

Dificultades en el uso de los descriptores en el proceso de evaluación de la expresión oral de ELE

Francisco del Moral Manzanares

Universidad de Verona.

francisco.delmoral@univr.it

del Moral Manzanares, F. (). Dificultades en el uso de los descriptores en el proceso de evaluación de la expresión oral de ELE. *Revista Nebrija de Lingüística Aplicada* () 16.

RESUMEN

El presente artículo se enmarca en un proyecto de investigación en curso que tiene como objetivo principal la adopción de una escala de descriptores para la calificación de la prueba oral de ELE de nivel C2 de la Universidad de Verona. Para ello se ha introducido, además del uso de una escala de descriptores, el de las fichas de calificación, que permite a los evaluadores anotar ejemplos representativos de la actuación de cada candidato, así como diferentes valoraciones de la misma basadas en la escala utilizada. Esta segunda herramienta sirve de ayuda a la hora de decidir la calificación y permite obtener una mayor transparencia del proceso. Sin embargo, al hacer las anotaciones, los evaluadores no siempre han reflejado fielmente los descriptores de la escala, sino que en muchas ocasiones los han adaptado, modificándolos en mayor o menor medida, lo cual puede suponer un problema para la fiabilidad de la evaluación.

En las siguientes líneas se presenta una clasificación de lo que se ha denominado “modificaciones de los descriptores” que los evaluadores han introducido en sus anotaciones y, a partir de ellas, se intenta sentar las bases para mejorar el proceso en sus etapas posteriores, concretamente en lo que se refiere a las características de la escala de descriptores, las de las fichas de calificación y las de la formación de los evaluadores.

Palabras clave: evaluación de la lengua oral, escalas de descriptores, fichas de calificación, formación de evaluadores, ELE

ABSTRACT

The present study is part of an ongoing research project focused on the adoption of a scale for assessing speaking performance in a sample of native Italian students who are learning Spanish as a foreign language. Scoresheets were also used in order for examiners to take notes of candidates' representative speech samples. This second tool contributed to the participants' final speaking evaluation and to a greater transparency of the evaluation process. However, while making their notes, the raters do not always reflect accurately the scale's descriptors, which can lead to reliability issues of the evaluation process.

In this paper a classification of these “scale's descriptors modifications” is presented. Moreover, an attempt was made to introduce some improvements regarding the evaluation process and in particular the characteristics of the scale, the scoresheets and the raters' training.

Keywords: Oral Language Assessment, Rating Scales, Scoresheets Rater Training, Spanish as a Foreign Language

1. INTRODUCCIÓN

Este trabajo se plantea poner de relieve una serie de dificultades detectadas durante un proyecto centrado en la evaluación oral de ELE, relacionadas con el uso de las dos herramientas utilizadas a tal fin: la escala de descriptores y las fichas de calificación(1). Más concretamente, durante la prueba de expresión oral, los examinadores aplicaban una escala de descriptores basada en las de los DELE y adaptada al propio contexto educativo y, al mismo tiempo, iban realizando anotaciones en unas fichas de calificación concebidas para dar mayor transparencia al proceso, y que habían de contener tanto citas textuales de las actuaciones de los candidatos como fragmentos de los descriptores que reflejaran el grado de pertinencia alcanzado por los mismos. Pues bien, es en este proceso de transferencia desde la escala hasta las fichas de

calificación donde, en numerosas ocasiones, se produce alguna transformación de los descriptores (filtrados, probablemente, por la mente del evaluador y, las necesidades de celeridad propias del momento). Estos cambios, que quedan registrados sobre el papel de cada una de las fichas mencionadas y han sido denominados *modificaciones de los descriptores*, pueden suponer un riesgo para la fiabilidad de la evaluación, pues atentan de alguna manera contra la función principal de una escala de descriptores: dar una referencia invariable y clara de lo que significa alcanzar cada uno de los niveles de dominio lingüístico establecidos.

Las siguientes líneas tienen como principal objetivo presentar una clasificación de dichas modificaciones de los descriptores, de modo que en futuras fases del proyecto se pueda utilizar una versión mejorada de la escala, así como de las fichas de calificación, y, con ello, minimizar en la medida de lo posible los riesgos para la fiabilidad que pudiera derivarse de ellas.

1.1. CONTEXTO Y ESTRUCTURA DE LA INVESTIGACIÓN

Como ya se ha dicho, el proyecto se está llevando a cabo en el Centro Lingüístico de la Universidad de Verona. Se basa en las pruebas orales de nivel C2 grabadas durante la sesión de julio de 2012. Los candidatos que realizaron la prueba fueron 37, de los que se seleccionaron 18 por “muestreo aleatorio estratificado” (León y Montero, 1997:101). Los evaluadores implicados son dos: la profesora oficial del curso y el propio investigador.

En su primera fase, ya finalizada, se trató de averiguar si la utilización de una escala de descriptores para la evaluación de la prueba de expresión e interacción orales de nivel C2 permitiría obtener mayor fiabilidad en la evaluación, en comparación con una evaluación holística de la actuación de los candidatos sin más herramienta que una lista en la que figuran los datos personales de estos y el espacio destinado a la calificación de la prueba oral. Los resultados permitieron concluir que la fiabilidad aumenta cuando se usa la escala de descriptores propuesta, y que el aumento de la fiabilidad externa (interevaluadora, entre evaluadores) es mayor que el de la interna (intraevaluadora, de un evaluador consigo mismo) (2).

En un primer momento, los evaluadores (llamados E1 y E2) calificaron cada prueba oral dos veces por medio de una valoración holística sin utilizar escalas de descriptores ni fichas de calificación (que se denominaron calificación 1 –C1- y calificación 2 –C2-), y en un segundo momento, realizaron otras dos calificaciones, pero esta vez con ayuda de ambas herramientas (se trata de la calificación 3 –C3- y calificación 4 –C4-). Podemos resumir el proceso de la primera fase con el siguiente esquema:

C1E1 C2E1 C1E2 C2E2	C3E1 C4E1 C3E2 C4E2
Hoja de datos del candidato y nota holística de la prueba oral	Escala de descriptores y fichas de calificación

Tabla 1. Esquema general de la primera fase de la investigación

Además de las conclusiones que se refieren a la influencia del uso de la escala en la fiabilidad de la prueba, la primera fase de nuestro proyecto determinó que, para las etapas sucesivas, podrían extraerse informaciones muy útiles del análisis de los datos del diario de investigación y las fichas de calificación.

Durante la segunda fase, actualmente en curso, se utiliza toda la información procedente tanto del diario de investigación como de las fichas de calificación para crear una segunda versión de las herramientas de evaluación (escala y fichas). Es en esta etapa del proyecto donde se detectan las modificaciones de los descriptores a que hace alusión el presente trabajo. Una vez que ya se haya obtenido una versión mejorada de las fichas de calificación y la escala de descriptores, se someterán al juicio de un comité de expertos.

En la tercera fase del proyecto, se comprobará la eficacia de la versión definitiva de las herramientas con nuevas sesiones de evaluación, tras las que se volverá a medir la fiabilidad. Los índices resultantes se compararán con los obtenidos en la primera fase, de modo que se pueda establecer cómo ha influido la validación de la escala en la fiabilidad de la prueba.

1.2. FIABILIDAD Y ESCALAS DE DESCRIPTORES

El objetivo principal de nuestra investigación es mejorar la fiabilidad de la prueba oral. Se entiende por fiabilidad la “consistencia en la evaluación” (Bordón, 2006:62), o la “estabilidad (...) en la obtención de resultados” (*Diccionario clave de términos de ELE*), es decir, la capacidad de una prueba determinada para arrojar los mismos resultados independientemente del contexto. Cuando se trata de pruebas de corrección subjetiva (las de expresión oral o escrita), se suele diferenciar entre *fiabilidad interna* (intraevaluadora o de un mismo evaluador), para referirse a la correspondencia entre los resultados de dos evaluaciones de la misma prueba realizadas por un solo evaluador en momentos diferentes, y *fiabilidad externa* (interevaluadora o entre evaluadores), para aludir a la correspondencia entre la evaluación otorgada a una misma prueba por dos evaluadores diferentes (cfr. Alderson, Clapham y Wall, 1998:128).

Otro concepto básico en nuestra investigación es el *dedescriptor*. Por tal se entiende cada una de las explicaciones que se añaden a las diferentes puntuaciones (numéricas o verbales) para aclarar el significado de un determinado juicio. La serie completa y ordenada de todas ellas constituye una *escala de descriptores* (cfr. Luoma, 2004:59).

Nuestra escala de descriptores está basada en la de los DELE y consta de cinco criterios (coherencia, fluidez, capacidad crítica, corrección y alcance) y de cinco bandas de nivel (50, 65, 75, 85, 95), determinadas por, entre otros factores, nuestra necesidad de dar una puntuación sobre cien.

La escala de descriptores es una herramienta fundamental para garantizar un mínimo de fiabilidad en la evaluación de pruebas de corrección subjetiva. A pesar de que los expertos apuntan que la creación de las escalas habría de realizarse simultáneamente a la de las pruebas para cuya evaluación se van a utilizar (cfr. Luoma, 2004), en nuestro caso se introdujeron cuando la prueba ya estaba implantada desde hacía años e, igualmente, la influencia en su fiabilidad fue positiva.

2. CARACTERÍSTICAS Y USO DE LAS FICHAS DE CALIFICACIÓN

Las fichas de calificación se han concebido como una herramienta destinada a facilitar la tarea del evaluador de expresar una valoración justa de la actuación del candidato que tuviera en cuenta sus rasgos fundamentales y los cinco criterios utilizados a tal fin. Por ello se diseñó como una tabla en la que las columnas se correspondían con los criterios y las filas horizontales, con cada una de las bandas de nivel. En los espacios resultantes, el evaluador había de escribir todo aquello que le resultara útil para alcanzar su objetivo (v. fig. 2).

	<u>Coherencia</u>		<u>Fluidez</u>		<u>Cap. crítica</u>	<u>Corrección</u>	<u>Alcance</u>
95							
85							
75							
65							
50							

Fig. 1. Ficha de calificación

La fig. 2 muestra un ejemplo de ficha de calificación completa. En ningún caso los evaluadores añadieron la fecha de la prueba ni se identificaron en el espacio reservado para ello. Sin embargo, se añadió un código en la parte superior izquierda que identifica tanto el orden de la calificación como al evaluador (en este caso, se trata de la calificación 4 del evaluador 2, C4E2). En el espacio destinado a los cinco criterios aparecen sendas puntuaciones encerradas en un círculo. La media aritmética de todas ellas constituye la calificación final, añadida en el espacio correspondiente, debajo de la tabla (en este caso, 74). Como se puede observar, los espacios destinados a las anotaciones relativas a la coherencia y la fluidez aparecen, a su vez, divididos en dos. Están destinados a diferenciar entre el monólogo y el diálogo (partes de la prueba), en el caso de la coherencia, y entre la pronunciación y la fluidez propiamente dicha, en el de la fluidez.

C4E2

Fichas de calificación
Prueba oral-CZLM

Nombre: _____
Número de matrícula: 362608
Fecha: _____

85 75 75 85 70

95	Coherencia		Fluidez		Cap. Crítica	Corrección	Alcance
	10.0	5.0	10.0	5.0			
85	buen estruct. pero poca de 85	85	buen fluidez 75	75	profundidad 85 capacidad de recurrir a quien es más el poder		Atento a los cambios de pronunciación a suelta espaldas
75		80	75				
65			algun errores placidos 75/85			65	
50			BA malta respon			prácticamente en dicho tema concen- tando en un de si que ocurre (una condición)	pregunta/respuesta lo media los publicista aprovechando habilitando condiciones

Nota media: 74
Calificador: _____

En todos en
situación de
poco
buena

Fig. 2. Ficha de calificación completa

En principio se indicó que los apuntes susceptibles de ser incluidos en las fichas (que se han llamado *anotaciones*) podían ser tanto ejemplos concretos de la actuación del candidato como fragmentos de los descriptores que reflejaran el modo en que este se expresaba. Precisamente en este segundo tipo se centra el presente trabajo, ya que se ha constatado que, al realizar las anotaciones que deberían reflejar el contenido de los descriptores, los evaluadores introducen una serie de cambios de diversa índole (denominados *modificaciones de los descriptores*). Estos pueden responder a una posible dificultad meramente mecánica surgida en el momento de escribir, mientras se escucha al candidato, pero también a la necesidad de adecuar un descriptor determinado a la calidad del discurso que se está calificando, o a un simple descuido, entre otros motivos.

3. MODIFICACIÓN DE LOS DESCRIPTORES: DEFINICIÓN Y EJEMPLOS

En el análisis de las irregularidades(3) cometidas por los examinadores, se han establecido dos grupos: las que suponen una irregularidad en la aplicación de la escala (divididas a su vez en irregularidades de valoración e irregularidades de procedimiento) y, por otra parte, las causadas por el contenido de los descriptores. En el primer grupo, se consideran de valoración las puntuaciones incorrectas, los errores en la asignación de banda y los errores en la asignación de criterio, mientras que se consideran de procedimiento las anotaciones ambiguas, poco claras o ininteligibles, las anotaciones realizadas entre dos bandas y las modificaciones de los descriptores (4).

Se produce una modificación de los descriptores cuando un evaluador realiza una anotación en las fichas sin atenerse literalmente a lo que estos dicen. En el momento de escribir una anotación, los evaluadores deberían reflejar los descriptores con la mayor fidelidad posible. La importancia de esta característica del procedimiento reside en que el evaluador ha de comparar continuamente lo que dice el descriptor con la actuación de los candidatos, pues el hecho de que esta referencia permanezca invariable durante todo el proceso es lo que va a permitir que la fiabilidad de la prueba sea alta. Si el evaluador introduce en sus anotaciones cambios de mayor o menor entidad respecto a las expresiones contenidas en los descriptores, existe el peligro de que, conforme va examinando, se altere progresivamente su referencia de lo que debería ser la actuación correcta de los candidatos. Es decir, en su mente se produciría una progresiva transformación de la idea de “actuación correcta”, que partiría de lo expresado por los descriptores e iría pareciéndose cada vez más a lo que ha ido anotando en las fichas, sobre todo si ha optado por el mismo cambio en más de una ocasión.

Si bien hemos partido de estas consideraciones al clasificar las modificaciones de descriptores encontradas en las fichas de calificación, no todas ellas han sido consideradas error del evaluador (v. nota 3). Concretamente, dentro de este grupo, hemos establecido una diferenciación entre tres tipos de modificaciones: las hechas por medio de mínimas variaciones, por expresiones equivalentes y por expresiones de significado distante al del descriptor.

3.1. VARIACIONES MÍNIMAS, NO SIGNIFICATIVAS

Se trata de uno de estos casos.

- la mayor parte de la anotación es literal.
- el evaluador añade, suprime o sustituye una palabra o una expresión sin que varíe significativamente el sentido del descriptor original.

Por ejemplo:

Anotación (5) 1, C4E2, alcance 65:	Descriptor Alcance 65
“a veces presenta dificultad para extenderse”	“presenta dificultad para extenderse”

Anotación 4, C4E1, fluidez 75:	Descriptor Fluidez 75
“temas difíciles obstaculizan ritmo”	“Los temas conceptualmente difíciles pueden obstaculizar la fluidez natural de su expresión”

Tablas. 2 y 3. Ejemplos de anotaciones con variaciones mínimas no significativas(5)

En nuestro análisis, estos casos no han sido considerados *irregularidades*.

3.2. EXPRESIONES EQUIVALENTES (IMPRECISIONES) (?)

Son los casos en que el significado de la anotación está próximo al del descriptor, pero existe entre ellos una diferencia suficiente para que la fiabilidad de la evaluación pueda resultar perjudicada. En este sentido, la “equivalencia” ha de entenderse más como una distancia que dificulta la identificación plena entre anotación y descriptor, y no tanto como una analogía completa. Se trata de las anotaciones en que el evaluador:

- Añade un adjetivo valorativo (*bueno, malo, aceptable*) o sustituye con ellos otros presentes en el descriptor. La interpretación de estos adjetivos depende mucho de la subjetividad de cada evaluador.

- Utiliza sinónimos u otros términos de significado similar a los originales. Por ejemplo:

?

Anotación	Descriptor
1, C3E1, alcance 95:	Alcance 95
"léxico muy específico"	"dispone de un repertorio muy rico, preciso y variado"

Tabla 4. Ejemplo de anotación con expresión equivalente (imprecisión)

- Hace referencia a lo que hemos llamado "conceptos parciales" incluidos en los que aparecen en los descriptores. Por ejemplo, alusiones a una parte del discurso para referirse a un discurso bien estructurado:

?

Anotación	Descriptor
Varias	Coherencia 95
"introducción" "guión" "premisa" "tesis"	"discurso propio bien estructurado"

Tabla 5. Anotaciones con expresiones equivalentes (imprecisiones)

Estas expresiones pueden llegar a ser perjudiciales para la fiabilidad, sobre todo si se usan continuamente, por eso han sido consideradas imprecisiones.

3.3. EXPRESIONES DE SIGNIFICADO DISTANTE AL DEL DESCRIPTOR (ERRORES) (?)

- Anotaciones que expresan algo contrario a lo que dice el descriptor. Por ejemplo:

?

Anotación	Descriptor
5, C3E1, fluidez 75	Fluidez 75
"no varía entonación"	"con variaciones de entonación"

Tabla 6. Anotación de significado distante al del descriptor (error)

- Que hacen referencia a aspectos no contemplados en este. Por ejemplo:

?

Anotación 10, C4E1, coherencia 95	Descriptor Coherencia 95
"justifica la elección del tema"	No hace referencia a la elección del tema, sino el uso de las fuentes, la estructura, extensión y registro del discurso, la naturalidad y destreza en la interacción y la habilidad en los turnos de palabra.

Tabla 7. Anotación de significado distante al del descriptor (error)

- Guardan más relación con el descriptor de otra banda o de otro criterio.
- No guardan relación con el descriptor correspondiente, o bien esta es remota, o no se percibe con claridad. Por ejemplo:

?

Anotación Varios E2, coherencia 85	Descriptor Coherencia 85
"estructura no marcada" "estructura no muy marcada" "estructura marcada parcialmente"	"Extrae información adecuada de las diversas fuentes y la presenta con una estructura clara y ordenada, aunque a veces muestre cierta dependencia de las fuentes".

Tabla 8. Anotaciones de significado distante al del descriptor (errores)

En este caso el problema es que el significado *demarcada* no está claro y no ha sido consensuado previamente.

Otro caso que incluimos en este apartado es el de la anotación "pregunta-respuesta" (o "pregunta / respuesta"), que no se corresponde con los descriptores de los criterios y bandas en que se recoge (v. tabla 9). De hecho, a pesar de que haya sido siempre utilizada por el evaluador 2, su significado es difícil de desentrañar si tenemos en cuenta los diferentes criterios y bandas en que aparece. En la primera de ellas habría que considerarla reflejo de una actuación positiva indicadora de una "estructura clara y ordenada", mientras que en las dos últimas se correspondería con una carencia. Se trata de una prueba evidente de que salirse de lo que marcan los descriptores pueden ser perjudicial para la claridad y la fiabilidad de la evaluación.

Anotación 10, C4E2, coherencia 85	Descriptor Coherencia 85
"pregunta-respuesta"	"Extrae información adecuada de las diversas fuentes y la presenta con una estructura clara y ordenada".
Anotación 11, C3E2, alcance 65	Descriptor Alcance 65
"pregunta / respuesta"	"presenta dificultades para extenderse y suele responder con frases muy cortas o incluso monosílabos"
Anotación 11, C4E2, alcance 50	Descriptor Alcance 50
"pregunta / respuesta"	No hace alusión al tipo de respuestas

Tabla 9. Anotaciones de significado distante al de los descriptores (errores)

3.4. MATIZACIONES

Un caso especial lo constituyen aquellas modificaciones de los descriptores que hemos llamado *matizaciones*. En estos casos, el evaluador percibe que la actuación del candidato no se corresponde exactamente con un determinado descriptor, por lo que en su anotación añade una pequeña variación para redondear hacia arriba o hacia abajo el grado de dominio alcanzado según un determinado criterio. Los ejemplos de matizaciones han sido, a efectos estadísticos, añadidos a dos de los grupos anteriores, a saber:

- Imprecisiones (?): si la decisión del evaluador se considera apropiada.
- Errores (?): si se considera que la decisión del evaluador se basa en una decisión equivocada.

Veámoslo detenidamente, sirviéndonos de una agrupación diferenciada por criterios:

3.4.1. COHERENCIA

Anotación 1, C4E2, coherencia 85	Descriptor Coherencia 95
"hilo se sigue bastante bien" (i)	"discurso propio y fácil de seguir".
	Descriptor Coherencia 85
	No hace referencia al concepto fácil / difícil de seguir

Tabla 10. Matización a la baja (imprecisión)

Se observa la intención del evaluador de matizar a la baja esa idea presente en 95 por medio del adverbio *bastante* y, la decisión consecuente de realizar la anotación en la banda inferior.

Anotación 2, C4E1, coherencia 75	Descriptor Coherencia 85
"no extrae mucha <u>info</u> fuentes" (ε)	"extrae información adecuada de las diversas fuentes"
	Descriptor Coherencia 75
	No hace referencia al uso de las fuentes

Tabla 11. Matización a la baja (error)

Se observa la intención de matizar a la baja, como en el caso anterior. Se considera error (?) puesto que el descriptor de 85 no hace alusión a la cantidad de información que el candidato extrae de las fuentes, por tanto no da a entender que esta tenga que ser "mucho". Es decir, una actuación que se corresponda con la anotación realizada por el evaluador ("no extrae mucha info fuentes") se corresponde igualmente con el descriptor de 85.

Anotación 8, C3E1, coherencia 85 10, C3E1, coherencia 85	Descriptor Coherencia 95
"no varía el curso de la interacción" (ε)	"Muestra la capacidad, si la situación lo requiere, de tomar iniciativas para variar el curso de la interacción"
Anotación 11, C3E2, coherencia 85	
"bien diálogo sin tomar iniciativas" (ε)	
Anotación 14, C4E1, coherencia 85	
"No toma iniciativas, pero referencia y alusiones a <u>interl</u> natural" (ε)	

Tabla 12. Matizaciones a la baja (errores)

Se observa que el evaluador piensa que el hecho de que el candidato no tome la iniciativa en ningún momento, impide que su actuación sea considerada de nivel 95. Sin embargo, del descriptor 95 se desprende que esa capacidad no es una condición indispensable para alcanzar dicho nivel, pues solo habrá de demostrarse "si la situación lo requiere". Por lo tanto, se consideran errores (?).

Anotación 13, C4E1, coherencia 75	Descriptor Coherencia 75
"Uso controlado de mecanismos de cohesión (algo limitado)" (I)	"uso controlado de (...) mecanismos de cohesión"
	Descriptor Coherencia 65
	"Utiliza un número limitado de mecanismos de cohesión"

Tabla 13. Matización a la baja (imprecisión)

Se observa un intento de matización a la baja, obteniendo como resultado una formulación que no se corresponde completamente con ninguno de los dos niveles (ni con 75 ni con 65). En este caso, el evaluador ha optado por situarlo en la banda superior. Se considera imprecisión (?), pues el cambio introducido en la anotación nos permite considerarlo expresión equivalente, y no distante, al descriptor original de la misma banda.

3.4.2. FLUIDEZ

Anotación 5, C4E1, fluidez 85	Descriptor Fluidez 85
"ritmo uniforme" (ε)	"Se expresa espontánea y detalladamente con fluidez natural y coloquial"
Anotación 2, C4E1, fluidez 75 12, C4E1, fluidez 75	Descriptor Fluidez 75
"ritmo uniforme" (I) "ritmo uniforme pero con pausas y vacilaciones" (I)	"Se expresa con fluidez y espontaneidad sin apenas esfuerzo"
Anotación 15, C3E1, fluidez 65	Descriptor Fluidez 65
"ritmo uniforme"(ε)	"ritmo bastante uniforme, pero hace algunas pausas largas y puede dudar mientras busca estructuras"

Tabla 14. Varias matizaciones del criterio de fluidez.

En el caso de las anotaciones realizadas en la banda 75 (2, C4E1 y 12, C4E1), se observa un intento de matización al alza por medio de la supresión del adverbio *bastante* del descriptor de 65. La anotación resultante ("ritmo uniforme") se puede considerar equivalente al descriptor 75, por lo que consideramos imprecisiones (?) ambos ejemplos. En el caso de 12, C4E1 existe una segunda matización al alza por medio de la supresión del adjetivo *largas* del descriptor de 65 ("con pausas y vacilaciones" en lugar de "hace algunas pausas largas y puede dudar mientras busca estructuras").

Toda vez que hemos considerado la anotación "ritmo uniforme" equivalente al nivel 75, señalamos como error (?) los casos en que se recoge la misma anotación en otra banda (5, C4E1 y 15, C3E1).

3.4.3. CAPACIDAD CRÍTICA

Anotación 11, C3E1, fluidez 85	Descriptor Capacidad crítica 95
"Hace aportaciones nuevas" (ε)	"demostrando capacidad de escucha y elaboración de las opiniones del interlocutor, siendo flexible y coherente"
	Descriptor Capacidad crítica 85
	"Sus aportaciones van siempre encaminadas a sostener su idea inicial"
	Descriptor Capacidad crítica 75
	"No hace aportaciones nuevas"

Tabla 15. Matización al alza (clasificación subjetiva).

Se observa una corrección al alza del descriptor de 75 y la colocación de la correspondiente anotación en el nivel inmediatamente superior (85). Sin embargo, leyendo atentamente el descriptor de 85 observamos que este se correspondería no solo con una actuación en que se hacen aportaciones nuevas, sino también con aportaciones que van siempre encaminadas a sostener la idea inicial, lo cual quiere decir que la anotación "hace anotaciones nuevas" podría corresponderse también con el descriptor 95, según el cual las aportaciones personales no van siempre encaminadas a sostener la idea inicial, sino que el candidato muestra "capacidad de escucha y elaboración de las opiniones del interlocutor, siendo flexible y coherente". La consideramos, pues, una matización equivocada (error).

Anotación 11, C3E2, fluidez 95	Descriptor Capacidad crítica 95
"No hace "derroche" de elaboración del interlocutor" (puntuación 90) (ε)	"demostrando capacidad de escucha y elaboración de las opiniones del interlocutor, siendo flexible y coherente"

Tabla 16. Matización a la baja (error).

Se observa una matización a la baja de la capacidad a la que se refiere el descriptor y el consecuente redondeo de la calificación, de 95 a 90. Lo consideramos una matización equivocada porque el descriptor de referencia no incluye consideraciones sobre el grado de elaboración de las opiniones del interlocutor. Por lo tanto, a efectos de nuestra clasificación es un error (?).

3.4.4. CORRECCIÓN

Anotación 2, C4E1, corrección 75 "No arriesga" (ε)	Descriptor Corrección 85
Anotación 4, C4E1, corrección 75 "Buen control pero no riesgos" (ε)	"Los errores (...) Pueden deberse a la voluntad de asumir riesgos"
Anotación 11, C4E1, corrección 75 "pocos riesgos" (ε)	
Anotación 15, C4E1, 65-50 "pocos riesgos para evitar errores" (ε)	No hace referencia a la voluntad de asumir riesgos

Tabla 17. Matizaciones a la baja (errores).

En los casos anteriores, el evaluador ha tomado como una condición indispensable para alcanzar el nivel 85 la voluntad por parte del candidato de asumir riesgos, es decir, de intentar construir enunciados con estructuras de nivel alto, a pesar de que supone el peligro de una mayor comisión de errores. Sin embargo, de la lectura del descriptor 85 se desprende que los errores cometidos por el candidato situado en este nivel de corrección "pueden deberse" a esta razón, pero no necesariamente, por lo que la ausencia de este rasgo no justifica por sí solo que una actuación determinada no sea incluida en la banda 85.

Así pues, estos tres casos se consideran error (?). En 4, C4E1 se da, además, la circunstancia de que "buen control" se corresponde con el descriptor 65, frente a "alto grado de corrección gramatical", propio de la banda 75. En 15, C4E1 la matización a la baja es, asimismo, excesiva, pues se recoge en 65-50, y la anotación se hace en dos bandas consecutivas, lo que hemos considerado, en principio, por sí solo, irregularidad de procedimiento(6).

Anotación 18, C4E2, corrección 75-65 "repetidamente me parece excesivo" (i) (puntuación 70)	Descriptor Corrección 75 "Los errores son escasos"
Anotación 11, C3E2, corrección 65 "no llega a cometer errores repetidamente" (i) (puntuación 70)	Descriptor Corrección 65 "comete repetidamente errores y deslices"

Tabla 18. Matizaciones consideradas lógicas (imprecisiones).

En la tabla 18 se muestran dos intentos de ponderación de la cantidad de errores. Se expresa explícitamente que la actuación del candidato no se corresponde con una comisión de errores repetidos y, al mismo tiempo, de ambas anotaciones se deduce que el evaluador tampoco percibe que los errores sean escasos. Parece, pues, justificada la matización (en la que se otorga una puntuación intermedia ¿70?). A pesar de que la anotación se ha recogido entre dos bandas en 18, C4E2 (lo que señalamos normalmente como una irregularidad de procedimiento ¿v. nota 5?), la consideramos en este caso una imprecisión (?).

3.4.5. ALCANCE

Anotación 1, C4E1, alcance 85	Descriptor Alcance 95
"No consciencia" (ι)	"mostrando consciencia de las implicaciones de carácter sociolingüístico y sociocultural"
Anotación 3, C4E1, alcance 85	Descriptor Alcance 85
"poca consciencia?" (ε)	"comete repetidamente errores y deslices"

Tabla 19. Matizaciones del criterio de alcance.

En este caso, consideramos que responde a una deducción apropiada clasificar en 85 aquellas actuaciones que no cumplen el requisito expresado en el descriptor 95, por lo que la anotación de 1, C4E1 se señala como imprecisión. Por otra parte, la anotación de 3, C4E1, añade además de una gradación intermedia entre el descriptor ("mostrando consciencia") y la anotación de 1, C4E1, ("no consciencia"), un signo de interrogación, que le resta claridad. De ahí que la marquemos como error (?) (v. nota 4).

4. RESULTADOS

En relación con la frecuencia en que los evaluadores han realizado una modificación de los descriptores en sus anotaciones, la fig. 3 nos muestra una comparación entre estas y las anotaciones generales, entendiendo por tales todas aquellas que no constituyen una *puntuación de criterio* (7). El 88% de "otras" incluye tanto anotaciones generales apropiadas, como aquellas que incurren en alguna irregularidad diferente a las modificaciones de los descriptores (v. apartado 3).



Fig. 3. Comparación entre el porcentaje de modificaciones de los descriptores y anotaciones generales.

Por lo que se refiere a los diferentes tipos de modificación de los descriptores, la tabla 20 nos muestra la distribución de todas ellas en las categorías anteriormente establecidas. Las que hemos llamado *matizaciones* han sido a su vez distribuidas en dos de las tipologías básicas, concretamente en expresiones de significado equivalente (imprecisiones) y expresiones de significado distante (errores). Esta doble nomenclatura se debe a que todas las irregularidades encontradas en las fichas de calificación (cfr. Del Moral Manzanares, 2014) han sido clasificadas como *errores, imprecisiones o clasificaciones subjetivas*, lo que permite hacer una estadística general que engloba todas ellas.

Modificación de los descriptores		
	No significativa	17
Matizaciones	Equivalente (I)	72
	Distante (ε)	136

Tabla 20. Clasificación de las modificaciones de descriptores

Por último, la fig. 4 muestra el porcentaje de las tres subcategorías establecidas dentro de las modificaciones de descriptores.

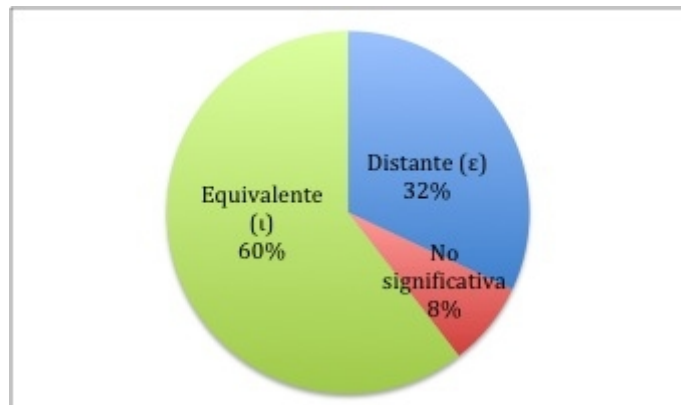


Fig. 4. Porcentaje de las diferentes modificaciones de descriptores.

5. CONCLUSIONES

Teniendo en cuenta todo lo observado en el apartado 3, podemos apuntar las siguientes recomendaciones para las sucesivas fases del proyecto:

5.1. EN RELACIÓN CON LOS DESCRIPTORES(8)

- Suprimir gradaciones difíciles de apreciar en las actuaciones de los candidatos (por ejemplo, descriptor fluidez 65: “ritmo bastante uniforme”).
- Suprimir frases condicionales (“si la situación lo requiere”) o posibles razones de la aparición de un determinado rasgo en el discurso del candidato (“los errores (...) pueden deberse a la voluntad de asumir riesgos”). Las frases deben ser afirmativas y responder a una descripción objetiva fácilmente contrastable con la realidad (por ejemplo, en fluidez 65, en lugar de “puede dudar mientras busca estructuras”, es preferible “en ocasiones duda mientras busca estructuras”).
- Evitar los descriptores negativos.

5.2. EN RELACIÓN CON LAS ANOTACIONES

- No realizar anotaciones suprimiendo o añadiendo a los descriptores adverbios o adjetivos, pues pueden alterar sustancialmente el significado original (por ejemplo, anotar “con pausas”, en lugar de “hace algunas pausas largas”, de fluidez 65, o “amplio”, en lugar de “muy amplio”, de alcance 85).
- Evitar realizar matizaciones de los descriptores.
- En el caso de percibir que la actuación de un determinado candidato está entre dos bandas, identificar los rasgos de cada una de estas que se corresponden con dicha actuación y plasmarlo en las anotaciones.
- Evitar hacer anotaciones que no se correspondan con el contenido de la banda correspondiente.
- No utilizar términos diferentes a los que aparecen en los descriptores, aunque se puedan sentir muy próximos a ellos (por ejemplo, “hilo” en lugar de “discurso”) o estén incluidos en otras bandas (por ejemplo, si “ritmo uniforme” aparece únicamente en la banda 65 no podrá ser utilizado para las anotaciones de ninguna de las demás).
- No introducir en la gradación de adjetivos o adverbios, variaciones inexistentes en los descriptores. Por ejemplo, “se sigue bastante bien”, en lugar de “fácil de seguir” (coherencia 95), ni suprimir las que están presentes.
- No negar total ni parcialmente los descriptores. Por ejemplo: “no extrae mucha información”, en lugar de “extrae información” (coherencia 85), aunque se sitúe el descriptor en una banda inferior a aquella en la que aparece la frase afirmativa.
- No realizar anotaciones que reflejen parcialmente un aspecto de los descriptores. Por ejemplo, en lugar de escribir en capacidad crítica 85 “Sus aportaciones van siempre encaminadas a apoyar su idea inicial”, anotar “hace aportaciones nuevas”.
- No incluir signos que puedan dificultar una correcta interpretación de las anotaciones, como interrogaciones y exclamaciones.

Por último, habría que apuntar que la observación detenida de las modificaciones de los descriptores realizadas por los evaluadores puede servir para mejorar las escalas, incluso tomando como buenas para futuras fases del proyecto algunas de las señaladas como imprecisiones, en caso de que consideremos que pueden facilitar la tarea del evaluador sin desvirtuar la referencia original de lo que es una correcta actuación del candidato.

?

NOTAS

(1) Dicho proyecto se enmarca en el programa de Doctorado en Lingüística Aplicada a la Enseñanza del Español como Lengua Extranjera de la Universidad Antonio de Nebrija de Madrid y se desarrolla en el Centro Lingüístico de la Universidad de Verona.

(2) Cfr. Del Moral Manzanares, 2013.

(3) Se han denominado *irregularidades* las anotaciones inapropiadas, incorrectas o defectuosas en relación con las normas establecidas. Todas ellas se consideran *errores, imprecisiones o clasificaciones subjetivas*, independientemente del subgrupo en el que se hayan clasificado.

(4) Para mayor detalle, v. Del Moral Manzanares (2014).

(5) En expresiones del tipo “1, C4E2, alcance 65”, el primer número (del 1 al 18) indica el candidato, después se alude a la calificación (en este caso, calificación 4 del evaluador 2), a continuación el criterio (alcance) y, por último, el nivel (65).

(6) V. Apartado 3 y, para mayor detalle, Del Moral Manzanares (2014).

(7) Para los cálculos estadísticos, se ha preferido separar las valoraciones numéricas de los cinco criterios existentes (denominadas *puntuaciones de criterio*) de las demás anotaciones (*anotaciones generales*). Así pues, el porcentaje de anotaciones con modificación de los descriptores se calcula sobre el total de *anotaciones generales*. Las puntuaciones de criterio, numéricas, no son susceptibles de introducir modificaciones de los descriptores.

(8) Es interesante apuntar que estas recomendaciones ya se tuvieron en cuenta antes de confeccionar la escala de descriptores utilizada. Aún así, su aplicación ha determinado la subsistencia de rasgos inapropiados y ha confirmado la conveniencia de tomar en consideración, en este sentido, los consejos de los expertos (cfr. Luoma 2004:63 y MCER:197).

REFERENCIAS BIBLIOGRÁFICAS

Alderson, J.C., Clapham, C. y Wall, D. (1998). *Exámenes de idiomas*. Madrid: Cambridge University Press.

Bordón, T. (2006). *La evaluación de la lengua en el marco de E/L2: Bases y procedimientos*. Madrid: Arco Libros.

Del Moral Manzanares, F. (2013). Escalas de descriptores y fiabilidad de la evaluación de la expresión e interacción orales del usuario competente. *Revista Nebrija de Lingüística Aplicada*, 13. <http://www.nebrija.com/revista-linguistica/revista-linguistica-nebrija13/htm/MoralFrancisco%20del.htm>

_____ (2014, en prensa). Las fichas de calificación en el proceso de validación de escalas de descriptores para la evaluación de la expresión oral de ELE. *Actas del XXVIII Congreso de la Asociación de Jóvenes Lingüistas*.

Instituto Cervantes (1997-2013). *Diccionario de términos clave de ELE*. http://cvc.cervantes.es/ensenanza/biblioteca_ele/diccio_ele/default.htm/

_____ (2002) (trad). *Marco común europeo de referencia para las lenguas: aprendizaje, enseñanza, evaluación*. Madrid: MEC-ANAYA. (Council of Europe (2001) *Common European Framework of Reference for Languages: Learning, Teaching, Assessment*. Strasbourg, Council of Europe). <http://cvc.cervantes.es/obref/marco/>

León, O. y Montero, I. (1997). *Diseño de investigaciones. Introducción a la lógica de investigación en Psicología y Educación*. Madrid: MacGraw-Hill.

Luoma, S. (2004). *Assessing Speaking*. Cambridge: Cambridge University Press.