

FRANCISCO DEL MORAL MANZANARES
UNIVERSITÀ DEGLI STUDI DI VERONA
UNIVERSIDAD ANTONIO DE NEBRIJA
LAS FICHAS DE CALIFICACIÓN EN EL PROCESO
DE VALIDACIÓN DE ESCALAS DE DESCRIPTORES
PARA LA EVALUACIÓN DE LA EXPRESIÓN ORAL DE ELE

1. INTRODUCCIÓN

El presente artículo se centra en el papel que han tenido las fichas de calificación en un proyecto aún en curso de validación de escalas de descriptores para la evaluación de la expresión e interacción orales de ELE. 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.

En la 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, realizada 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.

Así pues, en la primera fase de la investigación fue fundamental la utilización de una escala de descriptores, pero no fue esta la única herramienta que se añadió de modo experimental, sino que también se utilizaron unas fichas de calificación, concebidas en un principio para facilitar a los evaluadores una valoración global de sus apreciaciones, y que más tarde han sido una importante fuente de datos sobre el modo en que estos llevan a cabo su tarea. La observación de las mencionadas fichas ha arrojado información importante sobre cómo los evaluadores entienden y usan la escala, así como sobre el contenido de la misma (pertinencia de los descriptores, calidad, lagunas, etc.).

² Para una mayor información sobre los resultados de la primera fase del proyecto, v. Del Moral Manzanares (2013).

1. 1. CONTEXTO Y ESTRUCTURA DE LA INVESTIGACIÓN

Como ya se ha apuntado, el proyecto se está llevando a cabo en el Centro Lingüístico de la Universidad de Verona. Se basa en las pruebas orales de nivel C2³ 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 & Montero, 1997: 101). Los evaluadores implicados son dos: la profesora oficial del curso y el propio investigador. Cada uno de ellos evaluó cada prueba oral dos veces por medio de una valoración holística sin utilizar escalas de descriptores, y otras dos con ayuda de estas y de las fichas de calificación. Podemos resumir el proceso con el siguiente esquema⁴:



Fig. 1. Esquema de la primera fase de la investigación

1. 2. FIABILIDAD Y ESCALAS DE DESCRIPTORES

El objetivo principal de nuestra investigación es mejorar la fiabilidad de la prueba oral. Para ello se introdujeron en la primera fase las escalas de descriptores y las fichas de calificación, y en fases sucesivas se modificarán las escalas utilizadas según la información aportada por el análisis de las mencionadas fichas. Asimismo, se someterán después a una validación por parte de expertos y se utilizarán de nuevo en la evaluación de pruebas orales, para determinar el comportamiento de los índices de fiabilidad.

Se entiende por *fiabilidad* la “consistencia en la evaluación” (Bordón, 2006), o la “estabilidad (...) en la obtención de resultados” (Instituto Cervantes, 1997-2013), 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 & Wall, 1998).

³ El nivel C2 de nuestro centro equivale a 440 horas de clase repartidas en cuatro años académicos, durante los cuales el alumno también asiste a clases impartidas en la Facultad, que pueden tratar temas varios de gramática, traducción, literatura y cultura.

⁴ Para referirnos a cada serie de calificaciones, utilizamos un código que primero indica el número de calificación (C1, C2, C3, C4) y luego identifica al evaluador (E1, E2), de modo que, por ejemplo, C2E1 significa ‘calificación 2 del evaluador 1’; C3E2, ‘calificación 3 del evaluador 2’, etc.

Otro concepto básico en nuestra investigación es el de *descriptor*. 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).

La nuestra 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)⁵.

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 ha 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.

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 del análisis de los datos procedentes del diario de investigación y las fichas de calificación podrían extraerse informaciones muy útiles para las fases posteriores. Es precisamente este aspecto el que se desarrolla a continuación.

2. CARACTERÍSTICAS Y USO DE LAS FICHAS DE CALIFICACIÓN

Nuestras fichas de calificación se concibieron como una herramienta destinada a facilitarle al evaluador la tarea de expresar una valoración justa de la actuación del candidato, de manera que tuviera en cuenta sus rasgos fundamentales y los cinco criterios utilizados a tal fin⁶. Por ello se diseñaron en forma de tabla (las columnas se corresponden 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 facilitara la tarea. Se indicó que dichos apuntes podían ser tanto ejemplos concretos de la actuación del candidato como fragmentos de los descriptores que reflejaran el modo en que este se expresaba (v. fig. 2).

⁵ Dichas puntuaciones responden a la necesidad de nuestro centro de dar una puntuación sobre cien.

⁶ Como se puede observar en la fig. 2, el espacio destinado al criterio de coherencia aparece dividido en dos, para diferenciar entre la actuación del monólogo y la del diálogo (las dos partes de la prueba). De un modo análogo, la división del espacio dedicado a la fluidez se hace para facilitar las anotaciones que se refieren a la fluidez propiamente dicha, por una parte, y la pronunciación, por otra.

NOMBRE:
NÚMERO DE MATRÍCULA
FECHA:

	Coherencia		Fluidez		Cap. Crítica	Corrección	Alcance
95							
85							
75							
65							
50							

NOTA MEDIA:
CALIFICADOR:

Fig. 2. Ficha de calificación en blanco

En nuestra convocatoria, para la identificación del candidato, se utilizó simplemente el número de matrícula, con el fin de evitar que el evaluador pudiera asociar un nombre a una calificación determinada de una evaluación previa, y este hecho pudiera perjudicar consecuentemente su fiabilidad⁷.

En la fig. 3 aparece una 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 agregó un código en la parte superior izquierda (en nuestro ejemplo, destacado en azul claro) que identifica tanto el orden de la calificación como al evaluador (en este caso, se trata de la calificación 3 del evaluador 1, C3E1). 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, 89).

⁷ En la segunda fase del proyecto (aquella en la que se centra el presente trabajo), se ordenaron los números de matrícula y se sustituyeron por el número correspondiente del 1 al 18, de modo que fuera más fácil hacer referencia a una calificación determinada. Por ejemplo, "13, C3E1, coherencia", alude al candidato 13, calificación 3 del evaluador 1, criterio de coherencia.

C3E1

Fichas de calificación
Prueba oral-C2LM

Nombre:
Número de matrícula: 365268
Fecha:

	Coherencia	Fluidez	Cap. Crítica	Corrección	Alcance
95	introducción clara eficaz nivel apropiado fácil de seguir cohesión de	fluidez natural (95)	buen capacidad de escuchar		
85		solo se interrumpe a dar explicaciones	expresa opiniones ya enriquecer contenido (85)	errores poco signif. (85)	repertorio amplio (85)
75					
65					
50		consonante doble verdad =		el tema es lo mismo =	

Nota media: 89
Calificador: 89

Fig. 3. Ficha de calificación completa modificada con colores

3. ANÁLISIS DE LAS FICHAS DE CALIFICACIÓN

Antes de proceder a la clasificación de los fenómenos observados, nos resulta imprescindible definir dos términos a los que haremos continuamente referencia: *anotaciones* e *irregularidades*.

Consideramos *anotaciones* cada frase, palabra, grupo de palabras o signo independientes que el evaluador escribe en las fichas de evaluación. Nuestro trabajo en esta fase del proyecto ha consistido principalmente en el análisis de todas ellas, teniendo en cuenta los acuerdos tomados previamente por los evaluadores y el contenido de la escala de descriptores. Así pues, llamaremos *irregularidades* a las anotaciones inapropiadas, incorrectas o defectuosas, en relación a las normas establecidas.

3. 1. IRREGULARIDADES DE LOS EVALUADORES EN LA APLICACIÓN DE LAS ESCALAS

El primer grupo de nuestra clasificación lo forman aquellas irregularidades motivadas por que el evaluador ha pasado por alto o malinterpretado alguna información que le ha sido dada y que, de hecho, en otras ocasiones ha usado correctamente. Distinguimos en este apartado entre las irregularidades *de valoración* y las *de procedimiento*.

3. 1. 1. De valoración

Incluimos dentro de este grupo aquellos casos en que el evaluador hace una ponderación irregular de rasgos clasificados correctamente, o bien aquellos en que, después de detectar un rasgo digno de ser anotado, hace una clasificación equivocada.

3. 1. 1. 1. Puntuación incorrecta

Cuando, dentro de un criterio de evaluación determinado, el evaluador registra anotaciones en bandas diferentes⁸, se encuentra con la dificultad de dar una puntuación global a todas ellas. Por ejemplo, un evaluador que registrara anotaciones en los niveles 85, 75 y 50, podría dudar entre decidirse por una puntuación de 80, 75 o 70. Partiendo de la imposibilidad de crear un mecanismo de ponderación exacta, hemos considerado correcta cualquier puntuación comprendida entre el valor máximo y el mínimo (en este caso ficticio, entre 85 y 50), sin incluir los extremos. Consecuentemente, se han considerado irregularidades las puntuaciones situadas:

- Por encima de la banda de mayor valor
- Por debajo de la banda de menor valor
- En uno de los extremos, cuando existan anotaciones en más de una banda.

Teniendo en cuenta lo expuesto anteriormente, la fig. 4 muestra el porcentaje de puntuaciones de criterio, sobre un total de 360⁹, y diferencia entre:

- Beneficiosas: refleja un valor excesivamente alto, teniendo en cuenta las anotaciones recogidas.
- Perjudiciales: refleja un valor excesivamente bajo, teniendo en cuenta las anotaciones recogidas.
- No justificadas: se trata de los casos en que no se han realizado anotaciones, por lo que la corrección de la puntuación otorgada no puede comprobarse.

⁸ A pesar de que este tipo de irregularidades es mucho más frecuente en los casos en que el evaluador realiza anotaciones en más de una banda, también se da en algunos casos en los que tan solo se ha registrado una anotación (por ejemplo, el evaluador hace solo una anotación en 50 y da una puntuación de 95). Esto puede deberse, sin duda, a que el evaluador está teniendo en cuenta para otorgar la puntuación del criterio correspondiente no solo lo que anota sino también lo que escucha pero no añade a la ficha.

⁹ Como ya hemos explicado, nuestra escala de descriptores contempla cinco criterios (coherencia, fluidez, capacidad crítica, corrección y alcance). Cada prueba oral recibe, pues, cinco puntuaciones de criterio. Si tenemos en cuenta que hay dieciocho candidatos, tenemos un total de noventa (5x18=90) puntuaciones de criterio en cada serie de calificaciones. Y al ser cuatro las series de calificaciones que se realizan con ayuda de la escala de descriptores y las fichas de calificación (C3E1, C3E2, C4E1 y C4E2), obtenemos un total de 360 (90x4=360) puntuaciones de criterio.

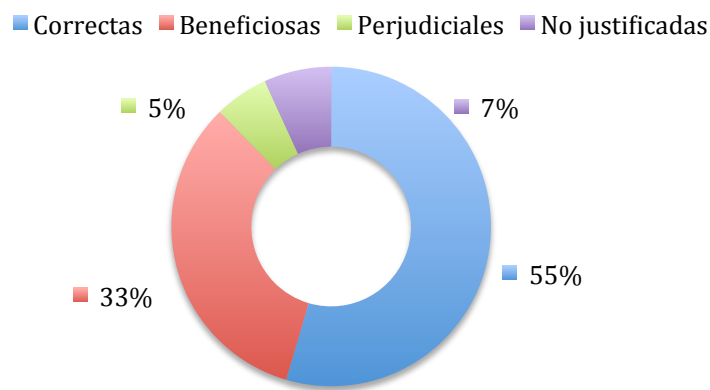


Fig. 4. Puntuaciones de criterio

Se observa, pues, que en los casos en que los evaluadores cometen irregularidades al decidir una puntuación de criterio, existe la tendencia de beneficiar al candidato¹⁰.

3.1.1.2. Error en la asignación de banda

Se trata de aquellos casos en que el evaluador se ha equivocado en la atribución de una banda concreta, dentro del criterio adecuado. Por ejemplo, en nuestra escala, el error que consiste en pronunciar “v” como /v/, por influencia del italiano, está situado en la banda 65, por lo que los casos en que algún evaluador lo ha anotado en otro nivel diferente, (por ejemplo, en 50) se han considerado irregularidad.

3.1.1.3. Error en la asignación de criterio

Este tipo de irregularidades se produce cuando se ha atribuido un fragmento de la actuación del candidato (se trata casi siempre de errores) a un criterio que no corresponde a lo establecido por la escala de descriptores (por ejemplo, fluidez en lugar de alcance).

3.1.2. De procedimiento

En este caso, es en el propio proceso de anotación donde se produce la irregularidad.

3.1.2.1. Modificación de los descriptores¹¹

Cuando el evaluador opta por anotar un fragmento del descriptor que refleja la actuación del candidato, en ocasiones no respeta lo que aparece escrito en la escala. En este sentido, se distingue entre:

- No significativa (no se considera irregularidad)
- Significativa (irregularidad)

¹⁰ Es interesante apuntar que Karavas y Delieza (2009) han observado una tendencia parecida en la actuación de los examinadores de pruebas orales. Concretamente, según el citado estudio, las intervenciones de los examinadores / interlocutores suelen estar encaminadas a facilitar el *output* del candidato.

¹¹ Para ampliar este punto, v. Del Moral Manzanares (2014).

Este tipo de irregularidad es importante porque, en el momento de escribir una anotación en las fichas, los evaluadores deberían reflejar los descriptores con la mayor fidelidad posible, 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, tal vez por razones de eficacia, introduce en sus anotaciones cambios de mayor o menor entidad respecto a las expresiones contenidas en los descriptores, existe el peligro de que, a medida que va examinando, se desvirtúe 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.

3.1.2.2. Falta de claridad

Este apartado incluye aquellas anotaciones incomprensibles o de sentido impreciso, así como signos o símbolos no acordados previamente cuyo significado solo es evidente para el evaluador en el momento de su utilización.

3.1.2.3. Anotación entre dos bandas

Este tipo de irregularidades se comete cuando, al escribir, el evaluador se sale del espacio reservado para un determinado criterio y banda, causando un problema en la interpretación de la anotación.

3.2. IRREGULARIDADES CAUSADAS POR EL CONTENIDO DE LOS DESCRIPTORES

Por lo que se refiere a este segundo grupo, el análisis de la información contenida en las fichas de calificación nos ha permitido detectar que algunas de las irregularidades se deben a que el evaluador necesita información que no está disponible en la escala, o que, si lo está, no es lo suficientemente clara. Esta circunstancia lo lleva a tomar decisiones basándose en su propio criterio (clasificaciones subjetivas), o bien en datos que por sí mismos no deberían bastar para identificar la actuación del candidato con una banda concreta (clasificaciones realizadas con información insuficiente). Concretamente, las irregularidades de este grupo pueden dividirse en:

3.2.1. Clasificaciones subjetivas

3.2.1.1. Por la ausencia de rasgos correctos en la escala

Se trata de fragmentos correctos de la actuación del candidato (construcciones gramaticales, léxico apropiado, conectores, etc.) que el evaluador intenta clasificar sin la ayuda de la escala de descriptores, pues no aparecen en ella.

3.2.1.2. Por la ausencia de errores en la escala

En este caso, son errores cometidos por el candidato no contemplados por la escala de descriptores y que el evaluador trata de encajar en una banda determinada, aplicando su propio criterio.

3.2.2. Clasificaciones realizadas con información insuficiente

Se deben a que hay fragmentos de un descriptor que aparecen en varias bandas y a que el evaluador utiliza uno de estos fragmentos como única referencia para decidir inscribir esa característica de la actuación del candidato en un nivel determinado. Sin embargo, debido a que se trata de rasgos presentes en más de un nivel, para decidirse por uno de estos, el evaluador debería haber tenido en cuenta, además, algún otro rasgo exclusivo de la banda en la que ha colocado la anotación.

4. CONCLUSIONES

La figura 5 representa gráficamente el porcentaje de anotaciones apropiadas e irregulares, así como el tipo de irregularidad de cada una de estas últimas, según la clasificación desarrollada más arriba, excluyendo las puntuaciones de criterio¹². Como puede verse, los grupos de irregularidades más numerosos son la modificación significativa de los descriptores y las clasificaciones subjetivas de segundo tipo (debidas a que algunos errores frecuentes en nuestros estudiantes no están contemplados en la escala utilizada).

Del análisis de las fichas de calificación que hemos realizado se desprenden tres tipos de conclusiones: las que nos sirven para mejorar la fase de formación de los evaluadores, las que podemos aplicar al perfeccionamiento de la escala de descriptores y las que resultan útiles para desarrollar las fichas de calificación.

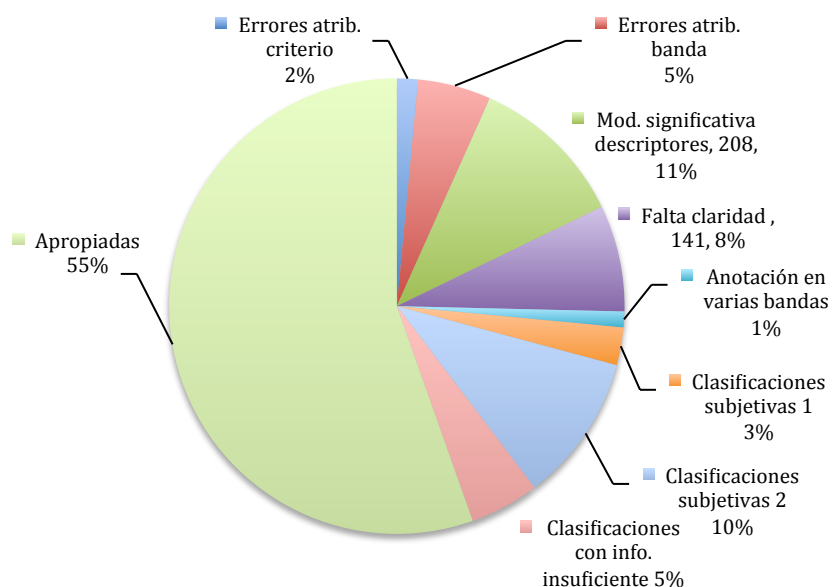


Fig. 5. Cuadro de anotaciones generales

¹² Para presentar los datos estadísticos, hemos preferido separar las irregularidades que se refieren a las puntuaciones de criterio de las demás irregularidades, que hemos llamado *generales*.

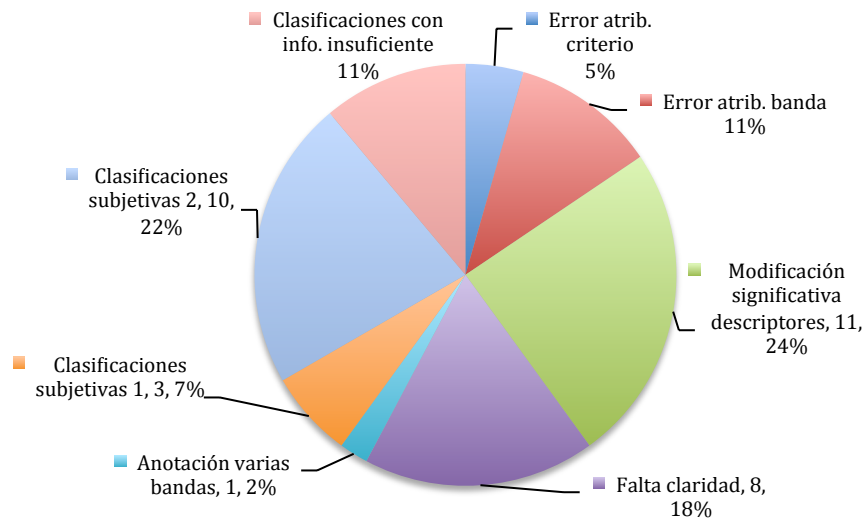


Fig. 6. Cuadro de irregularidades generales

4.1. PARA LA FORMACIÓN DE EVALUADORES

En lo que se refiere a la fase de instrucción de los evaluadores, la observación de las fichas de calificación nos ha permitido establecer dos ejes de actuación.

El primero es la introducción de nuevos elementos: criterios de anotación y criterios de ponderación de las anotaciones.

Antes de llevar a cabo la primera serie de calificaciones no se hizo una reflexión suficientemente profunda sobre cuáles eran los elementos susceptibles de ser anotados y, sobre todo, cómo había que realizar las anotaciones. Es necesario reafirmar de cara a la continuación del proyecto que los descriptores han de ser la base de las anotaciones y, asimismo, hay que establecer normas claras sobre dónde anotar ejemplos de la actuación de los candidatos que planteen dudas (por ejemplo, casos de errores fronterizos entre léxico y fonética). Igualmente, sería muy útil crear un código fácil de utilizar en caso de que el evaluador necesite rectificar una anotación, eliminarla o cambiarla de ubicación, así como indicar la repetición de un rasgo determinado, todo ello para evitar las anotaciones inapropiadas por falta de claridad (18% de las irregularidades generales) y modificaciones significativas de los descriptores (24% de las irregularidades generales).

Otra deducción que nos han permitido las fichas ha sido la necesidad de decidir unos mínimos criterios de ponderación de las anotaciones para puntuar los diferentes criterios. Por un lado, hay que tener en cuenta la calidad general de la actuación y, por otro, la cantidad y la entidad de los errores. Es importante que las anotaciones reflejen ambos aspectos y que todos ellos sean tenidos en cuenta para otorgar cada una de las puntuaciones de criterio. También es necesario plantearse si, en los de corrección y

alcance, debemos establecer una jerarquía que nos permita contabilizar errores fosilizados o logros situados en un nivel superior a la actuación general del candidato.

El segundo eje de actuación en la formación de los evaluadores es la mejora de algunos aspectos ya tenidos en cuenta en la primera fase del proyecto. Concretamente, hemos de insistir en la importancia de que los evaluadores tengan un buen conocimiento de la escala y los criterios de evaluación antes de tener que utilizarla. Han de estar muy familiarizados con los rasgos de la expresión e interacción orales considerados en cada criterio de evaluación, el significado exacto y consensuado de los términos clave y los aspectos distintivos de cada nivel. Para conseguir una evaluación lo más correcta posible no basta con que los evaluadores demuestren cierta destreza en la utilización de las escalas y las fichas de calificación, sino que además han de tener previamente un profundo conocimiento de las mismas. De este modo, creemos que se reducirán al máximo los errores de atribución de banda y criterio, que juntos suponen el 16% de las irregularidades generales.

4.2. PARA LA MEJORA DE LA ESCALA DE DESCRIPTORES

Hemos advertido que la aplicación de las escalas sería más fácil si en ellas se introdujeran dos tipos de modificaciones: en primer lugar, la supresión de una serie de rasgos, como son gradaciones sutiles (“ritmo bastante uniforme”), oraciones condicionales (“si la situación lo requiere”), posibles causas (“los errores pueden deberse a...”) o la excesiva longitud de muchos de los descriptores¹³. Todo ello haría más fácil identificarlos con la actuación de un determinado candidato y creemos que reduciría también la tendencia de los evaluadores a modificar los descriptores en las anotaciones, para adaptarlos a sus necesidades.

También es necesario suprimir la repetición de fragmentos de descriptores en bandas diferentes y sustituirlos por otros que sean exclusivos de cada nivel, de modo que todos los logros incluidos en una determinada banda se consideren, por defecto, propios no solo del nivel en que aparecen, sino de todos los superiores, sin necesidad de que esté explícitamente expreso en los descriptores correspondientes¹⁴. Esta mejora evitaría lo que hemos llamado *clasificaciones realizadas con información insuficiente* (11% de las irregularidades generales).

En segundo lugar, también hemos observado la necesidad de añadir algunos elementos: ejemplos representativos tanto de actuaciones correctas como de erróneas, criterios de

¹³ Es interesante subrayar que varias de estas recomendaciones fueron ya tenidas en cuenta antes de la creación de la escala de descriptores, pues forman parte de lo aconsejado por los estudiosos, pero después de su utilización han cobrado mayor importancia.

¹⁴ Por ejemplo, el rasgo “pronunciación clara y natural” que se repite en las bandas 75, 85 y 95 de nuestro criterio de fluidez, debería aparecer solo en el más bajo de ellos (75) y considerarse un rasgo propio de 85 y 95, sin que este aparezca en sus respectivos descriptores. De este modo, el evaluador deberá decidirse por la banda 85 o 95 basándose en otros rasgos que no sean propios también de niveles más bajos.

evaluación de conectores (coherencia) y criterios de evaluación de léxico (alcance), que reducirían la aparición de clasificaciones subjetivas (29% de las irregularidades generales).

4.3. PARA EL DESARROLLO DE LAS FICHAS DE CALIFICACIÓN

Por último, creemos que, para evitar futuros errores en la atribución de banda y criterio, nuestras fichas de calificación podrían incluir una referencia a rasgos concretos (tanto correctos como erróneos), cuya existencia en la actuación del candidato podría indicarse con una simple marca gráfica (por ejemplo, una casilla que rellenar en el caso de observar la presencia del rasgo correspondiente, según se muestra en la fig. 7). Ello reduciría el riesgo de equivocarse en el momento de anotar una determinada característica del discurso del candidato. Si modificamos en este sentido nuestras fichas, compartirían ciertas características con las llamadas *listas de control*, si bien estas, según el MCER, responden a qué contenidos de un elenco determinado (de un curso, por ejemplo) ha adquirido el candidato.

65	/z/ <input type="checkbox"/> /ʎ/ <input type="checkbox"/>
50	ceceo <input type="checkbox"/> acentuación <input type="checkbox"/> cons. dobles finales <input type="checkbox"/>

Fig. 7. Posible modificación de las fichas de calificación (fluidez)

5. REFERENCIAS BIBLIOGRÁFICAS

Alderson, J.C., Clapham, C. & 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. Disponible en la web:
<http://www.nebrija.com/revista-linguistica/revista-linguistica-nebrija13/htm/MoralFrancisco%20del.htm>

Del Moral Manzanares, F. (2014), "Dificultades en el uso de los descriptores en el proceso de evaluación de la expresión oral de ELE", *Revista Nebrija de Lingüística Aplicada*, 8 (13).

Dörnyei, Z. (2010), *Research Methods in Applied Linguistics*. Oxford: Oxford University Press.

Instituto Cervantes (1997-2013), *Diccionario de términos clave de ELE*. Disponible en la web:

http://cvc.cervantes.es/ensenanza/biblioteca_ele/diccio_ele/default.htm

Instituto Cervantes (2002) (trad.), *Marco común europeo de referencia para las lenguas: aprendizaje, enseñanza, evaluación*. Madrid: MEC-D-Anaya. (Council of Europe (2001). *Common European Framework of Reference for Languages: Learning, Teaching, Assessment*. Strasbourg: Council of Europe). Disponible en la web:

<http://cvc.cervantes.es/obref/marco>

Karavas, E. & Delieza, X. (2009), "On site observation of KPG oral examiners: Implications for oral examiner training and evaluation", *Apples-Journal of Applied Language Studies*, 3 (1): 51-77.

León, O. & 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.