1942) proponía que en toda oración debía reconocerse algún componente que responda al contenido de modalidad, y esta es una categoría lingüística transversal, es decir, que aglutina lo gramatical, la semántica y la pragmática (Bybee & Fleischman 1995), en las escalas tenemos una prueba fehaciente de que esto es así, de que las modalidades lógicas repercuten sobre el contenido de cada descriptor de las bandas y redundan en que los términos son poco concretos o aseverativos y,

por tanto, crean dudas en el calificador a la hora de moverse entre las bandas que configuran las escalas. Por ejemplo, las bandas se redactan, generalmente, para la calificación holística con la misma estructura. Esto es, núcleo del sintagma verbal (*escribe, aporta, redacta, se emplea, etc.*) con elementos lingüísticos (*adverbios, adjetivos*) que marcan la graduación de los diferentes niveles. Es aquí donde, debido a la modalidad, detectamos los posibles puntos de desencuentro de los calificadores a la hora de interpretar estos elementos utilizados para matizar. Adverbios como *sobradamente, parcialmente, adecuadamente, suficientemente, ocasionalmente*, o sintagmas preposicionales como *de forma sencilla, sin aparentes limitaciones*, sumergen al calificador en ese limbo de medición del que no siempre sale convencido. En las certificaciones que cuentan con doble corrección, se asegura la fiabilidad de la puntuación con la calificación del segundo corrector, pero no sucede así en el caso de exámenes en los que la puntuación es otorgada por un solo calificador, como es el caso del examen SIELE: por tanto, una escala poco precisa con un solo calificador debería evitarse en aras de una calificación justa, especialmente, en exámenes de alto impacto.

En segundo lugar, también hemos observado que la redacción de las escalas, esta vez en las analíticas, crea una predisposición a calificar las actuaciones de los candidatos desde la perspectiva de la penalización más que desde la valoración positiva de lo que estos han sido capaces de hacer. Decimos esto porque, tomando como ejemplo la escala analítica de calificación de las pruebas de expresión e interacción escritas en el DELE B1, concretamente, la categoría que corresponde a *Adecuación al género discursivo*, vemos que, en primer lugar, la extensión de la redacción de las distintas bandas de esta categoría (de 3 a 0) no es homogénea, pues, las bandas 3 y 2, que corresponden a la calificación de «APTO», son mucho más sintéticas que las bandas 1 y 0, que corresponden a la calificación «NO APTO». La banda 3 tiene 34 palabras, la banda 2 tiene 53, y las bandas 1 y 0 contienen 101 y 128 palabras respectivamente, es decir, se dedica más texto a las bandas en las que el candidato no pasa la prueba. Por tanto, la longitud en la descripción de las bandas no tiene que dar como resultado una calificación más fiable, pues esto depende más de la redacción y de la experiencia del calificador, tal y como iremos detallando.

Desde el punto de vista semántico, y siguiendo, por ejemplo, con la categoría de *Adecuación al género discursivo*, se observa también que, en la redacción de las bandas, aparece un número mayor de verbos y adjetivos negativos que positivos, como puede verse ilustrado en la siguiente tabla.

TABLA 5. Verbos y adjetivos de las escalas DELE Nivel B1
(guia_examen_dele_b1_1.pdf)

VERBOS		ADJETIVOS	
Positivo (+)	Negativo (-)	Positivo (+)	Negativo (-)
Amplía	Obliga a	Claros	Básicos/as
Emplea	Faltan	Precisos	Breves
Desarrolla	Menciona ¹	Adecuado	Desordenada
Utiliza	Se limita	Sencillos	Incompleta
Trata	Comete	Cotidiana	Personal
	Resulta		Elementales
	Adapta ²		Aisladas
	Se aleja		Incomprensible
	No se ajusta		

Teniendo en cuenta que esta categoría, la *adecuación*, es la primera de la escala analítica que tiene que manejar un calificador para evaluar una tarea de EIE del B1, consideramos que, redactada de esta forma, puede provocar un halo que afecta tanto a la calificación de la tarea como a la duda de los calificadores, aunque estos tengan experiencia incluso en la corrección de pruebas de EIE en estos exámenes. Sin embargo, nada de esto ocurre en las escalas de SIELE o Usal esPro en la categoría que mide lo mismo, ni en el resto de las categorías de la escala analítica del examen DELE, por tanto, podrán reformarse siguiendo el ejemplo de estas dos escalas.

Por otra parte, hemos encontrado que se toman cuestiones lingüísticas, la ortografía, por ejemplo, como parámetros para superar las bandas APTO/NO APTO que no quedan claras a la hora de ser interpretadas por los calificadores, puesto que no ayudan a discriminar el paso de la banda de NO APTO (Banda 0-1) a la banda de APTO (Banda 2). Así, en el caso del examen DELE, en el criterio *Corrección*, recogemos de la Banda 1 «errores ortográficos sistemáticos que en algunos casos dificulten la comprensión», cuando nos parece poco probable que un candidato que se presenta a un nivel de B1 no tenga fijada una ortografía medianamente aceptable, de ahí que esta descripción que se incluye para marcar el paso de una banda a otra resulte también un tanto vaga. En la Escala del examen SIELE se tiene en cuenta también la ortografía expresada de forma poco precisa: «la ortografía es razonablemente correcta» (Banda 4, equivalente al nivel B2. En el examen Usal es-

¹ El verbo *mencionar* que aparece en la banda 1 es negativo con respecto a *desarrollar*, que es el verbo que aparece en las bandas 2 y 3.

² «Comete errores de formulación que *adapta* al discurso». Sirva de ejemplo, la despedida de un correo electrónico de ámbito personal con la expresión poco adecuada «Gracias».

Pro no se hace alusión a cuestiones ortográficas porque, como venimos señalando, califica de forma más general las producciones.

Para evaluar la cohesión y coherencia de las producciones, vemos cuestiones muy dispares dependiendo de con qué escala y criterios trabajemos. En DELE se habla del criterio *coherencia*, en SIELE de *cohesión* y en USAL esPro de *organización del texto*. Centrándonos en elementos de cohesión como los conectores, en el examen DELE y SIELE se detallan los que debería usar un candidato del nivel, y es aquí donde el calificador puede verse de nuevo más desorientado, porque duda entre elegir una banda u otra si se ajusta literalmente a la escala. Por ejemplo, los conectores *muy sencillos* y *básicos* que se mencionan en la Banda 1 del examen DELE solo son tres «y, pero, porque», es decir, que la producción del candidato que contenga solo estos tres conectores no podría superar la Banda 1, aunque el texto sea breve y cohesionado, como se describe en la banda 2. Sin embargo, sabemos que el *Plan Curricular del Instituto Cervantes (PCIC)*, documento referencia para la creación de los exámenes DELE y SIELE, incluye muchos más para el nivel el B1, por lo tanto, recoger solo estos tres conectores en esa banda, no parece ajustarse a lo que realmente pueden hacer los candidatos de ese nivel (Santiago Guervós, 2020). Esta escala debería acercarse más a la del examen SIELE, pues es más abierta, ya que incluye gran parte de los mencionados en el *PCIC*. En el examen Usal esPro la escala holística se acerca más la descripción que hace el *MCER* (2001: § 4.4.1.2., 64) en el apartado de Expresión escrita en general, y lo traslada a la Banda del nivel B1 como *Texto sencillo y lógicamente secuenciado*.

Por último, y en lo que se refiere al reflejo de la corrección gramatical en las bandas de los diferentes exámenes que estamos estudiando, en el examen DELE B1, aparecen términos como *buen control gramatical*, *control razonable*, *estructuras gramaticales sencillas* y control *limitado*, que no permiten al calificador discriminar entre las bandas del APTO y NO APTO porque resultan poco específicas. En el examen SIELE, por ejemplo, para la banda del nivel B1, encontramos «control razonable de elementos lingüísticos sencillos» y detalla a qué se refiere el control razonable: distinción en el uso de *ser* y *estar*, uso de los pasados, perífrasis verbales más frecuentes, etc., lo cual no es aclaratorio para el calificador, pues dudará entre seguir literalmente lo que dice la banda o lo que dice la banda anterior.

En el examen USAL esPro, en la categoría Recursos lingüísticos, es decir, la equivalente a Corrección de DELE y SIELE, se lleva a cabo de una manera mucho más general y vaga, con el uso de adverbios y adjetivos abstractos: *cierta corrección*, *repertorio adecuado*, *algunas incorrecciones gramaticales*, etc. que, igualmente, solo ayudan si el calificador es experto.

4. CONCLUSIONES

Como mencionamos al principio de este artículo, nuestra intención ha sido llegar a conclusiones provechosas para ayudar a los calificadores de las pruebas de EIE de certificación lingüística.

Creemos necesario que los calificadores de este tipo de pruebas tengan un conocimiento profundo del *MCER* (2001), del tipo de tareas que configuran las pruebas de *Expresión e interacción escritas*, además de formación y experiencia como docentes de ELE. Del mismo modo, deben estar familiarizados con los Niveles de Referencia del *Plan Curricular del Instituto Cervantes* (1997-2025) en el caso del SIELE y DELE. No es necesario este último punto para el examen Usal esPro, pues se trata de un examen de fines profesionales.

En nuestra opinión, convendría que hubiera revisiones periódicas de las bandas que conforman las escalas tanto analíticas como holísticas de las pruebas de EIE por parte de expertos en evaluación. El objetivo de estas revisiones sería evitar que se produzcan ambigüedades, vacíos, limitaciones, descripciones por exceso o por defecto que confundan a los calificadores y los desvíen del principal propósito, que no es otro que otorgar una calificación lo más objetiva posible a cada candidato, calificación que se vería reforzada siempre con la doble corrección.

Asimismo, con el fin de mejorar la redacción de las escalas, se podrían elaborar encuestas semidirigidas que se pasarían a los calificadores para recoger datos sobre los puntos más dudosos que encuentran en las descripciones de las escalas y que entorpecen la calificación de las pruebas. Con este trabajo de campo, los datos recogidos ayudarían a los responsables de certificación lingüística a mejorar, concretar y definir con mayor precisión la redacción de las diferentes bandas.

REFERENCIAS BIBLIOGRÁFICAS

- ALDERSON, Charles & CLAPHAM, Caroline & WALL, Dianne (1998): *Exámenes de idiomas*, Madrid, Cambridge University Press.
- BALLY, Charles (1942): «Syntaxe de la modalité explicite», en *Cahiers de Ferdinand de Saussure* 2, 3-13.
- BYBEE, Joan & FLEISCHMAN, Suzanne (1995): *Modality in Grammar and Discourse*, Amsterdam/Philadelphia, John Benjamins Publishing Company.
- CONSEJO DE EUROPA (2001): *Marco común europeo de referencia para las lenguas: aprendizaje, enseñanza, evaluación* [en línea], https://cvc.cervantes.es/ensenanza/biblioteca_ele/marco
- CRUZ TRAPERO, Joaquín Manuel (2025): *Diseño y validación de exámenes de dominio de lengua*, MEFPD, Dirección General de Evaluación y Cooperación Territorial.

- Exámenes DELE, SIELE y USAL esPro* (s.f.): <https://exámenes.cervantes.es/es/dele/exámenes/b1>, <https://exámenes.cervantes.es/es/siele/preparar-prueba> y <https://cursosinternacionales.usal.es/USALesPro/es/>
- KNOCH, Ute (2009): *Diagnostic assessment of writing: A comparison of two rating scales*, en *Language Testing* 26(2), 275-304. <https://doi.org/10.1177/0265532208101008>
- MARTÍNEZ BAZTÁN, Alfonso (2011): *La evaluación de las lenguas: garantías y limitaciones*, Granada, Octaedro/Mágica.
- INSTITUTO CERVANTES: véase *PCIC* (1997-2025).
- MCER: véase Consejo de Europa (2001).
- PCIC: *Plan curricular del Instituto Cervantes: Niveles de referencia para el español (1997-2025)*, Madrid, IC. [en línea]. https://cvc.cervantes.es/ensenanza/biblioteca_ele/plan_curricular/
- SANTIAGO-GUERVÓS, Javier (2020): «El uso de los marcadores discursivos en español LE/L2: estudio de un corpus de aprendices», en Javier Santiago-Guervós (coord.), *Marcadores del discurso y enseñanza del español como lengua extranjera*. *Archiletras Científica* IV, 23-32.
- WEIGLE, Sarah (2002): *Assessing Writing*, Cambridge, Cambridge University Press. <https://doi.org/10.1017/CBO9780511732997>

