



Universidad
Nebrija

Facultad de Lenguas Aplicadas y Humanidades
Departamento de Lenguas Aplicadas

**LA OBJETIVACIÓN DE LOS DESCRIPTORES
DE EVALUACIÓN DE LA EXPRESIÓN ORAL EN
LA ENSEÑANZA DE E/LE A ALUMNOS
ADULTOS EN INMERSIÓN LINGÜÍSTICA**

Masteranda: Raquel Gutiérrez Acosta

Tutora: Susana Martín Leralta

**Máster Universitario en Lingüística aplicada
a la enseñanza del español como lengua extranjera**

Curso académico 2009/11

ÍNDICE

1. INTRODUCCIÓN	4
1.1. Justificación de la investigación	5
1.2. Objetivos	10
1.2.1. Objetivos generales	10
1.2.2. Objetivos específicos.....	10
2. ESTADO DE LA CUESTIÓN	11
2.1. Precisiones terminológicas	11
2.1.1. Expresión oral: definición del constructo.....	11
2.1.2. “Evaluar” <i>versus</i> “examinar”	14
2.1.3. “Criterios de evaluación” <i>versus</i> “criterios de calificación”	16
2.1.4. “Descriptor de capacidad lingüística” <i>versus</i> “escalas de niveles”	17
2.1.5. Competencia comunicativa: conocimiento y uso de la lengua	21
2.2. Características de las herramientas de evaluación	30
2.2.1. Fiabilidad, validez y viabilidad	30
2.2.2. Autenticidad, interactividad e impacto.....	37
2.2.3. Otros factores condicionantes en la evaluación	38
2.3. Proyectos y estudios recientes	40
2.4. Conclusiones	45
3. MARCO TEÓRICO	47
3.1. Elaboración de pruebas de evaluación de calidad	47
3.1.1. Pruebas de evaluación de la expresión oral.....	49
3.1.2. Criterios de evaluación de la expresión oral	57
3.1.3. Obtención y tratamiento de datos.....	62
3.2. La formación de los evaluadores	63
3.3. Dimensión ética de la evaluación	66
3.4. Conclusiones	68
4. ESTUDIO EMPÍRICO	69
4.1. Planteamiento del estudio empírico	69
4.2. Metodología y diseño	70
4.2.1. Paradigma de investigación.....	70
4.2.2. Variables de nuestra investigación.....	73
4.2.2.1. Variables de contexto	74
4.2.2.1.1. EOI Santa Cruz de Tenerife.....	74
4.2.2.1.2. Organización de las enseñanzas oficiales de idiomas en la Comunidad Autónoma Canaria	75
4.2.2.2. Variables de presagio	76
4.2.2.2.1. Departamento de Español para extranjeros	76
4.2.2.2.2. Tipo de alumnado que estudia español en la EOI Santa Cruz de Tenerife.....	76

4.2.2.3. Variables de proceso	78
4.2.2.3.1. Estructura de los estudios de español para extranjeros y metodología de enseñanza y evaluación.....	78
4.2.2.3.2. Metodología en las clases de español	79
4.2.2.3.3. Proceso de comprobación, análisis y aprendizaje de la expresión oral de los alumnos	80
4.2.2.4. Variables de producto.....	82
4.2.2.4.1. La evaluación en la EOI Santa Cruz de Tenerife	82
4.2.2.4.2. Elaboración y evaluación de las pruebas de certificación de idiomas de la Comunidad Autónoma Canaria	88
4.2.3. Herramientas utilizadas en nuestro estudio.....	89
4.2.4. Desarrollo del estudio empírico: procesos para la evaluación de la expresión oral atendiendo a las variables.	91
4.2.4.1. Ciclo 1.	91
4.2.4.2. Ciclo 2	100
4.2.4.3. Ciclo 3	107
4.2.4.4. Ciclo 4	117
4.3. Limitaciones de la investigación	126
4.4. Conclusiones de nuestra investigación.....	128
4.5. Futuras investigaciones	132
5. CONCLUSIONES.....	134
6. BIBLIOGRAFÍA.....	138
7. ANEXOS	143

1. INTRODUCCIÓN

El interés suscitado por la expresión oral durante nuestra observación y experiencia en el aula en la enseñanza del español y los problemas que conlleva evaluarla nos ha llevado a realizar esta investigación, que pretende encontrar maneras de evaluar de forma más precisa y objetiva la producción oral de los candidatos. Durante varios cursos escolares hemos observado las continuas quejas de los profesores y alumnos sobre la forma de evaluar la expresión oral; muchos profesores se limitan a recoger todos los errores que comete el alumno y, a partir de su intuición o de la interpretación subjetiva de los criterios de evaluación existentes, determinan un resultado, provocando en el alumno cierta sensación de falta de fiabilidad y validez en ellos. En la mayoría de las ocasiones estos resultados están “impregnados” de subjetividad provocada, la mayor parte de las veces, por la vaguedad en la formulación de los descriptores de evaluación existentes, así como por la falta de relación entre las tareas que se utilizan en la evaluación, los criterios de evaluación y el constructo que las sustenta. De esta manera, en nuestra investigación tenemos por objetivo comprobar si mediante la formulación de descriptores y puntuaciones más precisas, relacionadas con las pruebas de evaluación de la expresión oral, se puede atenuar la subjetividad en la evaluación de ésta y hacer que los resultados sean más acordes, válidos y fiables con la producción oral real de los candidatos.

Así, en el primer capítulo explicamos detalladamente el problema que hemos observado del que surge esta investigación y los objetivos que nos hemos planteado, para en el segundo capítulo centrarnos en el estado de la cuestión donde trataremos de aclarar algunos términos utilizados en nuestro estudio y las teorías lingüísticas que subyacen en ellos (Hymes (1971), Canale y Swain (1980), Canale (1983) Bachman y Palmer (1990 y 1996) y Bordón (2006), entre otros), así como explicaremos algunos proyectos y estudios recientes relacionados con la evaluación de la expresión oral.

En el capítulo tres detallaremos el marco teórico en el que se sustenta nuestro estudio empírico citando a los principales investigadores y teorías lingüísticas relacionadas con nuestra investigación, centrándonos fundamentalmente en cómo elaborar pruebas y criterios de calidad y garantizar la idoneidad de estos. Por último, en el capítulo cuatro explicaremos el desarrollo de nuestra investigación donde se expone el estudio empírico llevado a cabo para la investigación, adoptando el modelo de Elliott (1993) basado en el

proceso de investigación-acción orientándolo fundamentalmente a la evaluación de la expresión oral del español de nivel B2 en la Escuela oficial de idiomas (a partir de ahora EOI) de Santa Cruz de Tenerife. Para ello, estudiamos las pruebas y los criterios de evaluación aplicados a ellas tanto en clase como en las pruebas oficiales de certificación y elaboramos nuevas pruebas y criterios según lo observado en el resultado de nuestras investigaciones, con el objeto de comprobar su validez y fiabilidad. Asimismo, diseñamos las pruebas orales de certificación del último curso escolar 2010/2011 y las evaluamos aplicando los criterios de evaluación existentes estableciendo las correspondientes conclusiones. Todo ello se experimentó en el aula con 10 alumnos oficiales de español de nivel B2 y en las pruebas de certificación aplicadas a estos mismos alumnos. En este último capítulo explicaremos la metodología y herramientas utilizadas, las variables implicadas, el análisis de resultados, así como las limitaciones y conclusiones de dicho estudio empírico. Finalmente añadimos la bibliografía utilizada y los anexos con algunos de los instrumentos de recogida de datos empleados y explicados de manera más detallada.

1.1. Justificación de la investigación

La elaboración y evaluación de exámenes, fundamentalmente los de gran escala como son los de certificación, y otras muchas circunstancias observadas en la enseñanza hacen que el docente que ya tiene cierta experiencia advierta en su práctica diaria una serie de dificultades que, dependiendo de múltiples variables, suelen ser muy diferentes. Sin embargo, en la observación diaria de la práctica docente se aprecia una dificultad común que encuentra gran parte de los profesores no sólo de enseñanza de E/LE sino también de otros idiomas: la subjetividad y fiabilidad de los criterios de evaluación de la expresión oral. En la mayoría de las ocasiones, las pautas de medición están expresadas de manera vaga y desconocida por los candidatos y evaluadores.

Esta investigación surge de esta observación en el aula así como del descontento que, sobre este aspecto, manifiestan los profesores de enseñanza de idiomas de la Comunidad Canaria en general y el propio alumnado en particular, quien percibe que la calificación de la prueba de expresión oral, fundamentalmente en los exámenes oficiales de certificación, depende más de interpretaciones o apreciaciones hechas por alguien de las muestras de actuación obtenidas, que de datos objetivos que ellos conozcan y que puedan demostrar su verdadero nivel en esta capacidad. Además, los mismos alumnos

muestran su preocupación con su dominio de la lengua oral a pesar de haber superado con éxito el curso escolar o las pruebas de certificación. Ellos mismos consideran que su nivel de expresión oral es inferior al determinado por el resultado del examen y en otras ocasiones ocurre lo contrario: creen que su dominio en dicha capacidad es superior y que la puntuación asignada a la expresión oral como resultado de la evaluación no se corresponde con su dominio real del idioma.

Nuestra investigación se enmarca en la EOI Santa Cruz de Tenerife y el modelo que sigue para la enseñanza de L2 es comunicativo, por lo que la manera de evaluar también lo debe ser para determinar la competencia de un aprendiz de manera idónea, esto es, si un alumno es competente en una lengua cuando posee un dominio suficiente de su gramática, vocabulario, funciones comunicativas, recursos retóricos, variedades dialectales y sociales, registro, organización y cohesión, fluidez y la interpretación correcta de varios componentes (aspectos de la competencia comunicativa) la evaluación y los criterios de evaluación también deben estar hechos en función de estos conocimientos.

Siguiendo el *Marco común europeo de referencia de las lenguas* (MCER, 2002) la evaluación de una lengua debe ser analítica: es necesario que se valoren separadamente¹ distintos elementos, en este caso, de la producción oral. En este sentido, los exámenes de las EOI de Canarias sí creemos que cumplen con este requisito, la dificultad se plantea cuando debemos determinar qué escala es la que debemos utilizar para delimitar qué grado de dominio tiene un candidato en cada uno de estos aspectos. Además, ¿cómo sabemos que su grado de dominio, por ejemplo, en el léxico es muy alto, alto, medio, suficiente o bajo? Se recurre de nuevo a la subjetividad del profesorado.

Aquí es donde se hace necesaria una actuación que haga fiables los resultados creando una escala clara y objetiva, ya que para que un examen sea idóneo debe cumplir tres criterios: validez, fiabilidad y viabilidad (Canale y Swain (1980), Bachman y Palmer (1990 y 1996), MCER (2002), Bordón (2006)). Casi todos los profesores elaboran exámenes con validez por lo menos se cumple la validez de contenido y aparente, es decir, los confeccionan en la medida que se supone que miden lo que deben de medir, y la viabilidad siempre se intenta garantizar. Pero en la expresión oral es muy difícil que

¹ Aunque sabemos que todos ellos están relacionados y que no se pueden dar los unos sin los otros, a la hora de evaluar se pueden observar de manera más detallada mediante una escala analítica en lugar de holística.

se cumpla el requisito de fiabilidad en cuanto a la consistencia de su medición, a no ser que los criterios estén bien elaborados.

El problema de la expresión oral es que la producción del alumno no es una respuesta cerrada que permite calificarla como correcta o incorrecta, sino que el candidato produce respuestas abiertas y originales que requieren la intervención de una persona para valorarlas. La dificultad está en cuánta subjetividad utiliza el evaluador para calificar la prueba y si es posible eliminarla o, al menos, atenuarla en la evaluación de la expresión oral.

Por otro lado, el impacto (Bachman y Palmer 1996: 30), es decir, el efecto que produce el examen en los individuos (“*a micro level*”) y en la sociedad y en el sistema educativo (“*a macro level*”), que genera un examen de certificación es tal que los criterios de evaluación que aseguren resultados reales y fiables son muy importantes en la evaluación de la competencia de la expresión oral de un candidato, ya que implica a los individuos que diseñan las pruebas y los criterios para esos exámenes, a los candidatos que los realizan y a los profesores que preparan a esos candidatos². Además, este tipo de examen repercute muchísimo en la enseñanza del español en los cursos pares fundamentalmente (Básico 2, Intermedio 2 y Avanzado 2), ya que una parte del curso se orienta a preparar a los candidatos para esta prueba aunque no estén obligados a hacerla. Actualmente, como ya hemos dicho anteriormente, el alumno puede elegir, cuando llega a ese curso, realizar la prueba de certificación y conseguir, así, un título o promocionar y pasar al siguiente nivel sin realizar la prueba de certificación, siempre que haya aprobado como alumno oficial el nivel anterior o que haya hecho una prueba de clasificación que lo sitúe en el nivel que quiere cursar. Por todo ello, es fundamental “garantizar la calidad de la herramienta de medición” ya que “es una manera de asegurar la validez del constructo, y, en definitiva, de dotar de validez, fiabilidad y viabilidad al examen” (Bordón 2006: 67).

En esta investigación partimos de la hipótesis de que los problemas que observamos en la evaluación de la expresión oral se pueden deber, además, a que la corrección de las pruebas de expresión oral, debido a su propia naturaleza, son mucho más difíciles de medir que las de comprensión, por ejemplo, y siempre entrañan el riesgo de subjetividad, ya que los procedimientos llevados a cabo por la evaluación no suelen ser

² En el caso de que sean alumnos oficiales dentro del contexto de nuestra investigación, es decir, alumnos que estudian o han estudiado el idioma en alguna institución educativa.

mecánicos ni objetivos. Pero ¿se puede evaluar la expresión oral mediante procedimientos mecánicos y objetivos que permitan una mayor fiabilidad de la prueba? o, dicho de otra manera, ¿cómo podemos evaluar de una manera más objetiva muestras de lengua que ofrecen unos resultados que deben ser interpretados? Ya Bordón (2006: 237) planteaba esta misma pregunta y daba algunas soluciones teóricas a ello, pero sería muy interesante investigarlo en la acción en el aula y en la evaluación de los exámenes de certificación, para comprobar la eficacia de las soluciones que propone, así como la aportación de otras que surjan de dicha investigación. Igualmente el MCER (2002: 189) señala algunos pasos a seguir para reducir la subjetividad en la evaluación y aumentar la validez y la fiabilidad de las pruebas:

- Desarrollar una **especificación del contenido** de la evaluación, basada por ejemplo en un marco de referencia común al contexto de que se trate.
- Utilizar **valoraciones compartidas** para seleccionar el contenido y evaluar las actuaciones.
- Adoptar **procedimientos normalizados** respecto a la forma de realizar las evaluaciones.
- Proporcionar **claves definitivas de puntuación** para las pruebas indirectas y fundamentar las valoraciones de las pruebas directas sobre **criterios específicos definidos**.
- Exigir múltiples valoraciones y, en su caso, la ponderación de distintos factores.
- Realizar una **formación** adecuada en relación con las **directrices de evaluación**.
- Comprobar la calidad de la evaluación (validez, fiabilidad) **analizando los datos de evaluación**.

Creemos que si los descriptores de corrección están bien definidos, si se formulan de manera tan precisa que los evaluadores no puedan salirse de esa escala y, sobre todo, si se adecuan al contexto y al tipo de alumnado, la subjetividad se podría mitigar.

Además, la fiabilidad está muy ligada al modelo de examen y a cómo están formuladas sus preguntas. Aunque los profesores intenten o piensen que hacen pruebas válidas, ¿estas miden realmente lo que un alumno sabe y los ítem que en ellas se proponen son realmente “justos” para todos los estudiantes? Los alumnos, dependiendo de sus costumbres y su cultura, aprenden de distinta manera; así, por ejemplo, un asiático necesita más tiempo para contestar las preguntas que un alemán o, en una interacción oral, un francés o italiano se sentirá más cómodo que un asiático y no reaccionará de la misma manera. ¿Estas apreciaciones no debemos tenerlas en cuenta a la hora de confeccionar exámenes y de evaluar?, o dicho de otro modo, ¿los aspectos socioculturales no deberían tener cabida en los descriptores de la evaluación de la expresión oral?. Desde nuestro punto de vista, hasta ahora no se ha tratado este aspecto con la suficiente profundidad, y creemos que es un punto muy importante dentro de la

evaluación que mejoraría la validez de los exámenes en general y de las pruebas orales en particular.

A este respecto, Yagüe (2010: 23) apunta varias posibles soluciones que son dignas de investigar para comprobar su efectividad y

posiblemente no sería posible adoptar decisiones inmediatas en lo referido a la calificación del examen, pero sí respecto a la evaluación en general y al diseño mismo de pruebas de medición. En cualquier caso, los diseños de pruebas con este perfil son complejos y se deberían asentar sobre investigaciones y estudios de campo relevantes.

No podemos permanecer ajenos a la influencia del componente sociocultural en la elaboración y en la medición de los exámenes y los ítems de corrección deberían reflejar esta complejidad.

A partir de estas hipótesis generales intentamos mejorar la práctica docente y la evaluación de la expresión oral diagnosticando el problema mediante una explicación comprensiva de la situación actual. Para ello, intentaremos “obtener evidencias que sirvan como punto de partida y de comparación con las que se observan de los cambios o efectos del plan de acción” (Latorre, 2008: 43). Partimos del procedimiento llevado a cabo hasta ahora e intentamos mejorarlo. Nos centramos en elaborar criterios de evaluación objetivos que puedan ser aplicados a exámenes o a otras herramientas de evaluación, fijándonos fundamentalmente en los actuales exámenes de certificación de la Comunidad Autónoma Canaria y los criterios de evaluación que para ellas ha establecido la Consejería de Educación de Canarias. De esta forma, pretendemos implementar unos descriptores de evaluación más completos y rigurosos adecuados al contexto y al impacto de la evaluación que nos sirvan en nuestra práctica educativa para los exámenes de certificación, siendo extensibles, con algunas modificaciones, a otros instrumentos de evaluación tanto del proceso de aprendizaje de esta destreza como a los resultados finales. Se pretende investigar los efectos de estos descriptores, en relación con la viabilidad de las pruebas elaboradas, a través de una espiral de ciclos de investigación-acción puestos en marcha durante un curso escolar, para introducir mejoras.

Consideramos, así, citando las palabras de Yagüe (2010: 3-4),

que los exámenes -el diseño de los mismos- están lejos de los enfoques metodológicos actuales, lo cual genera una enorme distancia entre la realidad del aula, la docencia misma, y la medición de la competencia (...) los exámenes no parecen cumplir adecuadamente en todos los casos (o no con garantías suficientes) su cometido de medición de la competencia sincrónica de un aprendiz.

Esto lleva aparejado el que los ítem y descriptores no estén bien elaborados y que no respondan a los criterios de validez y fiabilidad mencionados anteriormente, por lo que creemos que los criterios de evaluación y escalas determinados por la Consejería de Educación son mejorables.

De esta manera, creemos que con esta investigación no sólo se puede mejorar la práctica educativa, sino también la actitud del alumno ante la evaluación de la expresión oral y el conocimiento de su dominio en este aspecto, ya que la utilización de descriptores le proporciona las pautas para determinar sus capacidades y así poder adscribirlas a un determinado nivel, repercutiendo, asimismo, en el mejor desarrollo de su proceso de enseñanza-aprendizaje, puesto que la evaluación también permite rendir cuentas de éste. Igualmente, y si se observan resultados favorables, el cambio introducido en la forma de evaluar podría ser asimilado e incorporado a la organización educativa del Centro donde se lleve a cabo la investigación.

1.2. Objetivos

1.2.1. Objetivos generales

- Conocer la actual validez y fiabilidad de los exámenes de certificación de Español en las EOI de la Comunidad Canaria.
- Determinar la influencia de la aplicación de criterios y pruebas válidos y fiables para establecer el dominio de nivel lingüístico en ELE.
- Proponer unos criterios de evaluación y descriptores bien desarrollados para reducir la variación e inconsistencia en la evaluación de la expresión oral.

1.2.2. Objetivos específicos.

- Proponer criterios de corrección que permitan hacer más válidos y fiables los actuales exámenes de Español de expresión oral de certificación aplicados a candidatos de nivel avanzado en inmersión lingüística de una EOI de Tenerife.
- Conocer cómo influye la correcta elaboración de criterios de corrección y de pruebas de evaluación de la expresión oral en la adecuada determinación del nivel de dominio avanzado de español y en la disminución de la subjetividad en la evaluación.
- Verificar si las herramientas propuestas son válidas y fiables, y permiten subsanar los problemas detectados en la evaluación de la expresión oral de alumnos de nivel avanzado en inmersión lingüística de las EOI de Canarias.

2. ESTADO DE LA CUESTIÓN

Teniendo en cuenta los objetivos perseguidos con este trabajo, el mismo puede enmarcarse dentro de las investigaciones relacionadas con la evaluación, en concreto con la evaluación de la expresión oral en exámenes de alto impacto como son las pruebas de certificación, que tienen por objeto reflexionar sobre el diseño, la elaboración y la corrección de herramientas de evaluación para garantizar la eficacia, calidad, fiabilidad, validez y viabilidad de estas. A continuación se exponen las reflexiones que se recogen en algunos de los estudios más recientes en torno a las dificultades que existen en relación con el diseño de estas pruebas y de los criterios de evaluación, así como algunos aspectos que se relacionan indirectamente con la evaluación pero que son indispensables a la hora de evaluar la actuación de un candidato.

El hecho de que en nuestro trabajo reiteremos con frecuencia determinados conceptos tiene la única finalidad de fijar su amplio valor en el terreno que nos ocupa (la objetivación de los descriptores de evaluación de la expresión oral), otorgándoles la profundidad nocional que su importancia, a nuestro juicio, merece. No obstante, antes de proseguir con la exposición es conveniente realizar algunas precisiones terminológicas.

2.1. Precisiones terminológicas

Seguidamente, vamos a definir algunos términos con los que vamos a trabajar a lo largo de nuestra investigación y que muchas veces se han llegado a confundir con otros que se han considerado similares. En primer lugar definiremos lo que entendemos por expresión oral dentro de nuestra investigación para luego poder detallar mejor las diferencias entre conceptos que, en ocasiones, su uso como semánticamente equivalentes es erróneo.

2.1.1. Expresión oral: definición del constructo

Para evaluar la expresión oral es necesario determinar qué entendemos por ésta, debido a que de esta definición depende tanto el enfoque que adoptemos para su aprendizaje y enseñanza como para su evaluación. La expresión oral se encuentra unida a la comprensión auditiva ya que en una conversación, por ejemplo, los usuarios de la lengua asumen los papeles de emisor y receptor teniendo que escuchar e interpretar lo

que dice su interlocutor (comprensión auditiva) para luego responderle (expresión oral), y junto a la comprensión y expresión escritas forman las cuatro destrezas lingüísticas básicas: hablar, escuchar, leer y escribir. Aparte de estas, el MCER va más allá y señala, además, la interacción y la mediación. La interacción se entiende como intercambio comunicativo entre dos o más participantes que se influyen mutuamente intercambiando acciones y reacciones verbales y no verbales, y la mediación se produce cuando el usuario se convierte en un canal de la transmisión de la información entre dos interlocutores que no pueden comunicarse directamente porque, por ejemplo, usan códigos distintos como cuando ocurre en una traducción.

Desde una concepción general, la expresión oral se identifica con hablar (MCER, 2002), pero esta identificación es necesario precizarla con mayor nitidez, puesto que una persona puede hablar sin sentido o fuera de contexto y, sin embargo, puede ser entendida, es decir, puede hablar cometiendo múltiples errores, lo que se conoce cotidianamente como “hablar mal” dentro del ámbito de un aprendiente de idiomas. Pero ¿expresarse correctamente de forma oral en otro idioma significa ser entendido por un interlocutor? Para muchos es así: no importa cometer errores si puede ser entendido. Si al entrar en una tienda decimos “yo querer manzanas”, seguramente nuestro interlocutor nos entiende e, incluso, se ha cumplido la intención comunicativa pero ¿eso significa que nos hemos expresado correctamente? Por nuestra experiencia hemos constatado que muchos profesores piensan así, y enseñan el idioma sin corregir a sus alumnos ya que consideran que si los corrigen no querrán hablar más, haciendo que a este alumnado se le pueda fosilizar el error siendo, más adelante, muy difícil de superar. Sobre el tratamiento del error hay muchas teorías en las que no nos vamos a detener ya que no es el objetivo de esta investigación, pero es un factor que influye en la enseñanza de la expresión oral y en su posterior evaluación.

Para otros profesores y/o evaluadores, si el candidato se expresa formalmente bien, su expresión oral es buena. Según este tipo de evaluadores la habilidad lingüística del candidato se mide según el mayor o menor número de errores gramaticales y léxicos detectados. De esta forma “se puede penalizar a candidatos con un mayor desarrollo de sus interlenguas frente a otros que sólo por una mayor corrección obtuvieran mejores notas” (Martínez, 1995: 366). Este tipo de evaluador parece anclado en las ideas estructuralistas de Saussure de *langue* y *parole*, o en las de Chomsky de *competence* y *performance*, de las que hablaremos posteriormente, en donde no se tenía en cuenta el contexto de uso de la lengua y en donde lo importante era el sistema, lo formal. En este

sentido, debemos advertir que aunque una persona se exprese correctamente en español desde el punto de vista formal, si el mensaje carece de sentido o de significado para su interlocutor, su producción oral no es válida, puesto que sólo utiliza correctamente el nivel formal, gramatical, o incluso léxico de la lengua pero su adecuación al contexto o su registro no son los correctos y son los que revisten enorme importancia para tener éxito en la comunicación oral.

Por lo tanto, siguiendo un enfoque comunicativo del estudio de la lengua, en el que nos detendremos más adelante, y que es el que nosotros seguiremos para esta investigación, el tener una correcta expresión oral significa no sólo saber expresarse oralmente bien en el idioma objeto de estudio, sino también que lo que se exprese tenga sentido, un significado para la persona que escuche o con la que se interactúe. Para ello, las actividades que utilicemos para la evaluación de la expresión oral tienen que ser lo que Littlewood (1998) denomina actividades de comunicación funcional donde se transmita el significado que se desea de un modo eficaz en una situación concreta. Pero, además, no sólo será importante que lo que se exprese tenga sentido en una situación determinada sino que sea, además, apropiado a la situación social en que se encuentra, comprobándose a través de tareas de interacción social donde los candidatos deben prestar especial atención a los significados funcionales y a los sociales que se transmiten mediante la lengua.

Así, por ejemplo, en un nivel A1 de español podemos proponer una tarea de evaluación oral pidiéndole a los candidatos que interactúen entre ellos presentándose los unos a los otros o, incluso, presentando a una tercera persona. Deben, así, elegir los recursos lingüísticos que tienen que utilizar, comunicándose de manera efectiva desde un punto de vista funcional; pero si además les indicamos que esas presentaciones se producen en el ámbito del trabajo, entonces también deberán adecuarse a las convenciones sociales que exige la comunicación. A la vista de esta concepción de la expresión oral se observa que su evaluación comprende no sólo la competencia lingüística del candidato sino también la sociolingüística y la pragmática, conformando, así, la competencia comunicativa de éste (Hymes, 1971; Canale y Swain, 1980; Widdowson, 1978; Bachman, 1990 y 1996; Vázquez, 2000; MCER, 2002).

En conclusión, el uso y aprendizaje de la lengua, y, por consiguiente, esto abarca a la expresión oral, se podría definir como recoge el MCER (2002: 9):

El uso de la lengua –que incluye el aprendizaje– comprende las acciones que realizan las personas que, como individuos y como agentes sociales, desarrollan una serie de competencias, tanto generales como competencias comunicativas lingüísticas, en

particular. Las personas utilizan las competencias que se encuentran a su disposición en distintos contextos y bajo distintas condiciones y restricciones, con el fin de realizar actividades de la lengua que conllevan procesos para producir y recibir textos relacionados con temas en ámbitos específicos, poniendo en juego las estrategias que parecen más apropiadas para llevar a cabo las tareas que han de realizar. El control que de estas acciones tienen los participantes produce el refuerzo o la modificación de sus competencias.

2.1.2. “Evaluar” versus “examinar”

Todos los expertos coinciden en que la evaluación es indispensable dentro de la enseñanza de idiomas, pero en ocasiones se confunden los términos “evaluar” y “examinar”.

Durante nuestra vida está presente la evaluación, no sólo cuando tenemos que realizar un examen sino también cuando desempeñamos una actividad, cuando tomamos una decisión o cuando elegimos la casa que queremos comprar. Desde un punto de vista general, la evaluación nos sirve para “analizar una serie de datos para, en consecuencia, adoptar la decisión adecuada”. (Bordón, 2006:18). Concretando esta definición en la enseñanza de idiomas la evaluación es la “valoración del grado de dominio lingüístico que tiene el usuario” (MCER, 2002: 177). Podemos decir que es necesaria desde el punto de vista práctico, porque “nos servimos de ella para tomar decisiones que resulten acertadas”, además de constituir “un incentivo para el estudiante, o la manera más “justa” para valorar sus conocimientos cuando hay que medirlos con los de otros” (Bordón, 2006: 57); y, desde el punto de vista didáctico, ya que forma parte del currículo de la enseñanza de L2 como herramienta para observar el proceso de aprendizaje del alumno donde no sólo se tienen que determinar los contenidos, objetivos, metodología y estrategias de enseñanza, sino también los criterios de evaluación que son los que determinan el éxito de la enseñanza. Igualmente, la evaluación resulta muy importante en la investigación sobre la adquisición de segundas lenguas y en el análisis de Interlengua.

Sin embargo, examinar es la manera en la que obtenemos información suficiente para realizar nuestra evaluación. Generalmente, obtenemos esta información mediante exámenes, aunque estos no son los únicos instrumentos, ya que también podemos recabar esta información mediante la observación en clase o la resolución de tareas, por ejemplo. El examen “es un instrumento que sirve para recoger evidencia suficiente que permite tomar decisiones acerca del dominio de lengua del evaluado” (Prati, 2007: 12). Con la información obtenida de los instrumentos de evaluación examinamos, es decir,

podemos obtener unos datos que podemos cuantificar y computar haciendo que la toma de decisiones sea más objetiva. No obstante, para que esta objetividad se produzca es necesario que nos aseguremos de que estos instrumentos de evaluación que utilizamos para luego evaluar (exámenes, notas de observación en clase, tareas, etc.) sean fiables y válidos con el objeto de que nuestras decisiones sean acertadas.

Prati (2007: 12) centrándose en el concepto de examen señala que según el punto de vista con el que se mire, el examen puede percibirse de distinta manera. Recoge estos puntos de vista diferentes en el siguiente cuadro sobre las representaciones de la palabra examen:

Cuadro 1.

para los estudiantes	<i>prueba</i> de dos sentidos: (1) pone a prueba las habilidades lingüísticas (2) da testimonio	(1) porque “ensayan” su capacidad en acciones que usarán en el futuro (laboral, académico, personal) (2) porque “demuestran” a todos lo que han aprendido.
para los docentes	<i>evaluación</i>	Porque recoge evidencia para <i>tomar decisiones</i> . Dicha evidencia puede tener efectos directos sobre la enseñanza y poner en marcha la investigación.
para la sociedad	<i>certificación</i>	Como documento sellado, requerido como habilitación para accionar en el mundo real, que da además prestigio y presencia institucional.

Como se puede ver, para los profesores un examen es evaluación, confundiendo los dos términos, porque consideran que el fin último de los instrumentos de evaluación que se

utilizan para examinar es evaluar, es decir, otorgar una valoración a partir de la actuación de la persona evaluada mediante la interpretación de los datos que proporcionan dichos instrumentos de evaluación. De ahí que se identifique evaluación con examen aunque no sean exactamente lo mismo. Además, en la evaluación se tiende a “generalizar a partir de una situación observada sobre otras situaciones no observadas” (McNamara, 1996, citado en Figueras, 2004: 67), es decir, aunque solamente examinemos a un candidato de un aspecto determinado dentro de los conocimientos y habilidades que debería mostrar en un acto de habla, la actuación producida se generaliza a la totalidad de los conocimientos y habilidades de dicha destreza.

De esta manera, los exámenes o el procedimiento de examinar se encuentra dentro de la evaluación o el procedimiento de evaluar, es decir, el concepto de evaluar engloba el concepto de examinar, ya que se puede evaluar mediante distintos procedimientos (un examen, datos de clase, etc.).

En nuestra investigación nos centraremos en la evaluación del dominio lingüístico de la expresión oral mediante las tareas facilitadas en los exámenes de certificación de la Comunidad Autónoma Canaria.

2.1.3. “Criterios de evaluación” versus “criterios de calificación”

Se puede decir que el concepto “criterio de evaluación”, al igual que el de “evaluar o evaluación” hacía con el de “examinar o examen”, engloba el de “criterios de calificación.

Los criterios de evaluación son las consideraciones que debe tener en cuenta el evaluador para determinar la actuación de un candidato y le permiten decidir si esa actuación es positiva o no y en qué medida. Los criterios de evaluación permiten determinar “cómo se interpreta la actuación” (MCER, 2002: 178), estos criterios determinan la consecución de un objetivo de aprendizaje; sin embargo, los criterios de calificación miden esa actuación en una prueba concreta. Por lo tanto, el concepto de criterios de evaluación es más amplio que el de calificación. Si queremos concretar aún más los criterios de calificación, podemos recurrir a los descriptores de calificación que determinan de una forma más detallada la medida de la actuación del candidato. Así, por ejemplo, un criterio de evaluación de la expresión oral de un alumno de B2 puede ser “saber argumentar de manera coherente y fluida con su interlocutor una idea defendida por este sobre un tema de actualidad”; los criterios de calificación pueden ser

medir la actuación del candidato determinada por ese criterio de evaluación, teniendo en cuenta las distintas competencias (lingüística, sociolingüística y pragmática), y, a su vez, los descriptores de calificación permitirían medir cada una de esas competencias de manera más detallada.

Por ejemplo, en el criterio de calificación de la gramática que se encuentra ubicado en la competencia lingüística, un descriptor podría ser “utiliza correctamente una gran variedad de estructuras y formas gramaticales propias del nivel” (Boletín Oficial de Canarias (BOC), nº 167. Resolución de 28 de julio de 2010). Es decir, partimos del concepto general de evaluación y, a partir de ahí, se va concretando cada vez más con el fin de que el evaluador pueda tomar una decisión adecuada, lo más objetiva posible, en relación con el resultado del candidato.

Pretendemos, en esta investigación, hacer más objetivos estos descriptores de calificación de los exámenes de certificación de las EOI de la Comunidad Autónoma Canaria, demostrando que en la actualidad no cumplen los requisitos de validez, fiabilidad y viabilidad.

2.1.4. “Descriptores de capacidad lingüística” versus “escalas de niveles”

Cuando establecemos criterios para el logro de los objetivos de aprendizaje necesitamos utilizar descriptores que contribuyan a la formulación de estos criterios. Los descriptores se utilizan como criterios de evaluación si se presentan como “una escala combinando descriptores de distintas categorías en un párrafo holístico por nivel” (MCER, 2002: 181).

Los descriptores de capacidad lingüística se pueden identificar con lo que el MCER (2002: 181) llama descriptores del dominio lingüístico referido a competencias concretas (gramatical, léxica, semántica, ortográfica y ortoépica). El progreso en el desarrollo de la capacidad que tiene un alumno de utilizar los recursos lingüísticos se puede clasificar por escalas de nivel dentro de cada competencia. Así, las escalas de niveles determinan lo que un alumno o candidato debe saber dentro de ese nivel y en relación con la competencia a la que pertenezca dicha escala.

La capacidad lingüística de un alumno en los exámenes de certificación tienen que estar relacionadas con un nivel, ya que un alumno se examina del idioma en función a ese nivel: de ahí que haya exámenes de certificación para nivel Básico, Intermedio o Avanzado. Las escalas de niveles nos sirven para determinar este nivel de una manera general y de una manera más específica en relación con cada competencia concreta. Así,

un descriptor de capacidad lingüística se puede concretar en una escala de nivel, por ejemplo, de la siguiente forma:

Cuadro 2.

COMPETENCIA LINGÜÍSTICA GENERAL	
C2	Saca provecho de un dominio amplio y fiable de un completo repertorio de elementos lingüísticos para formular pensamientos con precisión, poner énfasis, diferenciar y eliminar la ambigüedad. No manifiesta ninguna limitación de lo que quiere decir.
C1	Elige una formulación apropiada de una amplia serie de elementos lingüísticos para expresarse con claridad y sin tener que limitar lo que quiere decir.
B2	Se expresa con claridad y sin manifestar ostensiblemente que tenga que limitar lo que quiere decir. Dispone de suficientes elementos lingüísticos como para hacer descripciones claras, expresar puntos de vista y desarrollar argumentos, utilizando para ello algunas oraciones complejas y sin que se le note mucho que está buscando las palabras que necesita.
B1	Dispone de suficientes elementos lingüísticos como para describir situaciones impredecibles, para explicar los puntos principales de una idea o un problema con razonable precisión y para expresar pensamientos sobre temas abstractos o culturales, tales como la música y las películas.
	Dispone de suficientes elementos lingüísticos como para desenvolverse y de suficiente vocabulario como para expresarse con algunas dudas y circunloquios sobre temas como la familia, aficiones e intereses, trabajo, viajes y hechos de actualidad, pero las limitaciones léxicas provocan repeticiones e incluso, a veces, dificultades en la formulación.
A2	Tiene un repertorio de elementos lingüísticos básicos que le permite abordar situaciones cotidianas de contenido predecible, aunque generalmente tiene que adaptar el mensaje y buscar palabras.
	Produce expresiones breves y habituales con el fin de satisfacer necesidades sencillas y concretas: datos personales, acciones habituales, carencias y necesidades, demandas de información. Utiliza estructuras sintácticas básicas y se comunica mediante frases, grupos de palabras y fórmulas memorizadas al referirse a sí mismo y a otras personas, a lo que hace, a los lugares, a las posesiones, etc. Tiene un repertorio limitado de frases cortas memorizadas que incluye situaciones predecibles de supervivencia; suelen incurrir en malentendidos e interrupciones si se trata de una situación de comunicación poco frecuente.
A1	Tiene un repertorio muy básico de expresiones sencillas relativas a datos personales y a necesidades de tipo concreto.

(MCER, 2002: 107)

Estos descriptores se refieren a la capacidad lingüística general concretada en una escala de nivel. Sin embargo, dentro de la capacidad lingüística general podemos establecer, también divididos en escalas de nivel, descriptores de dominio de cada una de las capacidades que conforman la capacidad lingüística de un candidato. Así, por ejemplo, podemos establecer descriptores de dominio de la competencia léxica dependiendo del nivel, competencia que pertenece, a su vez, a la competencia lingüística; quedaría de la siguiente manera:

Cuadro 3.

RIQUEZA DE VOCABULARIO	
C2	Tiene un buen dominio de un repertorio léxico muy amplio, que incluye expresiones idiomáticas y coloquiales; muestra que es capaz de apreciar los niveles connotativos del significado.
C1	Tiene un buen dominio de un amplio repertorio léxico que le permite superar con soltura sus deficiencias mediante circunloquios; apenas se le nota que busca expresiones o que utiliza estrategias de evitación. Buen dominio de expresiones idiomáticas y coloquiales.
B2	Dispone de un amplio vocabulario sobre asuntos relativos a su especialidad y sobre temas más generales. Varía la formulación para evitar la frecuente repetición, pero las deficiencias léxicas todavía pueden provocar vacilación y circunloquios.
B1	Tiene suficiente vocabulario para expresarse con algún circunloquio sobre la mayoría de los temas pertinentes para su vida diaria, como, por ejemplo, familia, aficiones e intereses, trabajo, viajes y hechos de actualidad.
A2	Tiene suficiente vocabulario para desenvolverse en actividades habituales y en transacciones cotidianas que comprenden situaciones y temas conocidos. Tiene suficiente vocabulario para expresar necesidades comunicativas básicas. Tiene suficiente vocabulario para satisfacer necesidades sencillas de supervivencia.
A1	Tiene un repertorio básico de palabras y frases aisladas relativas a situaciones concretas.

DOMINIO DE VOCABULARIO	
C2	Utiliza con consistencia un vocabulario correcto y apropiado.
C1	Pequeños y esporádicos deslices, pero sin errores importantes de vocabulario.
B2	Su precisión léxica es generalmente alta, aunque tenga alguna confusión o cometa alguna incorrección al seleccionar las palabras, sin que ello obstaculice la comunicación.
B1	Manifiesta un buen dominio del vocabulario elemental, pero todavía comete errores importantes cuando expresa pensamientos más complejos, o cuando aborda temas y situaciones poco frecuentes.
A2	Domina un limitado repertorio relativo a necesidades concretas y cotidianas.
A1	No hay descriptor disponible.

(MCER, 2002: 109)

Como podemos comprobar, se ha creado un descriptor de capacidad lingüística centrado en la competencia léxica y se ha delimitado aún más en dos descriptores diferentes

como son el de la riqueza léxica y el de dominio del vocabulario, que, a su vez, se han concretado más dentro de una escala de nivel.

Estos descriptores son de gran ayuda para el evaluador y se pueden mejorar no sólo determinando lo que el candidato debe hacer sino también en qué medida lo hace bien o mal. Así, cada descriptor por competencia y nivel se puede subdividir en otros descriptores más pequeños que determinen la medida del descriptor cuantificándola en función de si es excelente, adecuada, suficiente, limitada o inadecuada y pueden también tomarse como criterios de calificación.

Así, por ejemplo, dentro de una tarea expositiva de expresión oral (monologada) de un examen de certificación podemos determinar para el léxico lo siguiente:

Cuadro 4.

Puntuación de léxico	LÉXICO
8	Utiliza un vocabulario, muy variado y preciso, adecuado a la tarea.
6	Utiliza un vocabulario adecuado a la tarea, variado y preciso.
4	Utiliza un vocabulario adecuado a la tarea, y suficiente.
2	Muestra un uso del vocabulario poco adecuado a la tarea, y limitado.
0	Utiliza un vocabulario inadecuado y muy limitado para la tarea.

(Boletín Oficial de Canarias (BOC), nº 167: 22223. Resolución de 28 de julio de 2010, por la que se dictan instrucciones para la elaboración de las pruebas comunes de obtención de los certificados de enseñanzas de idiomas de régimen especial, en la Comunidad Autónoma de Canarias).

El problema estriba, como ya he dicho anteriormente, en que no todos los evaluadores tienen un mismo criterio para considerar algo excelente, adecuado, suficiente, limitado o inadecuado, por lo que las puntuaciones suelen variar de un evaluador a otro. De ahí que los descriptores deban ser lo más exactos y claros posibles para que no hayan dobles interpretaciones y se pueda obtener la fiabilidad de los resultados.

De todas maneras, tenemos que tener en cuenta que el dominio de la capacidad lingüística del alumno es sólo uno de los aspectos que demuestran su competencia comunicativa o lo que Bachman (1990) llamó “habilidad lingüística comunicativa”, ya que la competencia comunicativa comprende no sólo la lingüística sino también la sociocultural y la pragmática (MCER, 2002: 106), como veremos en el siguiente punto.

2.1.5. Competencia comunicativa: conocimiento y uso de la lengua

Es necesario disponer de una concepción de la naturaleza de la lengua para poder elaborar y evaluar una prueba de lengua. Como dice Spolsky (1989: 132): “uno no puede desarrollar una buena prueba de lengua sin tener un método para definir lo que significa el saber una lengua, ya que hasta que no se ha decidido lo que se está midiendo no se puede afirmar haberlo medido”.

En la enseñanza de idiomas, el profesor hace una evaluación del nivel de lengua objeto de estudio por parte del alumno, que en el caso que nos ocupa es el español, pero ¿qué modelo de lengua vamos a evaluar? La respuesta a esta pregunta la debe tener muy clara el examinador-evaluador, ya que, si no, los resultados de la evaluación carecerían de validez.

Este aspecto ha sido de tal importancia que durante los años han surgido un gran número de teorías lingüísticas a este respecto y aunque éstas son ampliamente conocidas, creemos con sumo interés citarlas en nuestro estudio ya que el evaluador debe tener un conocimiento claro de ellas para poder saber qué es lo que va a evaluar y qué elementos va a analizar en las producciones orales siguiendo las escalas de evaluación. El proceso de construcción de pruebas de evaluación debe observarse desde su comienzo y se debe considerar principalmente la lengua que se quiere evaluar. Se pretende “*an emphasis on the centrality of constructs – what we are trying to measure – requires testers to consider what is known about language knowledge and ability, and ability to use the language. Language testing involves not only the psychometric and technical skills required to construct and analyse a test but also knowledge about language: testers need to be applied linguists, aware of the latest and most accepted theories of language description, of language acquisition and language use*” (Alderson *et al.*, 2002: 80). De esta forma, en nuestra investigación las teorías sobre la competencia comunicativa, fundamentalmente las de Bachman y Palmer (1996), son esenciales para la validación de las pruebas a través del concepto de lengua utilizada en su constructo. Para entender la evaluación actual de la lengua y la actitud de algunos profesores-examinadores ante ella debemos comprender cuál ha sido la evolución de este concepto hasta ahora, y es por ello por lo que expondremos brevemente cómo se ha entendido la lengua hasta ahora para su enseñanza y el modelo o modelos que se siguen en esta investigación y que suelen primar actualmente en la enseñanza y evaluación de la L2. Se pretende ofrecer con ello una visión general que explique la actitud y práctica de algunos examinadores ante la evaluación y la

aplicación del modelo elegido para esta en este estudio empírico.

Los postulados de la lengua como sistema tienen su origen en el *Curso de Lingüística general* de Saussure (1993) quien distinguía entre lengua (*langue*) y habla (*parole*), surgiendo, así, la dicotomía lengua-habla propia del estructuralismo europeo, donde la lengua se refiere al sistema lingüístico, el conjunto de unidades y reglas del sistema que es común a todos los individuos de una comunidad lingüística. Según Saussure (1993: 79) “la lengua es la parte social del lenguaje, exterior al individuo, que por sí sólo no puede crearla ni modificarla, no existe más que en virtud de una especie de contrato establecido entre los miembros de la comunidad”. Sin embargo, el habla es individual; en el acto del habla de una persona, por ejemplo, intervienen los actos de fonación necesarios para ejecutar las combinaciones de la lengua.

Así, durante los dos primeros tercios del siglo XX el estructuralismo fue el movimiento predominante en la lingüística y, por extensión, en la enseñanza, elaborándose exámenes en donde lo importante era, sobre todo, la competencia lingüística, es decir, la gramática, el vocabulario y la pronunciación. De ahí que en las primeras pruebas de evaluación de la expresión oral se evaluara fundamentalmente el componente gramatical y léxico y, por consiguiente, la actitud de algunos profesores veteranos en la enseñanza de querer evaluar aún bajo este sistema.

La aplicación de los principios del estructuralismo a la didáctica de la L2 fue un gran avance para adoptar otra metodología en la enseñanza-aprendizaje del idioma y superar el enfoque de gramática-traducción. Su aplicación más importante se encuentra en el método audiolingüe, basado en la audición y en la repetición, que se encuentra en consonancia con el conductismo, en donde el aprendizaje se adquiere mediante hábitos y por medio de la repetición. Sin embargo, la concepción del lenguaje de Saussure considerado como un sistema autónomo de signos constituido por elementos interrelacionados es lo que diferencia al estructuralismo europeo, con su carácter mentalista sobre la concepción saussureana del lenguaje, del estructuralismo norteamericano, cuyo padre es Leonard Bloomfield, que es de carácter conductista y observacional, ya que para estos últimos recurrir a la mente estaba fuera de todo rigor científico.

La concepción de la lengua como sistema, a pesar de haber sido ya superada, todavía hoy se encuentra en algunos ejercicios, libros de texto e incluso tareas de examen.

Posteriormente, se introduce la idea de que el ser humano puede reproducir secuencias lingüísticas que nunca antes había escuchado (Chomsky, 1965) siguiendo este carácter

mentalista iniciado por Saussure. Todas las personas tenemos una Gramática Universal debido a la que poseemos competencia (*competence*), una gramática interior de nuestra lengua, y actuación (*performance*) que se corresponde con el uso de la lengua afectado por nociones irrelevantes como falta de memoria o distracciones, por ejemplo, pero consideradas aceptables. Parece que la dicotomía de lengua y habla de Saussure es similar a la de competencia y actuación de Chomsky, sin embargo para Saussure la lengua tiene un carácter social, mientras que para Chomsky es psicolingüístico y, por otra parte, el estructuralismo se centraba en la elaboración de clasificaciones, repertorios, como un inventario de formas lingüísticas. No obstante, para los generativistas, como es el caso de Chomsky, se abre una investigación más dinámica y no tan estática, ya que se observa la competencia como un conjunto de reglas capaz de explicar la creatividad lingüística. Se pretende, así, explicar la competencia lingüística o competencia gramatical, frente a la descripción lingüística propia del paradigma estructuralista.

En los años 70 se ponen en entredicho estas teorías, ya que al observar la lengua en uso es muy difícil poder explicar su correcta producción solamente mediante el conocimiento de la gramática. Con Hymes (1971) se introduce el componente sociolingüístico, es decir, se debe tener en cuenta el contexto de uso de la lengua ya que debe estudiarse dentro de una comunidad de habla. Propone, así, el concepto de competencia comunicativa donde intervienen no sólo el aspecto lingüístico, gramatical o aceptable de la lengua (Chomsky), sino que también hay que tener en cuenta el uso de la lengua en la producción de enunciados adecuados en el contexto discursivo. Así, como advierte Widdowson (1989: 84) la competencia de Chomsky no tiene nada que ver con la de Hymes, ambos “están jugando diferentes tipos de juego”, a Chomsky no le interesa el uso del lenguaje sino su gramática, para él:

la competencia es algo mucho más abstracto que el lenguaje; es el conocimiento de un sistema de reglas, de parámetros o principios, configuraciones en la mente para las que el lenguaje simplemente sirve de evidencia (...) Para Chomsky, entonces, la competencia es el conocimiento gramatical como un arraigado nivel mental por debajo del de lengua (...). Para Hymes, por otro lado, la competencia es la habilidad para hacer algo: para usar la lengua (...). Lo que Chomsky elimina de su definición con tanto cuidado Hymes lo restablece. (*Ibidem*)

Por lo tanto, hay que atender en el estudio de la lengua, además de a lo gramatical, que para Hymes (1971: 40) es lo posible, a lo factible, es decir, el conocimiento psicológico y su uso (memoria, mecanismos de percepción, etc.), que es lo que está dentro de lo aceptable para Chomsky, la actuación; a lo apropiado según las características

contextuales y, por último, a lo que se da en la realidad, a lo hecho, haciendo referencia a la convención y al hábito. Sin embargo, “puede que algo resulte posible, factible, apropiado, y que no llegue a ocurrir” (*ibidem*: 41), o mediante un ejemplo más clarificador: “una frase puede ser gramaticalmente correcta, estilísticamente torpe, socialmente diplomática y de uso poco frecuente” (*ibidem*: 37).

Estos cuatro sectores pueden ser analizados en ocho componentes del acto del habla o acontecimientos contenidos en la palabra inglesa SPEAKING (Hymes 1974: 55-61) :

- S (*setting y scene*) referido a la situación del acto del habla, tanto física (*setting*) como psicológica (*scene*), esta última implica aspectos culturales.
- P (*participant*) en lugar de hablante-oyente, puesto que a veces en un acto de habla se requiere de una audiencia o de un solo emisor.
- E (*ends*): metas y resultados.
- A (*act-sequence*): forma y contenido del mensaje.
- K (*key*): la clave referida al tono y la forma del acto de habla.
- I (*instrumentalities*): el canal por el que se transmite la información (escrito, oral, variedades dialectales o estilos)
- N (*norms*): las normas de interacción y de interpretación.
- G (*genres*): tipos, categorías, características de lengua. Se considera que mediante el género se reconocen características formales como, por ejemplo, el tipo de discurso.

De esta manera, Hymes introduce la dimensión pragmática en el uso del lenguaje, en la que los aspectos socioculturales resultan de determinados actos comunicativos, pues se considera que todo acto de habla se desarrolla en un contexto sociocultural y tiene, además, una función comunicativa.

En 1978, H. G. Widdowson, siguiendo una idea similar a la de Hymes, considera que se deben tener en cuenta las reglas de uso en situaciones sociales particulares, que según su criterio es la competencia comunicativa, diferente a la competencia gramatical pero en la que ambas forman parte de la competencia del hablante. De esta manera, establece la dicotomía entre sistema lingüístico (*usage*) y el uso o “valor en la lengua” (*use*) considerando que la competencia comunicativa se debe entender más como uso (*use*) que como sistema (*usage*). El uso demuestra el conocimiento de las reglas lingüísticas y está muy relacionado con la competencia de la que hablaba Chomsky, “*that which makes evident the extent to which the language user demonstrates his ability to use his knowledge of linguistic rules for effective communication*” (Widdowson, 1978: 3).

Por lo tanto, Widdowson sostiene que su distinción entre sistema (*usage*) y uso (*use*) está muy relacionada con la dicotomía presentada por Saussure entre *langue* y *parole* y, a su vez, con la distinción entre *competence* y *performance* de Chomsky (*ibidem*).

Posteriormente, en 1980, Canale y Swain amplían la noción de competencia comunicativa de Hymes y señalan que ésta, a su vez, está conformada por otras subcompetencias³ (Canale y Swain, 1980: 29-31):

- Competencia gramatical: referida al código lingüístico, en el que se incluyen el vocabulario, la formación de palabras y frases, la pronunciación, la ortografía y la semántica.
- Competencia sociolingüística: el contexto y las reglas socioculturales así como en qué medida las expresiones son producidas y entendidas adecuadamente en dicho contexto.
- Competencia discursiva: relacionada con los elementos de cohesión y las reglas de coherencia.
- Competencia estratégica: estrategias verbales y no verbales que se utilizan para suplir fallos en la comunicación.

Por lo tanto, “la competencia comunicativa se refiere tanto al conocimiento como a la habilidad para utilizar este conocimiento cuando se participa en una comunicación real” (Canale, 1983: 66), entendiéndose como habilidad a la utilización, correcta o no, de estos conocimientos en la comunicación real.

Además, Canale también distingue entre la enseñanza y evaluación orientadas al conocimiento y las orientadas a la habilidad. Hasta entonces se pensaba que si se tenían buenos conocimientos en una L2, se podía utilizar esa L2 sin problemas. Sin embargo, Canale (1983) considera que el conocimiento y la habilidad de una lengua están relacionados. Así, nos aporta el ejemplo de un programa de autoescuela que si se basa en los conocimientos y no en la práctica, no sabríamos cómo actuarían los aprobados en una situación real conduciendo un coche, lo cual sería totalmente inadecuado y peligroso. Lo mismo ocurriría si se ignorara el aspecto teórico u orientado al conocimiento, ya que si un aprendiente en la práctica (habilidad) comete una imprudencia o fallo, se podría diagnosticar, por ejemplo, a partir de su falta de conocimiento. Por ello, Canale (1983) considera que tanto la enseñanza y evaluación

³ Entiéndase como cada uno de los aspectos que componen una competencia y a la que le corresponden una serie de descriptores en la escala de evaluación.

orientada al conocimiento como orientada a la habilidad no se pueden dar de una manera separada, dando lugar a un enfoque integrador, comunicativo.

Siguiendo este modelo de competencia comunicativa de Canale y Swain (1980) y Canale (1983), Bachman (1990) propone su modelo de habilidad lingüística comunicativa que tiene los mismos componentes que el modelo de Canale y Swain pero organizados de distinta manera. Su modelo de habilidad lingüística comunicativa ha repercutido muchísimo en la manera de enseñar y evaluar una lengua. Bachman distingue dos áreas en la competencia lingüística del hablante: por un lado el conocimiento organizativo que a su vez se subdivide en el conocimiento gramatical (léxico, morfológico, sintáctico, fonológico y ortográfico) y el conocimiento textual (elementos de cohesión y organización retórica escrita y oral), y, por otro lado, el conocimiento pragmático referido al uso de la lengua, relación de oraciones, textos e intenciones, y compuesto, a su vez, por el conocimiento funcional, llamado en 1990 por Bachman “competencia ilocutiva”, que comprende una serie de funciones divididas en ideales, manipulativas, heurísticas e imaginativas, y por el conocimiento sociolingüístico, que hace referencia a los dialectos, registros, expresiones idiomáticas, referencias culturales y figuras del habla. Todo esto se puede observar de una forma más clara en el siguiente esquema que aportan Bachman y Palmer (1996: 68) sobre las áreas del conocimiento lingüístico (el subrayado es nuestro):

Cuadro 5.

Organizational knowledge

(how utterances or sentences and texts are organized)

Grammatical knowledge

(how individual utterances or sentences are organized)

Knowledge of vocabulary

Knowledge of syntax

Knowledge of phonology/graphology

Textual knowledge

(how utterances or sentences are organized to form texts)

Knowledge of cohesion

Knowledge of rhetorical or conversational organization

Pragmatic knowledge

(how utterances or sentences and texts are related to the communicative goals of the language user and to the features of the language use setting)

Functional knowledge

(how utterances or sentences and texts are related to the communicative goals of

language users)
Knowledge of ideational functions
Knowledge of manipulative functions
Knowledge of heuristic functions
Knowledge of imaginative functions

Sociolinguistic knowledge

(how utterances or sentences and texts are related to features of the language use setting) Knowledge of dialects/varieties
Knowledge of registers
Knowledge of natural or idiomatic expressions
Knowledge of cultural references and figures of speech.

La diferencia respecto al modelo de Canale y Swain estriba en que no incluye la competencia estratégica entre los constituyentes de la competencia lingüística pero sí la considera parte de la habilidad lingüística comunicativa. Para Bachman, la competencia estratégica no sirve para remediar fallos en la comunicación, como señalaban Canale y Swain, sino que se usa para llevar a cabo funciones de planificación, ejecución y valoración, explicando, de esta forma, cómo interaccionan los componentes de la competencia lingüística (gramatical, textual, funcional y sociolingüístico). Como se puede comprobar, Bachman amplía el modelo de Canale y Swain. El papel de la competencia estratégica fue ampliado, a su vez, posteriormente por Bachman y Palmer (1996: 70-75) concibiéndola como una serie de componentes o estrategias metacognitivas que operan en tres áreas: establecimiento de metas (decidir qué se va a hacer), valoración (establecer qué se necesita, con qué se ha de trabajar y cómo se ha hecho) y planificación (decidir cómo usar lo que se tiene).

Acercándonos a nuestro tiempo, el siglo XXI, el MCER (2002: 99-120) advierte que las competencias del usuario o alumno se pueden considerar aspectos de la competencia comunicativa en donde no sólo se incluyen las competencias relacionadas con la lengua sino también otras competencias generales. Por ello, el uso y aprendizaje de la lengua se podría definir como recoge el MCER (2002: 9):

El uso de la lengua –que incluye el aprendizaje– comprende las acciones que realizan las personas que, como individuos y como agentes sociales, desarrollan una serie de **competencias**, tanto **generales** como **competencias comunicativas lingüísticas**, en particular. Las personas utilizan las competencias que se encuentran a su disposición en distintos **contextos** y bajo distintas **condiciones** y **restricciones**, con el fin de realizar **actividades de la lengua** que conllevan **procesos** para producir y recibir **textos** relacionados con **temas** en **ámbitos** específicos, poniendo en juego las **estrategias** que parecen más apropiadas para llevar a cabo las **tareas** que han de realizar. El control que de estas acciones tienen los participantes produce el refuerzo o la modificación de sus competencias.

En todas estas competencias se deben tener en cuenta tanto los conocimientos como las destrezas y habilidades de cada una de ellas. Así, en el MCER se deja claro que las competencias del usuario o alumno están formadas por:

1. Competencias generales en donde se incluyen :

a) El conocimiento declarativo (saber) compuesto por:

- El conocimiento del mundo, que se encuentra relacionado con el vocabulario y la gramática de la lengua materna. Aquí se debe tener especial relevancia en la enseñanza de idiomas en determinar el conocimiento del mundo respecto al país o países en los que se habla la lengua objeto de estudio.
- El conocimiento sociocultural: “el conocimiento de la sociedad y de la cultura de la comunidad o comunidades en las que se habla el idioma, es un aspecto del conocimiento del mundo” (MCER, 2002: 100). Se deben tener en cuenta las características de la sociedad europea en su vida diaria, condiciones de vida, relaciones personales, valores, creencias y actitudes respecto a distintas cuestiones, el lenguaje corporal, las convenciones sociales o el comportamiento ritual en determinadas áreas como, por ejemplo, en ceremonias y prácticas religiosas.
- La conciencia intercultural: hay que prestar atención a las diferencias y similitudes entre el país de origen y el extranjero teniendo en cuenta, por ejemplo, estereotipos o tópicos nacionales.

b) Competencia “existencial” (saber ser): relacionada con factores individuales (actitudes, motivaciones, valores, creencias, estilos cognitivos y tipo de personalidad).

c) Capacidad de aprender (saber aprender): se desarrolla en el curso de su propio aprendizaje. Aquí se incluyen la reflexión sobre el sistema de la lengua y la comunicación y reflexión sobre el sistema fonético y las destrezas correspondientes (destrezas de estudio, descubrimiento y análisis).

2. Competencia comunicativa:

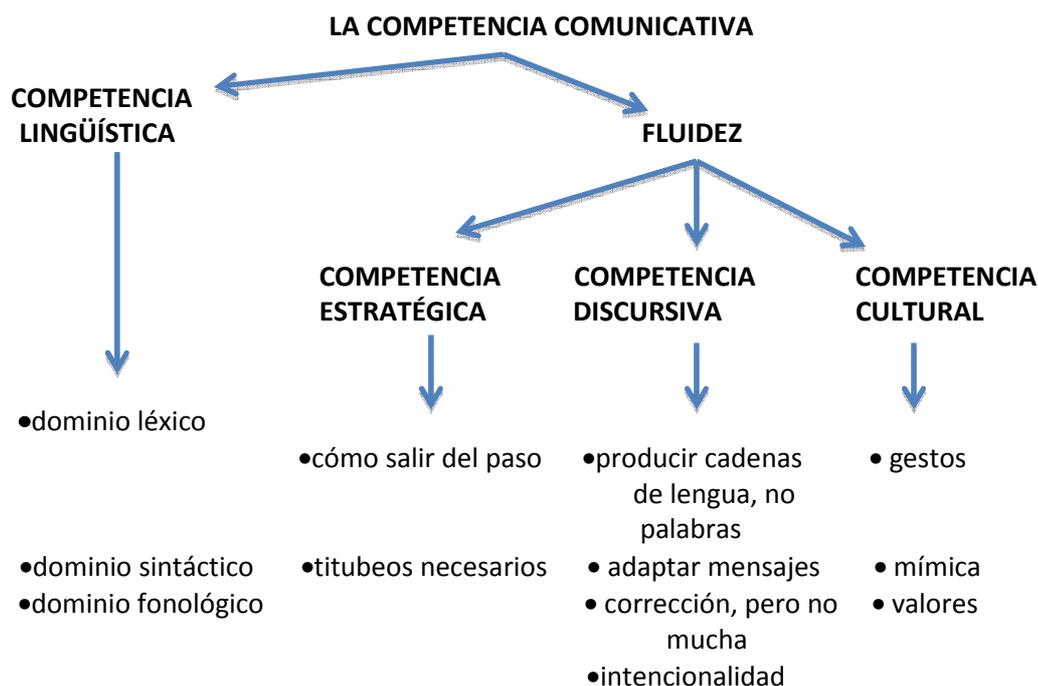
a) Competencia lingüística: compuesta por el conocimiento de recursos formales y su capacidad para usarlos. Incluye las competencias léxica, gramatical, semántica, fonológica, ortográfica y ortoépica.

b) Competencia sociolingüística: que tiene en cuenta los marcadores lingüísticos de relaciones sociales, normas de cortesía, expresiones de la sabiduría popular, las diferencias de registro, el dialecto y el acento.

c) Competencia pragmática: que incluye las competencias discursiva, funcional y organizativa.

Si todo esto lo aplicamos a la expresión oral podemos considerar que la competencia comunicativa en el habla comprende lo siguiente (Vázquez, 2000: 17):

Cuadro 6.



En la actualidad, las enseñanzas de idiomas se articulan siguiendo las directrices del MCER que a su vez han sido elaboradas teniendo en cuenta las teorías lingüísticas mencionadas anteriormente. En este sentido, cabe señalar que los profesores de las EOI de Canarias siguen también los modelos teóricos aquí expuestos, entendiéndose no cada uno en su totalidad sino que se ha escogido de ellos lo que se adecua más a la enseñanza de una L2 basándose en la experiencia y en el contexto de actuación.

Ineludiblemente, de manera más relevante se han tenido como referente más cercano las últimas teorías lingüísticas aplicadas a la pedagogía como son las de Canale (1983), Bachman (1990), Bachman y Palmer (1996), y el MCER (2002), y en el caso concreto del español, el Plan curricular del Instituto Cervantes (PCIC, 2007). A la vista de estas teorías, está demostrado que el modelo de lengua a enseñar es el que conjuga los conocimientos y las habilidades siguiendo el enfoque comunicativo defendido por estos lingüistas y comenzando por Hymes en 1966 con su modelo de competencia

comunicativa, concepto de vital importancia en la enseñanza de lenguas, ya que no sólo debemos tenerlo en cuenta para saber cómo debemos enseñar la lengua en el aula sino que también nos sirve como reflexión de nuestra propia práctica docente. La propia actividad docente en sus diferentes fases de enseñanza, aprendizaje y evaluación es un acto comunicativo, por lo tanto “es dependiente de la competencia comunicativa como cualquier otra manifestación lingüística” (Llobera, 1995: 25).

2.2. Características de las herramientas de evaluación.

2.2.1. Fiabilidad, validez y viabilidad

Lo que todos los expertos tienen claro con carácter general (Bachman y Palmer, (1996), Bordón (2006), Prati (2007)) es que las herramientas de evaluación deben ser válidas, fiables y viables para que sean efectivas.

La **fiabilidad** consiste en la consistencia de la medición, es decir, si un alumno realiza un examen hoy y le da un resultado, si lo volviera a realizar mañana le tendría que dar el mismo resultado (Bachman y Palmer, 1996). Para el MCER (2002: 177) la fiabilidad es un concepto técnico:

Se puede decir que una prueba o un procedimiento de evaluación tiene validez en la medida en que pueda demostrarse que lo que se evalúa realmente (el constructo) es lo que, en el contexto en cuestión, se debería evaluar y que la información obtenida es una representación exacta del dominio lingüístico que poseen los alumnos o candidatos que realizan el examen.

No obstante, como ya apuntamos anteriormente, y señala Bordón (2006:171-172),

la subjetividad de la corrección puede incidir en la fiabilidad de la prueba o, mejor dicho, en su falta de fiabilidad (...). El mejor modo de paliar la subjetividad de correcciones realizadas por evaluadores es, en primer lugar, disponer de buenos criterios de evaluación que proporcionen descripciones claras y operativas de niveles de actuación y que asignen puntuaciones, que podrían ser computadas a los distintos niveles asignados.

De ahí la importancia de establecer descriptores claros, objetivos y específicos de cada una de las destrezas teniendo en cuenta el tipo de examen, el nivel y el alumnado al que vayan destinadas. Para comprobar la fiabilidad de una prueba debemos considerar dos tipos: fiabilidad interna que se produce cuando apenas existe variación en la puntuación de una actuación lingüística oral realizada en dos ocasiones distintas, y fiabilidad externa o entre correctores que se refiere a la similitud entre distintos examinadores a la hora de puntuar. Una manera de garantizar la fiabilidad entre correctores y de supervisar el trabajo de los evaluadores es mediante la técnica del muestreo donde un evaluador o jefe del equipo observa la administración de la prueba y puntúa de forma independiente

mientras que el examinador hace lo mismo por su parte; luego, se intercambian puntuaciones y discuten discrepancias. Esto es lo que se suele hacer en la EOI Santa Cruz de Tenerife, contexto de nuestra investigación. Otra forma es mediante la doble corrección de la prueba en su totalidad o de cada una de sus partes: cada examinador puntúa las partes de manera independiente y el jefe del equipo observa si las puntuaciones son parecidas o no, si lo son se halla la media de las dos puntuaciones y si no lo son los dos examinadores deberán volver a corregir la tarea y revisarán los criterios de evaluación relacionados con el nivel que se evalúa. Si aún así no se ponen de acuerdo, se pasará la corrección a otro examinador que puede ser incluso el jefe del equipo. Para todo ello, es necesario grabar las actuaciones de los candidatos.

También podemos recurrir para garantizar esta fiabilidad externa a los “ejercicios de fiabilidad” (Alderson *et al.*, 1998: 130) que se escogerán como los más representativos de entre todos los realizados en función a los distintos aspectos de la escala de puntuación y a los diferentes problemas a los que se enfrentan los examinadores, estos ejercicios ya habrán sido corregidos por un examinador jefe o por una comisión de evaluación. Para ello, se les pide a los examinadores que den una puntuación individual al ejercicio y luego se compara con la del jefe o comisión evaluadora. Si se observan muchas diferencias se reflexiona sobre la causa de estas o se replantea una nueva unificación de criterios.

Los “ejercicios de fiabilidad” y la doble corrección son más difíciles de llevar a cabo en las pruebas de examen oral, ya que sería cara la copia de audios de fiabilidad si hay muchos examinadores. Sin embargo, si los examinadores no son muchos se podrían llevar a cabo, y mediante la grabación de la actuación de los participantes se lograría, además, efectuar la doble corrección. Sin embargo, hay que tener en cuenta que las puntuaciones obtenidas de las actuaciones grabadas pueden ser distintas a las obtenidas mediante actuaciones en directo, ya que en las grabadas no se pueden advertir elementos no verbales que se producen en el discurso o en la interacción oral, por ello, generalmente las puntuaciones establecidas mediante grabaciones suelen ser inferiores a las de actuaciones en directo.

En cuanto a la fiabilidad interna, sólo se puede asegurar corrigiendo el examinador una vez más el ejercicio que ya ha puntuado anteriormente. De esta forma, advertirá si las puntuaciones asignadas en dos momentos distintos difieren mucho entre ellas o no, y si es así se podrán tomar medidas al respecto.

Por otro lado, es necesario la **validez** del constructo (Bachman y Palmer, 1996) en la medida en que podemos interpretar una calificación de la prueba como indicador representativo de la habilidad lingüística del candidato. Además, esto permitirá poder generalizar las tareas de examen a una serie de tareas de uso de la lengua a las que corresponden. Tenemos, así, que preguntarnos si el examen es representativo de la lengua meta del nivel que se pretende medir y si su puntuación está acorde con el objetivo del examen.

Al igual que la fiabilidad existen distintas formas de determinar la validez de una prueba. Así, siguiendo a Alderson *et al.* (1998: 166-180) podemos determinar dos tipos de validez: validez interna que se refiere a los estudios realizados sobre el contenido de una prueba y su efecto, y la validez externa que hace alusión a la comparación de los resultados de las distintas actuaciones de los alumnos y las medidas de su capacidad obtenidas fuera de la prueba.

La validez interna se puede comprobar de distintas maneras: una de ellas es determinando la validez aparente a partir de los comentarios y opiniones obtenidos sobre la valía de una prueba entre personas que no las han elaborado como los alumnos o administradores. A través de ella podemos observar la reputación de nuestros exámenes, pues esto influye en la validez de respuesta de una prueba. Así, un alumno que se toma una prueba “en serio” no contesta de la misma manera que si la considera poco importante. Esto es lo que Bachman y Palmer (1996) llama autenticidad, la cual está relacionada a su vez, con la validez de respuesta, de contenido y de constructo, ya que engloba aspectos que van desde el diseño de las tareas, pasando por la definición del dominio lingüístico a medir, la percepción que tienen los candidatos sobre la lengua que se mide, los materiales con los que se mide hasta las generalizaciones que se puedan establecer con los resultados. Para Bachman y Palmer la autenticidad depende de la definición del concepto de lengua y de las tareas en las que se use este tipo de lengua a evaluar, así como del uso de unos buenos criterios y escalas de evaluación que constaten que los resultados obtenidos reflejan la capacidad de la lengua que se pretende medir, en nuestro caso, la expresión oral.

Otra forma de determinar la validez interna es mediante la validez de contenido que se refiere a la adecuación del contenido de una prueba con lo que se pretende evaluar. En este caso son los redactores o expertos en la materia los que emiten juicios de valor al respecto. Esta validez de contenido está relacionada con las especificaciones de una prueba y con el currículo de la materia. Según Alderson *et al.* (1998) esta validez se

podría determinar mejor a través de la creación de un instrumento para la recogida de datos, por ejemplo, en forma de escalas.

Por último, la validez de respuesta también puede garantizar la validez interna. En ella se observa cómo los candidatos responden a la prueba mediante métodos de recogida cualitativa, por ejemplo, mediante la recogida de datos introspectivos preguntándole al alumno el porqué de su respuesta o manera de proceder, o también de manera retrospectiva. Mediante esta se pueden saber qué procedimientos (estrategias) usan los candidatos para realizar una tarea. Por su parte, Bachman y Palmer (1996: 143-144) no se centran tanto en la respuesta que producen las tareas sino en las características de los candidatos y las habilidades lingüísticas necesarias para que la tarea pueda ser respondida con éxito, es decir se plantean en qué medida interactúa la tarea con el candidato (interactividad).

Por otro lado, la validez externa se puede determinar también de distintas formas, por ejemplo mediante la validez concurrente que supone comparar los resultados de la prueba con otros resultados de estos más o menos el mismo periodo de la prueba. En nuestra investigación se pueden recoger resultados de otras pruebas de los alumnos que no estén relacionados por ejemplo, con la prueba de certificación. En estos casos hay que tratar los resultados de correlación con cautela debido a que es muy difícil comparar la prueba con otra u otras que estén reconocidas como válidas y fiables. Incluso se pueden recopilar las valoraciones que el profesor o profesores que conocen a dichos alumnos tienen sobre su capacidad lingüística comunicativa, no sólo con la capacidad que se evalúa (la expresión oral) sino también con el resto de ellas e, incluso, con la valoración del propio estudiante.

Otra manera es a través de la validez predictiva. En este caso, las medidas externas se recogen después de que la prueba se haya hecho. Así se puede administrar una prueba y posteriormente otra distinta que mida la misma capacidad lingüística comunicativa, aunque nos podemos encontrar con muestras truncadas debido a que las diferencias de resultados entre una prueba y otra puede variar a causa de la mejora en dicha capacidad del candidato al estudiar el idioma en inmersión, perseverancia en los estudios, etc. Se pueden estudiar correlaciones entre resultados de los alumnos durante el curso y los resultados del examen final o, incluso, de certificación con otras medidas de sus capacidades o con el éxito en la correcta clasificación de éste. En este último caso es muy fina la línea que separa la validación concurrente y la predictiva, ya que por ejemplo, en las EOI se hace a un alumno una prueba de clasificación y durante la

primera semana de clase se debe determinar si ese alumno está bien clasificado, antes de que dicho alumno haya tenido tiempo de mejorar, para que la valoración se pueda considerar concurrente y predictiva.

Por último, aparte de la validez interna y externa podemos considerar la validez de constructo que intenta determinar hasta qué punto la prueba se basa en la teoría que subyace, es decir, si la prueba mide la puesta en práctica de la teoría. Una forma de determinar la validez de constructo es observando las correcciones entre distintas partes de la prueba, de cada parte con su totalidad o eliminando la parte analizada. Se puede además, comparar los resultados con las características psicológicas y los datos personales de los alumnos para detectar la equidad de ésta mediante análisis más complejos como el análisis multirrasgo o la validación convergente y divergente⁴ (Alderson *et al.*, 1998: 179)

La fiabilidad y la validez de las herramientas de evaluación no siempre se producen de manera conjunta; no deberíamos tener validez sin fiabilidad, pero los límites no están claros. A este respecto, Alderson *et al.* (1998: 182) señalan la confusión que existe entre estos dos conceptos y consideran que aunque es muy difícil discernir entre ellos, “lo que importa en realidad es si la prueba da un resultado que puede considerarse como un reflejo justo y preciso de la habilidad lingüística del candidato”. Además ni la validez ni la fiabilidad se pueden considerar de manera pura, es decir, como valores absolutos ya que existen grados entre ellas que se deben complementar, pero dependiendo del objetivo de la prueba, de los candidatos a los que vaya dirigido y de las consecuencias de los resultados negativos de ésta en los candidatos se decidirá maximizar una en detrimento de la otra.

La estabilidad de los resultados depende también del rigor con el que se apliquen los criterios, sobre todo en los exámenes a gran escala donde no se pueden tener más muestras de lengua del candidato que en la situación del examen. Por ello, se hace necesaria la formación de los examinadores. De otra manera, si los evaluadores no mantienen los mismos criterios, la evaluación dependerá de la subjetividad del evaluador pudiendo obtener la misma persona evaluada por dos personas distintas en una misma prueba notas totalmente diferentes. En este sentido, el MCER considera que incluso más importante que la fiabilidad es la “precisión de las decisiones” (MCER,

⁴ Para una explicación con detalle consultar Kerlinger, FN. (1973, citado por Alderson *et al.*, 1998: 179). *Foundations of Behavioral Research*. Holt, Rinearth and Winston. Nueva York,

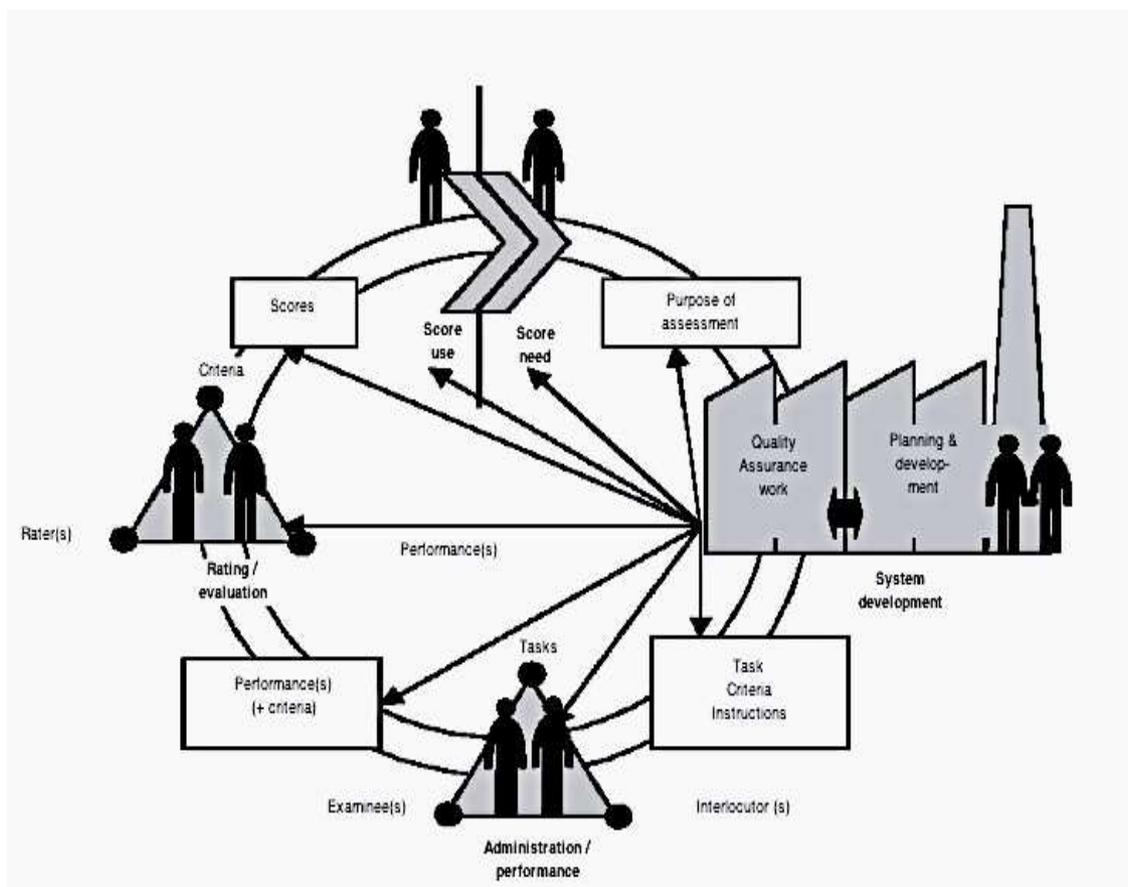
2002: 177). Lo ideal sería tener equipos de redacción de ítems que incluyan a redactores profesionales y a profesores con la experiencia adecuada, ya que las inexactitudes en los sistemas de evaluación y las interpretaciones de los examinadores son una de las fuentes que puede amenazar la solvencia de cualquier prueba. Además, todo esto se complica más en la evaluación de la expresión oral en el aula de idiomas puesto que, en muchas ocasiones, el profesor-examinador no sólo examina al candidato sino que también tiene que actuar de interlocutor haciendo preguntas, respondiendo al candidato e intentando evaluar la actuación, y todo ello a la vez. Por ello, es más necesaria la formación de los examinadores para poder “controlar” todas estas variables, además de comprender las bases que se encuentran en las escalas de puntuación con las que deben trabajar y el de interpretar los descriptores de forma coherente.

Asimismo, Parrondo (2004b: 38) también coincide en la importancia de formar a los examinadores y, basándose en el Código deontológico de la Asociación de entes certificadores de la competencia lingüística (ALTE, *Association of Language Testers in Europe*), establece las obligaciones de los evaluadores en cuatro ámbitos: utilidad, transparencia, equidad y derechos y responsabilidades.

Una forma de poder alcanzar la fiabilidad dentro de la evaluación de la expresión oral es la que advierte Luoma (2009: 4-5), quien considera que cuando percibimos la necesidad de evaluar se pone en marcha un ciclo de actividades para alcanzar este objetivo de la manera más rigurosa posible. Primero hay una etapa de planificación, donde los evaluadores deben definir lo que debe evaluarse. Todos deben tener clara esta definición para luego desarrollar, probar y revisar las tareas y los criterios de calificación y de administración para garantizar así su fiabilidad. Los evaluadores también deben llevar un control (observación) de todo lo que sucede en este ciclo que les permita detectar dificultades e introducir cambios. Luego comienza el periodo de evaluación de los participantes con la administración y calificación de la prueba a partir de los criterios elaborados según las necesidades detectadas en un principio. Si se observan más dificultades el ciclo puede volver a comenzar de nuevo. Así, vemos como Luoma, en el esquema que nos presenta, aparece el lugar de trabajo donde los evaluadores elaboran los criterios, las tareas, las instrucciones y todo lo necesario para la fiabilidad y viabilidad de la prueba como una fábrica, ya que se elabora un producto con unas garantías de calidad: “*As at any factory, quality assurance is an important aspect of the development work. It ensures that the testing practices being developed are good enough for the original purpose*” (2009: 5).

Todo este proceso Luoma lo representa de la siguiente manera, donde se muestra que el trabajo de control de calidad se extiende sobre el ciclo de evaluación en su conjunto (Luoma, 2009: 5):

Cuadro 7.



Otro requisito que debe tener toda herramienta de evaluación es la viabilidad, llamada por Bachman (1996: 35) *practicability*, quien considera que esta viabilidad se da cuando existen recursos tanto de tipo humano, académicos, económicos y estructurales para diseñar, desarrollar y hacer uso del examen, así como cuando existe la disponibilidad de los mismos, es decir, cuando disponemos de todos los recursos adecuados para llevar a cabo el examen. En ocasiones, esta viabilidad no depende del profesor o del evaluador sino de las instituciones educativas pertinentes; el profesor sólo puede hacer todo lo posible para garantizar la viabilidad en función de los medios y recursos disponibles no sólo de la institución académica sino también de sus propios medios y recursos como su tiempo fuera del trabajo, su propio ordenador ante la no disponibilidad de estos, pagarse cursos de perfeccionamiento sobre la evaluación o de informática, entre otros. Ante estas situaciones en donde no existe una formación previa en la evaluación el profesor,

desde su propia ética como profesional de la enseñanza, intentará que los resultados obtenidos sean todo lo adecuados y objetivos posible a la actuación del candidato pero, como ya he dicho, hay aspectos que son competencia de las instituciones educativas y que están fuera del alcance del profesor para que se den estos requisitos aquí mencionados.

2.2.2. Autenticidad, interactividad e impacto.

Los aspectos de autenticidad, interactividad e impacto también son requisitos fundamentales para determinar la calidad de una herramienta de evaluación (Bachman y Palmer (1996), Bordón (2006)).

Desde que se ha adoptado el enfoque comunicativo en la enseñanza de lenguas, las características de las tareas de examen deben corresponder con las características de las tareas de uso de la lengua meta. Según propone Bachman (1996: 23) *“we define authenticity as the degree of correspondence of the characteristics of a given language test task to the features of a TLU⁵ task”*. Así, si existe **autenticidad**, podemos determinar si el candidato puede desenvolverse satisfactoriamente en situaciones reales fuera del examen. En la actualidad, todos los libros de texto y exámenes tienen tareas que cumplen este requisito ya que las tareas de simulación de situaciones reales en el aula son bastante eficaces para acercar al alumno al uso de la lengua en acciones de la vida diaria. Así, por ejemplo, tanto en el aprendizaje del idioma como en su evaluación y teniendo en cuenta el nivel se proponen tareas, en el caso de la expresión oral que es el que nos ocupa, de exposición e interacción en situaciones que pueden ocurrir en la realidad como, por ejemplo, debates, diálogos en una entrevista de trabajo, argumentaciones sobre un tema, etc.

En las tareas de interacción oral se cumple otro de los requisitos que debe tener toda herramienta de evaluación: la **interactividad**. Con este tipo de tareas se puede observar la medida en que las habilidades y las características del alumno se usan para el cumplimiento de la tarea de examen. Estas habilidades y características según Bachman (1996) son el conocimiento lingüístico, la competencia estratégica y las estrategias metacognitivas. Así, mientras un alumno se está expresando oralmente está poniendo en marcha todas sus habilidades para llevar a cabo con éxito la tarea. Estas habilidades también las usa cuando el candidato se expresa mediante un monólogo, ya que en

⁵ Bachman (1996: 18) llama TLU a “target language use”, entendido en español como “uso de la lengua meta” (ULM).

realidad interactúa con una audiencia, pero sólo a través de la interacción cara a cara se pueden utilizar algunos conocimientos y estrategias como son el registro, la confirmación en el entendimiento de la información, las vacilaciones o recordar lo dicho anteriormente, por ejemplo (Luoma, 2009). Estos aspectos se deberían reflejar en los criterios de corrección y calificación de las tareas de expresión oral en interacción, porque si el hablante los utiliza demuestra que sabe expresarse de manera interactiva con otros interlocutores y no se limita a una mera exposición de su parte de la tarea para luego callarse y cederle el turno al interlocutor, como ocurre en muchas ocasiones con algunos candidatos en este tipo de pruebas.

Por último, es necesario percatarse del **impacto** que genera la prueba, denominado repercusión o efecto rebote, traducciones del inglés *backwash* o *washback*, entendido como el efecto que produce el examen en los individuos y en la sociedad o sistema educativo (Bachman, 1996). En el caso de nuestra investigación, el impacto de los exámenes de certificación es muy alto ya que afecta a los diseñadores del examen, a las personas que realizan el examen y a los profesores que preparan a los candidatos. Igualmente, repercuten en la metodología de la enseñanza, por lo que aún se hace más necesario que se cumplan todos estos criterios que hemos citado.

Todas estas características de fiabilidad, viabilidad, validez, autenticidad, interactividad e impacto que dan utilidad al examen no deben ser observadas y aplicadas separadamente, sino que se encuentran relacionadas entre sí. Desde el momento en que en un examen no se dé alguna de ellas, no se podría considerar un buen examen, ya que su resultado no representaría el dominio del candidato del idioma y nivel en el que se examina.

2.2.3. Otros factores condicionantes en la evaluación

Como hemos visto anteriormente, uno de los factores que influye en la objetividad de los resultados de la evaluación es la falta de fiabilidad. El problema está cuando, dependiendo del evaluador, un mismo candidato puede obtener calificaciones o percepciones de su habilidad diferentes, incidiendo en esta falta de fiabilidad y, por consiguiente, en la validez de la prueba.

Pero nos encontramos con más factores que pueden acentuar la subjetividad de la evaluación. Uno de ellos es la personalidad de los candidatos, y aunque es verdad que este aspecto no se puede controlar, los evaluadores deberían tenerlo en cuenta para que no les influya en su percepción. Así, podría parecer que un candidato extrovertido y con

buena pronunciación es muy competente frente a otro que es más tímido. En algunas ocasiones, esto ha hecho que algunos evaluadores puntúen con mayor nota al extrovertido que al tímido. Sin embargo, esto no puede suceder en la evaluación ya que influiría la personalidad del candidato en la nota final y, por tanto, la evaluación carecería de fiabilidad y validez. Consideramos que estas situaciones se producen debido a que la parte expositiva de la prueba, sobre todo en un nivel Básico, tiene una duración muy escasa de tres a cinco minutos, lo que no permite realmente obtener muestras de la lengua aceptables para evaluar la competencia del candidato.

Se ha demostrado, según estudios llevados a cabo a este respecto, que “el tiempo mínimo para obtener muestras relevantes de lengua está fijado en diez minutos” (Bordón, 2006: 214) por lo que una posible solución sería alargar estas tareas expositivas, como mínimo a diez minutos. No obstante, hay un factor decisivo en la temporalidad de la prueba que es la viabilidad de ésta, sobre todo en los exámenes de alto impacto en donde hay un mayor número de candidatos. Esto hace que se reduzca el tiempo de las tareas llevando a una obtención de los resultados de manera precipitada y, por lo tanto, a una falsa calificación del candidato.

Por otro lado, como advierten muchos lingüistas (Bachman y Palmer, 1996 y Bordón, 2006) la personalidad del candidato también influye en las tareas de interacción debido a que un carácter extrovertido tiende a acaparar toda la conversación, no dejando muchas oportunidades de demostrar su competencia a un candidato más tímido, a pesar de respetar los diez minutos como mínimo. Esto lleva a los evaluadores sin formación o sin experiencia a la falsa creencia que hemos citado anteriormente y que también ocurre en la tarea expositiva: el candidato más extrovertido es el más competente. Por ello, es importante poder asegurar el “entrenamiento” de los evaluadores que les permitirá detectar estas situaciones y actuar en consecuencia sin que su calificación esté influida por estos aspectos. En el mismo sentido “la lengua materna, la formación académica, la cultura o el sexo repercuten en la actuación del candidato en un examen” (Bordón, 2004b: 19)

Asimismo, Bachman (citado por Bordón, 2004b: 18-19) advierte que no sólo la lengua materna de la persona evaluada influye en la evaluación, sino también la lengua materna del evaluador, así como su formación y experiencia.

2.3. Proyectos y estudios recientes

Se exponen a continuación algunos estudios y proyectos relacionados con la elaboración de criterios de evaluación que suponen la certificación oficial de la competencia idiomática en los que se establecen unos objetivos y se determinan metodologías a seguir. Estos proyectos nos sirven de referencia y base para la elaboración de nuestros propios criterios de evaluación dentro de nuestra investigación.

Como ya hemos visto, el correcto desarrollo y diseño de los descriptores asociados a las escalas de nivel son de extrema importancia en la evaluación ya que proporcionan modelos con los que los candidatos pueden compararse para autoevaluarse, y son muy eficaces para aumentar la fiabilidad de una evaluación subjetiva como es la de la expresión oral. Pero debemos destacar que la mayor o menor exactitud en los descriptores de la escala, el procedimiento llevado a cabo o la aplicación de la metodología adecuada no es lo que hace que la evaluación de una tarea, por ejemplo, sea fiable sino que la dificultad reside en la finalidad para la que se usan estos criterios (Parrondo, 2004a), teniendo en cuenta el contexto, el tipo de alumnado y los objetivos de aprendizaje

Casi todas las investigaciones relacionadas con la evaluación y más concretamente con el desarrollo de criterios, descriptores o escalas dentro de esta han tenido como referencia los siguientes proyectos que se podrían considerar los más destacables (Parrondo, 2004a; Bordón, 2006; Prati, 2007):

En primer lugar, podemos nombrar al proyecto suizo, financiado por el Consejo superior de investigación científica suizo, basado en un banco de descriptores elaborados a partir de otras escalas ya existentes de medición de la competencia; a través de estas y mediante su experimentación con profesores, se realizaron descriptores más claros y útiles para los usuarios. Posteriormente, a partir de una selección de estos descriptores los profesores participantes les presentaron a sus alumnos un cuestionario de autoevaluación. Según Parrondo (2004a: 969):

este banco de descriptores se utilizó después para elaborar la primera edición de los niveles comunes de referencia en el *Marco común europeo de referencia* (2002) y crear los instrumentos de autoevaluación necesarios para la creación de un prototipo de *Pasaporte lingüístico europeo* o *Portfolio europeo de las lenguas*, documento ideado para recoger los logros, etapa por etapa, de los usuarios en el aprendizaje de otras lenguas.

Otro de los proyectos destacados es el Proyecto DIALANG que desarrolla escalas de evaluación con la finalidad de clasificar o diagnosticar al candidato en su nivel de competencia mediante una autoevaluación, diferentes pruebas de competencia

lingüística ofrecidas a través de Internet y su correspondiente retroalimentación. Se ofrecen escalas de evaluación en 14 idiomas y le confiere al alumno una autonomía en su propio proceso de aprendizaje.

Por su parte, el proyecto CAN-DO, desarrollado por ALTE tiene como objetivo la validación de descriptores de la capacidad lingüística con relación al MCER. De esta manera les es más fácil a los usuarios comprender la escala de niveles en relación con las habilidades reales de las que hace uso. Cabe destacar que la asociación ALTE tiene como objetivo el establecimiento de escalas y niveles comunes de competencia así como el establecimiento de formas de proceder comunes para la elaboración, desarrollo y distribución de sistemas de calificación y de comunicación de resultados así como el análisis e investigación de estos. Por ello creó un código deontológico o “Código de Buena Práctica” para los evaluadores en donde se especifican medidas y métodos de trabajo comunes en la evaluación.

En un intento por desarrollar lo que los usuarios deberían hacer cuando utilizan una lengua, el Consejo de Europa desarrolla una serie de especificaciones estableciendo una descripción de los niveles de competencia lingüística. Estas especificaciones fueron desarrolladas primero para inglés llamadas “nivel umbral” que, posteriormente, fueron adaptadas al francés (*un niveau-seuil*) y al español, desarrolladas estas últimas por Peter Slagter en 1979. A partir de estas primeras especificaciones han surgido nuevas ideas y ampliaciones, fundamentalmente en inglés, pero actualmente seguimos la descripción de niveles recogida en el MCER (2002), del que hablaremos posteriormente, y que en el caso del español no ha sido objeto de revisión hasta la fecha. La descripción de niveles de competencia lingüística se detalla a continuación:

Cuadro 8.

Avanzado	C2	Maestría
	C1	Dominio operativo eficaz
Intermedio	B2	Avanzado
	B1	Umbral
Principiante	A2	Plataforma
	A1	Acceso

En el caso de las EOI de Canarias seguimos esta descripción de niveles, aunque con la salvedad de que no impartimos clase del nivel avanzado del MCER (C1, C2); nuestro nivel avanzado es el B2.

De manera más reciente, surge el Marco común europeo de referencia (MCER, 2002) que también fue desarrollado por el Consejo de Europa, y que “proporciona una base común para la elaboración de programas de lenguas, orientaciones curriculares, exámenes, manuales, etc., en toda Europa” (MCER, 2002: 1). De esta manera se pretende “«conseguir una mayor unidad entre sus miembros» y aspirar a este objetivo «adoptando una acción común en el ámbito cultural»” (*ibidem*: 2). Actualmente, el MCER es una referencia ineludible en todas las enseñanzas de idiomas influyendo en la programación, metodología, evaluación y materiales de las instituciones educativas como son las EOI o el Instituto Cervantes cuyo Plan curricular, desarrollado a partir del MCER, es una referencia en la didáctica del español.

En relación con el MCER, la asociación ALTE estableció las bases para un modelo de Portfolio europeo de las lenguas (PEL) validado por el Consejo de Europa y que muchas instituciones educativas utilizan en sus enseñanzas. El PEL constituye el pasaporte lingüístico del candidato conformado por una carpeta de documentos de trabajos hechos por el alumno, reflexiones y experiencias sobre su proceso de aprendizaje. Todo ello sirve de complemento a los certificados oficiales y facilita a los usuarios la comprensión de los niveles de competencia que han adquirido en su proceso de aprendizaje en la lengua objeto de estudio. Además, también es útil para ofrecer información a terceras personas en una entrevista de trabajo, por ejemplo, sobre nuestra competencia lingüística de manera detallada y homologada.

También cabe destacar otras asociaciones de evaluación de lenguas que operan en Europa. Así, además de ALTE tenemos a la Asociación europea de evaluación y pruebas de lengua⁶ (EALTA, *European Association for Language Testing and Assessment*) o a la Asociación Internacional de pruebas de lengua⁷ (ILTA, *International Language Testing Association*). ILTA se estableció en el año 1994 y desde principios de los años 80 organiza una conferencia internacional relacionada con las pruebas de lengua, el Coloquio de investigación de pruebas de lengua (LTRC, en sus siglas en inglés). ILTA pretende mediante estas conferencias, entre otras metas, estimular el

⁶ Para más información se puede consultar su página web: www.ealta.eu.org.

⁷ Para más información se puede consultar su página web: www.iltaonline.com.

desarrollo profesional, promover la difusión y publicación de información relacionada con el área de pruebas de lengua, desarrollar el área de las pruebas de lengua y proporcionar personas capaces de liderar esa área y promover servicios profesionales a sus miembros. La LTRC de ILTA suele centrar sus ponencias en los exámenes a gran escala más que en las evaluaciones de aula. Por su parte, EALTA, creada en 2004, pretende acercar las pruebas de lengua y la evaluación al mayor número de profesionales en Europa. Para ello tiene por objetivo compartir la experiencia profesional en el campo de la evaluación y pruebas de lengua, incrementar la comprensión de la evaluación de dichas pruebas, promover el reconocimiento de la evaluación y pruebas de lengua como disciplina profesional en Europa, mejorar la práctica de los sistemas de evaluación de lenguas o formar a expertos en dichas pruebas. También, al igual que ILTA, ofrece conferencias todos los años para favorecer este intercambio de experiencias y dotar de formación a los profesionales de la enseñanza de lenguas, ya que también lleva a cabo seminarios específicos y talleres. Asimismo, debido a la importancia que tiene la formación de los evaluadores tanto ALTE como EALTA han desarrollado un código deontológico que sirve de guía. Por su parte, ALTE ha diseñado un Código de Buena Práctica para los elaboradores de pruebas y EALTA ha elaborado un código de deontología profesional que abarca directrices de buena práctica en tres áreas: la evaluación de lengua tanto en la formación de profesores que aún no se han incorporado a su práctica profesional como en la de los que se encuentran trabajando, la evaluación en el aula y las pruebas de lengua a una mayor escala, institucionales o nacionales.

Todos estos proyectos y asociaciones, y actualmente de una manera más concreta el MCER, han servido para crear y desarrollar unos criterios de evaluación y unos descriptores más precisos para conseguir unas pruebas de calidad y unos resultados fiables. Sobre estas bases se han hecho los Diplomas de Español como Lengua Extranjera (DELE) y las Pruebas de Certificación de las EOI (PCEOI).

Los DELE, que acreditan el conocimiento del idioma español, fueron creados por el Ministerio de Educación de España en 1989, por el Real Decreto 826/188, de 20 de julio, con el objetivo de difundir el español e incrementar la presencia del español en el mundo. Primeramente estos exámenes sólo se realizaban en el extranjero y eran elaborados por el Ministerio de Educación de Madrid; además, solo existía el Diploma Básico de Español e incluía, junto a tareas para evaluar las destrezas lingüísticas, tareas sobre comentarios de texto que medían los conocimientos literarios del candidato.

Posteriormente, en 1990 se creó el Diploma Superior de Español y se ajustó el tipo de pruebas centrándose más en lo lingüístico y eliminando lo cultural o literario.

En 1991 la Universidad de Salamanca, pasó a encargarse de la elaboración y calificación de estos exámenes por un convenio firmado con el Ministerio de Educación. Esta Universidad ya se había informado sobre cómo confeccionar pruebas de calidad a través de ALTE. En este mismo año nace el Instituto Cervantes y se convierte en la mejor institución para difundir la lengua española con numerosas sedes en el extranjero.

En 1993 la Universidad de Salamanca siguiendo las directrices del nivel umbral para el español establecidas por Slagter en 1979, crea el Certificado Inicial de Español. Posteriormente el Instituto Cervantes se convierte en la institución adecuada para hacerse cargo de la elaboración, distribución y calificación de las pruebas y, por lo tanto, en ente certificador, tal como está en la actualidad. Los certificados cambian de nombre y pasan a llamarse Diploma de Español nivel inicial, nivel intermedio y nivel superior. Actualmente estos diplomas se han actualizado siguiendo el MCER en Diploma de Español nivel A1, Diploma de Español nivel A2, Diploma de Español nivel B1 (antes inicial), Diploma de Español nivel B2 (antes intermedio), Diploma de Español nivel C1, Diploma de Español nivel C2 (antes superior).

Aparte del Instituto Cervantes, son muchas las instituciones educativas que han desarrollado sus pruebas de certificación, como la Universidad de Buenos Aires, que teniendo en cuenta las necesidades de los extranjeros que vivían en el país y partiendo del análisis de muestras de escritura y oralidad, fue delineando descriptores de nivel y especificaciones de exámenes que resultaron un formato probado y homologado entre las pruebas, dando lugar, así, a la Certificación de la Competencia Lingüística de ELSE en la Argentina (Prati, 2007).

En las EOI, en 1998 a través del Real Decreto 967/1998, de 2 de septiembre (BOE 218) se expedían certificaciones académicas, no diplomas, al finalizar los estudios conducentes al aprendizaje de un idioma. Así, al finalizar los estudios del idioma durante tres cursos se expedía el certificado de nivel elemental y al finalizar dos cursos más se expedía el certificado de nivel superior. Cuando ya el alumno tenía este certificado de nivel superior, el Ministerio de educación expedía un Certificado de aptitud del idioma objeto de estudio. Lo podemos ver resumido en la siguiente tabla:

Cuadro 9.

CURSO	CERTIFICADO
Primero	
Segundo	
Tercero	Certificado de nivel elemental
Cuarto	
Quinto	Certificado de nivel superior Certificado de Aptitud del idioma

A partir del Real Decreto 944/2003, de 18 de julio (BOE, nº 182) por la que se establece la estructuras de enseñanzas de idiomas de régimen especial, se pasa de estudiar un idioma en cinco cursos a hacerlo en seis cursos, haciéndolo coincidir con la descripción de niveles del MCER. Se establecen, así, tres niveles (Básico, Intermedio y Avanzado) y al finalizar cada uno de ellos se expide el certificado de nivel Básico (A2), de nivel Intermedio (B1) y de nivel Avanzado (B2), no contemplándose las enseñanzas de nivel C1 y C2.

En un principio, los exámenes de certificación los realizaba un equipo contratado para tal fin que normalmente eran profesores voluntarios de EOI. Pero durante los últimos dos cursos escolares son los distintos departamentos de las EOI de Canarias los que tienen que elaborar, distribuir y evaluar estos exámenes de certificación. De ahí que sea tan importante establecer criterios válidos, fiables y viables para este tipo de exámenes teniendo en cuenta el contexto y los alumnos implicados.

2.4. Conclusiones

Como se ha podido comprobar, en todo diseño, elaboración, distribución y evaluación de cualquier herramienta de evaluación es necesario que se cumplan determinados aspectos que hacen que se convierta en una herramienta de calidad como son: la validez, fiabilidad, viabilidad, autenticidad, interactividad e impacto (Bachman y Palmer, 1996). Estos aspectos parecen aún más difíciles de llevar a cabo en evaluaciones subjetivas donde interviene la acción humana como es el caso de la expresión oral, donde también inciden otros factores como son las características de los propios candidatos, la inmediatez de la información o la naturaleza cooperativa de toda acción verbal, que sigue unas normas de construcción compartida del discurso. Para ello, se hace necesaria la formación de los examinadores que mediante un código deontológico puedan realizar acciones y tomar decisiones comunes sin que se vea afectada la objetividad, así como

tener claro cuál es el objetivo de la evaluación e, incluso, de cada prueba, qué enfoque es el que estamos utilizando y qué lengua vamos a evaluar. Además, si conseguimos que los descriptores sean formulados con mayor claridad y detalle y, sobre todo, sean adecuados al contexto y al tipo de alumnado, creemos posible que pueda mitigarse la subjetividad en la evaluación (Bordón, 2006).

En el siguiente capítulo, referido al marco teórico, explicaremos con más detalle cómo se pueden conseguir estos aspectos citados por Bachman y Pamer (1996) mediante la creación de pruebas de evaluación de calidad, qué tipo de formación deben tener los evaluadores para ello y la presencia de una dimensión que consideramos imprescindible dentro de la evaluación: la ética.

3. MARCO TEÓRICO

Partimos de la creencia de muchos expertos en la investigación de la evaluación de la enseñanza de idiomas de que la evaluación cumple un doble papel, pedagógico y práctico, que permite clarificar objetivos tanto al profesor como al alumno, impulsar al alumno a tomar consciencia de su propio aprendizaje y proporcionar al profesor la toma de decisiones. Esto no sólo se consigue por medio de exámenes sino también a través de otras herramientas de evaluación que no tienen por qué estar enfocadas al resultado, sino al proceso que conducen a él, permitiendo averiguar las dificultades y los logros en la adquisición del aprendizaje de la segunda lengua. Para ello, son indispensables pruebas viables y descriptores fiables elaborados con rigor y validados empíricamente (Bordón (2006), Alderson, *et al.* (1998), Luoma (2009) o Yagüe (2010) entre otros).

Como se puede ver, la evaluación de la lengua es un campo bastante prolijo que durante años ha dado lugar a un gran número de líneas de investigación para profundizar en esta cuestión, tanto para poder determinar la habilidad lingüística de la forma más precisa posible como para profundizar en el conocimiento de su naturaleza. Como ya hemos visto, se han hecho imprescindibles conceptos como fiabilidad, validez y viabilidad de las tareas de examen o han surgido múltiples reflexiones acerca de los tipos de evaluación, sobre qué evaluamos o las herramientas y características necesarias para realizar un buen examen. Es evidente que la extensión del tema es bastante amplia y a pesar de que haremos referencia constante a muchos de estos conceptos en nuestro trabajo, como ya se han tratado anteriormente en el apartado correspondiente al estado de la cuestión, ahora nos centraremos en los aspectos teóricos relacionados con la elaboración de pruebas o exámenes de idiomas, así como la redacción de descriptores de evaluación.

3.1. Elaboración de pruebas de evaluación de calidad

A tenor de lo dicho anteriormente, podemos advertir que no sólo los descriptores de evaluación son indispensables para asegurar la transparencia de los resultados, sino que también estos criterios tienen que estar en consonancia con las pruebas que evalúan, por lo que el diseño y la elaboración de estas también debe de hacerse con el mismo rigor que el que se emplea para el de los criterios de evaluación.

Son muchos los estudios sobre la evaluación que tratan de estos aspectos pero nos vamos a centrar en dos aportaciones que nos parecen las más importantes y relevantes

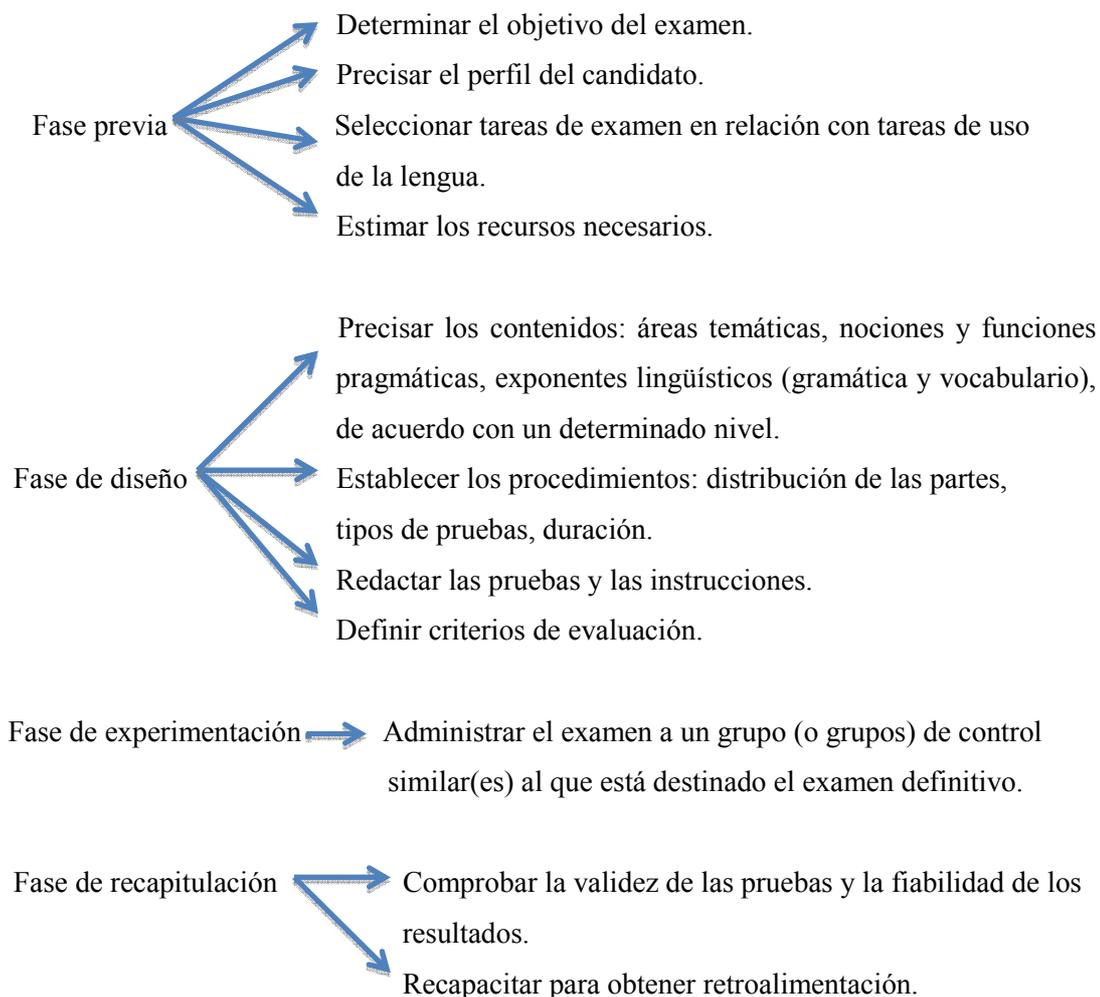
en relación a la elaboración de pruebas de evaluación de calidad e indispensables para nuestra investigación: la de Bachman y Palmer (1996) y la de Bordón (2006).

Las aportaciones de Bachman y Palmer (1996) sobre las características que debe tener toda prueba de evaluación de una lengua (fiabilidad, validez, viabilidad, interacción, autenticidad e impacto), ya las hemos desarrollado en el apartado del estado de la cuestión de este trabajo y no nos vamos a extender más en ellas. Sobre estas características se han basado muchos diseñadores de pruebas para garantizar la calidad de estas y nos servirán para hacer un análisis de cada una de las tareas que elaboremos para las pruebas de certificación⁸

Pero si bien Bachman y Palmer (1990 y 1996) nos proporcionan las características, Bordón (2006) nos propone algunas sugerencias para diseñar un buen examen. Así, considera que en una fase previa al diseño del examen se debe decidir cuál es el objetivo del examen lo que permitirá saber cuáles son los usuarios a los que va dirigido, el tipo de recursos necesarios o qué tareas vamos a emplear en la prueba. Posteriormente pasaremos al diseño de éste que, en el caso de un examen de certificación, se tendrán que tener en cuenta los contenidos de acuerdo con el nivel que proponga. Asimismo, se determinarán las partes de las que consta el examen, cuántas pruebas se incluirán en cada parte y de qué tipo serán dichas pruebas, se redactarán las pruebas y las instrucciones, así como los criterios de evaluación y la redacción de los descriptores. Luego es necesaria una fase de experimentación con candidatos de las mismas características a los que va destinado el examen. Esta fase permitirá comprobar la validez de la prueba y la fiabilidad de los criterios de evaluación en donde si se observara alguna dificultad se revisarían y se modificarían para mejorarlos. Por último, se debe hacer una reflexión a partir de los resultados obtenidos para poder mejorarlo y también para reflexionar sobre nuestra propia labor como evaluadores y diseñadores de pruebas. Podemos ver todo el proceso resumido en el siguiente esquema que nos proporciona Bordón (2006: 68):

⁸ Ver cuadro de análisis de tareas, anexo 6.

Cuadro 10.



Todas estas apreciaciones y características que nos ofrecen Bachman y Palmer (1996) y Bordón (2006) nos permiten generalizar modos de proceder y unificar criterios a la hora de elaborar pruebas y criterios de corrección y nos servirán de base en nuestra investigación.

3.1.1. Pruebas de evaluación de la expresión oral.

Además de los aspectos apuntados en el apartado anterior, en el caso de las pruebas de expresión oral debemos advertir otras consideraciones no menos importantes que garantizan la idoneidad de ésta.

Hemos visto que para poder evaluar la expresión oral se tienen que plantear tareas comunicativas donde el candidato pueda demostrar su habilidad lingüística. En ellas se tienen en cuenta no sólo los conocimientos sino también las habilidades del candidato para resolver con éxito la tarea. En este sentido, el MCER nos ofrece distintos tipos de

tareas que podemos llevar a cabo en la evaluación oral y distingue por un lado las actividades que denomina de expresión oral “hablar” en donde “el usuario de la lengua produce un texto oral que es recibido por uno o más oyentes” (MCER, 2002: 61), y por otro, las actividades de interacción oral donde el usuario debe conversar con uno o más candidatos.

Las actividades de expresión oral que cita el MCER es lo que nosotros llamaremos tareas monologadas o “monogestionadas” (Calsamiglia y Tusón, 2007: 19), donde el candidato produce él solo su actuación. Algunos ejemplos de actividades de este tipo que nos propone el MCER (2002: 61) en su capítulo cuatro son las siguientes:

- realizar comunicados públicos (información, instrucciones, etc.);
 - dirigirse a un público (discursos en reuniones públicas, conferencias universitarias, sermones, espectáculos, comentarios deportivos, presentaciones de ventas, etc.).
- Pueden suponer, por ejemplo:
- leer en voz alta un texto escrito;
 - hablar apoyándose en notas, en un texto escrito o en elementos visuales (esquemas, imágenes, gráficos, etc.);
 - representar un papel ensayado;
 - hablar espontáneamente;
 - cantar.

Las relaciones que se producen en este tipo de actividades es de persona a audiencia. Sin embargo, en algunas de estas actividades como leer en voz alta un texto escrito, hablar apoyándose en notas o cantar, “la expresión oral es sólo el canal, porque el verdadero discurso es escrito y el hablante “recita” de memoria o lee, pero no está produciendo lengua propia” (Bordón, 2006: 204). Como se puede ver, estas actividades son consideradas por algunos lingüistas no adecuadas para evaluar la producción oral ya que ésta se basa en la escrita y estiman que desde un enfoque comunicativo “la materialización “más auténtica” de la expresión oral es la que reviste la forma de una conversación” (*ibidem*). No obstante, consideramos que en estos casos se dan otros rasgos propios de la expresión oral como la inmediatez en la producción y la aparición de elementos imprevistos que pueden influir en la comunicación. Debemos recordar que la imprevisibilidad es una de las características fundamentales de la modalidad oral y por mucho que se “estudie” de memoria lo que se va a decir o cómo se va a actuar, lo imprevisto siempre puede suceder. Por lo tanto, si bien algunas de estas actividades monologadas están lejos de la conversación, comparten con ella aspectos de improvisación e incluso malentendidos, reformulación de lo que se iba a decir o de provocación de conflictos como ocurre en los intercambios dialogados o “plurigestionados” (Calsamiglia y Tusón, 2007).

Para estas actividades monologadas el MCER nos ofrece escalas de nivel que se pueden consultar en el capítulo cuatro del MCER (2002: 61-64) con descriptores para la expresión oral en general, para el monólogo sostenido como descripción de experiencias y como argumentación, por ejemplo en un debate, para declaraciones públicas y para hablar en público.

Estas escalas son de gran ayuda para el evaluador; sin embargo, se puede apreciar que en algunos niveles no existe descriptor disponible por lo que creemos que en estos casos el evaluador puede recurrir a lo que se considera que debe hacer el candidato en la escala general, es decir, por ejemplo, en la expresión oral en general; o también hacer él mismo el descriptor para esos niveles donde no existe basándose en su experiencia y/o formación.

Por otro lado, el segundo tipo de actividades que el MCER nos aporta son actividades y estrategias de interacción oral. La interacción oral se produce cuando el candidato actúa ante un interlocutor haciendo a su vez los papeles de hablante y oyente. Para ello, debe utilizar no sólo conocimientos lingüísticos sino también estrategias que exigen una cooperación entre los candidatos. Las actividades y estrategias de interacción oral que cita el MCER (2002: 75) son:

- transacciones;
- conversación casual;
- discusión informal;
- discusión formal;
- debate;
- entrevista;
- negociación;
- planificación conjunta;
- cooperación práctica centrada en los objetivos;
- etcétera.

Igualmente, ofrece escalas ilustrativas con descriptores por niveles para la interacción oral en general, para comprender a un interlocutor nativo, para la conversación informal y la conversación formal, por ejemplo en reuniones de trabajo, para colaborar para alcanzar un objetivo, para interactuar con el fin de obtener bienes y servicios, para intercambiar información y para entrevistar y ser entrevistado, que también se pueden consultar en el capítulo cuatro del MCER (2002: 75-81).

Aparte de darnos una tipología de actividades de expresión e interacción oral para la elaboración de las pruebas correspondientes a esta capacidad lingüística, el MCER va más allá y nos advierte que en la interacción intervienen toda una serie de estrategias como son:

- La planificación, donde se consideran los posibles intercambios que se van a llevar a cabo en la tarea, lo que el MCER denomina “encuadre”, y donde se activan una serie de esquemas mentales que desarrollan estas consideraciones (un “praxeograma”, es decir, un diagrama que representa la estructura de la interacción comunicativa). Además, también se produce una identificación de vacío de información y de opinión (condiciones de idoneidad), una valoración de lo que se puede dar por supuesto y, por último una planificación de los intercambios posibles o probables.
- La ejecución: durante la misma actuación de los candidatos se producen turnos de palabras que hacen referencia al momento en el que cada uno de los candidatos está interviniendo llevando la iniciativa en el discurso, cooperación interpersonal y de pensamiento para asegurar una cooperación en la tarea y una comprensión mutua, estrategias para enfrentarse a lo inesperado y petición de ayuda.
- La evaluación, que se origina cuando se produce un ajuste entre el esquema mental que se pensaba utilizar y lo que ocurre realmente (control de los esquemas y del “praxeograma”) así como cuando se valora el éxito, es decir, si todo ha salido como se pensaba (control del efecto y del éxito).
- La corrección, que se puede observar mediante estrategias que usan los candidatos para pedir u ofrecer aclaraciones, así como para reparar la comunicación cuando ésta se ha visto dañada al producirse malentendidos.

Estas estrategias también se miden por escalas descriptivas de nivel (MCER, 2002: 84-85) y las debe tener en cuenta el evaluador en la corrección de las pruebas de expresión oral.

Además de todos estos aspectos, la producción oral contiene otras variables que el evaluador percibe en la actuación del candidato y que no debemos pasar por alto. Así, una intervención oral puede ir acompañada por gestos y acciones no verbales que influyen en la comunicación (por ejemplo, señalar) o acciones paralingüísticas que conllevan significados comunes dentro de una misma cultura y que pueden ser diferentes en otras como, por ejemplo, gestos relacionados con el comportamiento de una persona (estar loco (el dedo haciendo círculos a la altura de la sien) o guiños de complicidad). Igualmente existen sonidos que llevan significados que se han hecho convencionales (sonidos extralingüísticos en el habla: “ay, bah, uuuuh”) o cualidades prosódicas como el tono de la voz, la cualidad, el volumen o a duración de ésta.

Como vemos, la expresión oral contiene innumerables variables que el evaluador debe

tener en cuenta a la hora de emitir un resultado, y las escalas que nos proporciona el MCER son de gran ayuda ya que sirven de guía para esta labor tan difícil. Es evidente que los últimos aspectos pertenecientes a la comunicación no verbal no los podemos medir y valorar por escalas pero también influyen en la comunicación y no se pueden apartar de la valoración de la actuación del candidato.

Una de las actividades que cita el MCER como actividad de interacción es la entrevista. Este tipo de actividad se ha utilizado tradicionalmente en muchas instituciones educativas bajo la creencia de que es la que mejor refleja la actuación del candidato, ya que “parece que esta prueba es la que reviste más visos de autenticidad en lo que respecta a la expresión oral (si estamos de acuerdo en que la forma más representativa de la expresión oral la constituye la conversación)” (Bordón, 2004a: 995). Sin embargo podemos comprobar que algunos científicos estiman, mediante algunos estudios hechos al respecto, que este tipo de actividad no es la adecuada para valorar la expresión oral. Figueras (2004: 63-84) hace un análisis crítico de la entrevista como tarea de examen. Según su estudio hecho a alumnos adultos de nivel Intermedio de inglés, la entrevista sólo permite un contexto único (dos profesores frente a un alumno), dando lugar a una muestra muy reducida de lengua, ya que el candidato sólo se limita a responder a las preguntas del evaluador. Esto supone que la conversación no sea natural, puesto que hay pocas ocasiones en las que el candidato puede empezar un turno, no existen las interrupciones o las respuestas imprevistas que se suelen dar en una interacción real y sólo alguna petición de aclaración o de repetición de la pregunta por parte de la persona evaluada. Por tanto, la muestra de lengua en este tipo de pruebas orales (la entrevista) es muy limitada, por lo que no se puede evaluar la lengua oral del candidato, ya que no constituye una muestra amplia del constructo a evaluar. Además, el evaluador marca una relación de poder que determina la actuación del candidato, y en el formato de examen se deben maximizar las capacidades de los candidatos y no limitarlas, como ocurre con la entrevista en donde se maximiza el rol del examinador.

Si aún así se quiere mantener este formato de entrevista para evaluar la expresión oral, ésta “debe estar estructurada cuidadosamente para que los aspectos de la prueba considerados importantes se cubran con cada estudiante, y cada estudiante sea evaluado de forma similar” (Alderson *et al.* 1998: 64) además, Bordón (2004a: 995) considera que si los entrevistadores-examinadores están entrenados en su técnica y la entrevista está muy bien estructurada y experimentada no tiene que plantear problemas. Lo que debemos garantizar es que los candidatos usen la lengua dentro de una misma situación

de dificultad, es decir, un candidato no puede limitarse a hacer preguntas simples mientras otro tiene que usar un lenguaje más complejo que puede resaltar sus deficiencias en la lengua a evaluar.

A la vista de los resultados de su investigación, Figueras (2004: 81) advierte que, en consonancia con lo que propone el MCER,

si los planteamientos de evaluación de la lengua oral incluyen tareas que generan tipos de discurso monológico y dialógico, actividades con distintos objetivos comunicativos y en distintos ámbitos, podremos emitir juicios sobre las competencias orales de los alumnos de forma mucho más válida y fiable,

es decir, que existen otras formas para evaluar la expresión oral, otros tipos de tareas que reflejan de una manera más fiable la capacidad lingüística del candidato.

No obstante, consideramos que la entrevista sería más relevante como introducción a la tarea o tareas de evaluación de la expresión oral, como una fase previa a ésta. De esta manera, el candidato puede sentirse más cómodo y reducir la tensión y el nerviosismo producidos en estas situaciones de examen. En este caso, la entrevista será concebida como una toma de contacto con el candidato y nunca basaremos en ella el resultado de la actuación de la persona evaluada; “de este modo se contribuye a la aceptabilidad de la tarea, lo cual incrementa el índice de fiabilidad de la prueba” (Bordón, 2006: 215).

Una vez que ya sabemos el tipo de actividades que podemos utilizar para elaborar una prueba de expresión oral, debemos determinar otros aspectos esenciales en relación con la elaboración de la tarea en sí misma. En este sentido, Bordón (2004a: 994, 2006: 205-206) considera que en las pruebas de expresión oral destinadas a evaluar la capacidad lingüística del candidato es aconsejable que tengan un apoyo gráfico que dependerá del nivel del idioma que se evalúa (una foto para una descripción física (nivel inicial), una secuencia de viñetas para relatar una historia (intermedio) o una secuencia de vídeo para proporcionar un resumen y/u opinión de ésta (avanzado)). Según Bordón este tipo de actividades con apoyo gráfico en las que no hay interacción reducen la tensión en el examen, por lo que se pueden concebir como actividades preparatorias. Por otra parte, en tareas de interacción debe haber un intercambio de vacíos de información donde un alumno posee información que el otro no conoce y tienen que negociar para poder entenderse y llegar a un acuerdo; simulaciones similares que se suelen producir en clase en donde se producen situaciones inventadas, similares a las de la vida real y en las que los alumnos deben interactuar; o por último, representación de papeles (*role playing*) en la que los alumnos reciben una ficha con una personalidad determinada y deben actuar con su interlocutor de acuerdo con el papel asignado. En este caso, se debe considerar

que en pruebas de alto impacto, si los candidatos no se conocen previamente, se puede ver afectada la validez de ésta, ya que existen factores como desconocimiento de la actividad de examen o características personales como la nacionalidad, cultura, edad o sexo que pueden llevar a no recoger una actuación lingüística adecuada. Por ello, Bordón advierte que en estos casos es mejor que la interacción se haga con un examinador.

Por su parte, Norris (1998, citado en Figueras, 2004: 79) incluye otros aspectos para la elaboración de la prueba de expresión oral:

- El número y el tipo de interlocutores, ya que esto determina la actuación dependiendo de si la persona evaluada tiene que interactuar con uno o más candidatos o con el profesor.
- La cantidad y el tipo de input, pues según cómo se presente la información de la tarea (en la lengua materna del candidato o en lengua extranjera), ésta puede modificar el discurso generado.
- La familiaridad con el contenido de la tarea y con los interlocutores.
- El tiempo de planificación.
- El tipo de intervención prevista: planificada previamente o espontánea.
- El objetivo comunicativo: según la situación que se plantee, por ejemplo, si hace falta llegar a un acuerdo o no en la interacción.

Igualmente, Vázquez (2000: 54) incide en que una prueba que evalúe la expresión oral debe ser planteada como un ejercicio comunicativo y para ello se debe producir una interacción en la que intervengan como mínimo dos personas, además, debe quedar claro quién habla con quién y para qué -es lo que denomina “explicitación”-, y, al igual que Bordón, considera que debe existir un vacío de información o de opinión, es decir, la respuesta que se desea obtener no debe ser conocida sino inesperada, y contener un mínimo de novedad para el interlocutor; por último, hay que tener en cuenta si hay más de una sola respuesta o no (predicción nula), que en el caso de la expresión oral se refiere, por ejemplo, si los interlocutores deben llegar a un acuerdo o no, si hay que aportar una solución a un problema, etc., como ya nos advertía Norris (objetivo comunicativo).

Cuando una persona se expresa oralmente en interacción, como señalaba el MCER, intervienen muchas variables y para evaluar esa interacción es indispensable incidir en ellas. Además de las especificadas en el MCER, Luoma (2009: 132-133) nos ofrece un ejemplo de criterios a tener en cuenta en una prueba de certificación (*A general purpose*

proficiency test) y plantea que la capacidad para interactuar de manera efectiva conlleva también la capacidad para interactuar en distintos contextos, para lo cual hay que observar el grado de formalidad, las convenciones de cortesía o la utilización del registro (formal o informal). También supone evaluar la claridad en la presentación (coherencia y cohesión) y la capacidad para expresar lo que se pide de manera correcta. Esto significa que el candidato debe demostrar, además, que tiene un buen conocimiento del vocabulario y las estructuras de la lengua, capacidad para acceder a los conocimientos relativamente rápido, una pronunciación comprensible y capacidad para comprender y reaccionar al discurso del interlocutor. Si alguno de estos aspectos estuvieran por debajo de lo que se considera adecuado para el nivel, la actuación de la persona evaluada se consideraría negativa. A la vista de estas consideraciones, Luoma propone evaluar de manera analítica todos los aspectos que forman parte de la competencia comunicativa y, además, advierte que aunque la competencia estratégica no se evalúe de forma directa, sí puede influir en la calificación porque interviene en la realización de la tarea.

Como se puede comprobar, la mayoría de estudios coincide en que la interacción es la mejor forma de evaluar la expresión oral ya que es la forma más característica en que las personas se relacionan y comunican. Pero debemos destacar que aunque se ha diferenciado entre las tareas monologadas, que llamaba el MCER de expresión oral, y las tareas interactivas, de interacción oral según el MCER, las monologadas también son interactivas, ya que si bien es verdad que, por ejemplo, en una conferencia la única persona que tiene el derecho y el deber de hablar es quien pronuncia el discurso, la audiencia que lo escucha interactúa con el ponente mediante toda una serie de aspectos no verbales que afectan al conferenciante. La diferencia más notoria entre ellas es que unas son más típicas o informales, las dialogadas o “plurigestionadas” (Calsamiglia y Tusón, 2007) que suelen ser más o menos espontáneas; y otras son más formales y preparadas, las monologadas o “monogestionadas”, pero ambas se dan en interacción. De esta forma, Calsamiglia y Tusón (2007: 18) señalan una serie de rasgos que se suelen dar en las producciones orales más cotidianas, rasgos que deberemos advertir en nuestras tareas de evaluación de la expresión oral para que la simulación producida en el examen sea lo más real posible.

La situación de enunciación oral prototípica se caracteriza, básicamente por los siguientes rasgos:

- En primer lugar, por la *participación simultánea* de las personas que intervienen en ella. Más que emisores y receptores, es preferible o más ajustado referirnos a ellas como

interlocutores.

- En segundo lugar, por la *presencia simultánea* de quienes interactúan, se comparte el espacio y el tiempo, los interlocutores participan **cara a cara**.
- En tercer lugar, porque los interlocutores activan, construyen y negocian en la **interacción** una *relación interpersonal* basada en sus características psicosociales: el estatus, los papeles o la imagen, por ejemplo.

Además de todos estos aspectos, la prueba de expresión oral al ser planteada como actividad comunicativa debe ser significativa, es decir, debe tener sentido para los participantes, por lo que es necesario que haya ese vacío de información al que aludíamos anteriormente. También debe ser auténtica para hacerla más cercana al alumno, abierta para que los candidatos puedan hablar y actuar con cierta libertad; es necesario que se incluya una retroalimentación significativa con el objeto de que los participantes sean informados de la causa del resultado de su actuación positivo o negativo, dosificar la dificultad en el desarrollo de la prueba dependiendo del nivel, siendo más recomendable administrar al principio las pruebas más sencillas y dejar para el final las más complicadas, y, por último, tener en cuenta la diferencia que puede existir entre los medios que posee el alumno para expresarse y lo que realmente quiere expresar, de ahí que sea importante en el aprendizaje del idioma introducir estrategias de comunicación (Robera Carreño, citado por Pinilla, 2004: 890)

Lo que resulta evidente es que todos estos aspectos que hemos señalado forman parte de la expresión oral entendida como acto comunicativo y tanto para diseñar, desarrollar como para evaluar las muestras de lengua de las personas evaluadas, debemos apreciar todas estas variables si queremos obtener pruebas de examen y resultados de calidad.

3.1.2. Criterios de evaluación de la expresión oral

Una vez que tenemos elaboradas las pruebas para evaluar la expresión oral de los usuarios y las hemos administrado, tenemos que valorar la habilidad lingüística comunicativa de los examinados dependiendo de los datos que se desprendan de ellas. El MCER (2002: 193-194) establece para la expresión oral catorce categorías cualitativas que debemos tener en consideración como son: estrategias de turnos de palabras, estrategias de colaboración, petición de aclaración, fluidez, flexibilidad, coherencia, desarrollo temático, precisión, competencia sociolingüística, alcance general, riqueza de vocabulario, corrección gramatical, control del vocabulario y control fonológico. Sin embargo, estas categorías deberían combinarse, reformularse o reducirse a un número más pequeño seleccionando aquellas que se adecuen más a las necesidades de los alumnos y al objetivo de la prueba, ya que se considera que el límite

máximo de categorías debería ser siete, puesto que por encima de este número se estima muy difícil de controlar por parte del examinador.

De esta forma, los resultados de una actuación procedentes de una prueba de lengua y la solvencia de estos se pueden medir dependiendo de las dudas que se tengan sobre las características de los candidatos, cómo se les va a evaluar, si las preguntas están relacionadas con las características de estos y, sobre todo, “inexactitudes en los sistemas de puntuación y variedad de interpretación y aplicación de estos por parte de los correctores” (Alderson *et al.*, 1998: 106).

En la evaluación de la expresión oral la puntuación que se utiliza es subjetiva, los evaluadores no pueden determinar su valoración en términos de verdadero o falso sino que deben valorar hasta qué punto la actuación de un candidato en relación con la tarea a desarrollar se puede considerar buena o mala. Para ello, se hacen necesarios unos criterios de evaluación llamados “plantillas de corrección” (*marking schemes*) o “escalas de valoración” (*rating scales*) que aunque existen algunas distinciones entre ellas, en español se suelen utilizar bajo la denominación de “criterios de evaluación” para ambas (*ibidem*: 105). Estos criterios contienen una serie de descriptores que entrañan una serie de afirmaciones sobre el tipo de actuación lingüística al que se refiere cada punto de la escala.

Esta escala en donde se incluyen los criterios de evaluación puede ser holística (*holistic scale*) o también llamada “escala de impresión general” (*impression scale*) y se utiliza cuando los evaluadores quieren determinar el resultado de la prueba de una manera rápida, sobre todo en exámenes de clasificación o en exámenes de progreso donde el evaluador tendrá más muestras de lengua que evaluar a lo largo del curso y, por lo tanto, no le hace falta ser tan específico en estos tipos de examen. Un ejemplo de escala holística es la siguiente⁹:

Cuadro 11.

18-20	excelente	Inglés natural con errores mínimos y realización completa de la tarea encomendada.
16-17	Muy bien	Más que un conjunto de frases simples, con buen vocabulario y estructuras. Algunos errores básicos.
12-15	Bien	Realización simple aunque precisa de la tarea encomendada con algunos errores serios.

⁹ Se toma aquí el ejemplo (en su versión traducida) que Alderson *et al.* (1998: 108) tomaron de UCLES (*International Examinations in English as a Foreign Language General Handbook*, 1987). Aunque hubiésemos preferido incluir aquí los baremos utilizados en la prueba oral de los DELE, estos no se encuentran aún publicados.

8-11	Suficiente	Razonablemente correcta aunque torpe y no comunicativa O tratamiento natural y justo del tema, con algunos errores.
5-7	Flojo	Vocabulario y gramática no adecuados al tema.
0-4	Muy flojo	Incoherente. Los errores muestran la ausencia de conocimientos básicos de inglés.

Pero, si lo que se quiere es juzgar varios componentes de una actuación lingüística por separado (gramática, vocabulario, registro, ortografía, coherencia y cohesión, etc.) se debe utilizar una escala analítica (*analytic scale*) en la que hay descriptores para cada componente de dicha escala. Este tipo de escala permite valorar con más detalle cada aspecto de la actuación lingüística pudiendo tener un mismo candidato puntuaciones distintas en dichos componentes determinando al final una puntuación global para la totalidad de la actuación o diferenciada en cada uno de sus componentes. A continuación mostramos un ejemplo de escala analítica¹⁰:

Cuadro 12.

TAREA 2 – 60 PUNTOS

CONSIGNA	Puntuación
Cubre ampliamente el objetivo de la comunicación, ajustándose al tiempo establecido y trata todos los puntos de la tarea con detalle.	10
Cubre el objetivo de la comunicación, ajustándose al tiempo establecido y trata todos los puntos de la tarea.	7
Consigue el objetivo de la comunicación de manera suficiente y trata la mitad, o más, de los puntos de la tarea.	5
El grado de cumplimiento de la comunicación es insuficiente y trata menos de la mitad de los puntos de la tarea.	3
No consigue el objetivo de la comunicación y trata pocos puntos de la tarea.	0
ORGANIZACIÓN Y COHESIÓN	Puntuación
Interactúa de manera natural, respeta las normas de comunicación y emplea marcadores variados del discurso de forma muy adecuada	10
Interactúa adecuadamente, tiene en cuenta las normas de comunicación, y hace un uso apropiado de los marcadores del discurso	7
Interactúa de manera coherente, tiene en cuenta la mayor parte de las normas de comunicación y hace un uso adecuado de los marcadores del discurso.	5
Interactúa de forma poco coherente y hace un uso limitado de los marcadores del discurso.	6
No interactúa adecuadamente, ni emplea de forma apropiada los marcadores del discurso.	0
LÉXICO	Puntuación

¹⁰ Esta escala es la que se utiliza en las EOI de Canarias para evaluar la prueba de interacción oral en español, recogida en la Resolución de 28 de julio de 2010 (BOC N° 167. Miércoles 25 de Agosto de 2010: 2224-2226).

Utiliza un vocabulario, muy variado y preciso, adecuado a la tarea.	12
Utiliza un vocabulario adecuado a la tarea, variado y preciso.	8
Utiliza un vocabulario adecuado a la tarea y suficiente.	6
Muestra un uso del vocabulario poco adecuado a la tarea y limitado.	4
Utiliza un vocabulario inadecuado y muy limitado para la tarea.	0
PRONUNCIACIÓN	Puntuación
La pronunciación es muy correcta y la entonación muy apropiada.	6
La pronunciación es correcta y la entonación apropiada.	4
La pronunciación y entonación son adecuadas.	3
La pronunciación es poco clara y la entonación poco apropiada.	2
La pronunciación es muy defectuosa y la entonación poco apropiada	0
FLUIDEZ	Puntuación
El diálogo se realiza con una gran fluidez y soltura, y con las pausas adecuadas.	6
El diálogo se realiza con fluidez y soltura, y las pausas apropiadas.	4
El diálogo se realiza con las pausas y la fluidez apropiadas.	3
El diálogo se realiza con escasa fluidez y pausas incorrectas.	2
El diálogo se realiza sin fluidez y con pausas incorrectas.	0
GRAMÁTICA	Puntuación
Utiliza correctamente una gran variedad de estructuras y formas gramaticales propias del nivel.	12
Utiliza correctamente estructuras y formas gramaticales propias del nivel.	8
Utiliza estructuras y formas gramaticales propias del nivel, de forma suficiente.	6
Utiliza de manera insuficiente las estructuras y formas gramaticales propias del nivel.	4
Utiliza un número muy limitado, y con numerosos errores, de estructuras y formas gramaticales propias del nivel.	0
REGISTRO	Puntuación
Se ajusta a la situación de comunicación durante toda la tarea y hace un uso correcto de los marcadores del contexto social.	4
Se ajusta a la situación de comunicación durante toda la tarea con algún error en el uso de los marcadores del contexto social.	2
No se ajusta a la situación de comunicación ni utiliza los marcadores apropiados del contexto social.	0

En esta escala las competencias que aparecen atienden a los siguientes aspectos para su evaluación (Resolución de 28 de julio de 2010 (BOC N° 167. Miércoles 25 de Agosto de 2010: 2221) :

Cuadro 13.

- Competencia pragmática

Consigna	Se refiere al grado de cumplimiento del objetivo de comunicación solicitado en la tarea: si se ha cumplido total o parcialmente o si no se ha cumplido en absoluto. Comprueba si se ha respetado la duración en la extensión de las intervenciones
----------	--

	que se solicitaba en la consigna de la tarea. Verifica si se han tratado todos los puntos de la tarea, así como la pertinencia del contenido del texto en conjunto o de determinadas partes de este.
Organización y cohesión	Se refiere al grado de adecuación en la organización del texto: la forma de presentar la información, la estructura interna, la ligazón entre las partes... Entre los mecanismos para la construcción textual, se prestará especial atención al empleo adecuado de conectores así como de los marcadores propios de la conversación.

- Competencia lingüística

Léxico	Verifica el grado de corrección en el vocabulario empleado, su variedad y la precisión del léxico, siempre según la tarea propuesta y dentro del nivel de competencia evaluado.
Pronunciación y fluidez	Se refiere a la corrección en la articulación de los sonidos, cuestión que se ponderará siempre en relación con el nivel evaluado y con el principio de efectividad en la comunicación. Se evaluará también la adecuación de la entonación, el ritmo y las pausas en función de la modalidad discursiva, la tipología textual y la función comunicativa.
Gramática	Se refiere al grado de corrección de todos los elementos relacionados con la morfología y la sintaxis: orden de los elementos, concordancias, formas y estructuras... En el análisis de los errores que la producción presente, será necesario considerar si estos corresponden al estadio de adquisición lingüística que se pretende certificar, si se derivan de la toma de riesgos y en qué grado obstaculizan la comunicación.

- Competencia sociolingüística

Registro	Se refiere a la consideración de cuestiones tales como la situación de comunicación (la real de la prueba y/o la creada a partir de las pautas consignadas en esta), el propósito de esta comunicación, los destinatarios, el mensaje y la relación -real o virtual- que les une a estas personas... Se deberá comprobar que estas variables han condicionado que, para el texto, se elija un determinado registro, estilo y nivel lingüístico, lo que se verbalizará a través de la utilización de unas formas marcadas sociolingüísticamente (en el léxico, en las estructuras...).
----------	---

Como se puede ver, en esta escala se sigue el modelo de competencia comunicativa que propone el MCER (2002) y no se da la misma puntuación a todas las categorías considerando que las más puntuadas tienen más importancia (léxico y gramática), es decir, aparecen ponderadas. No obstante, “la ponderación diferenciada de ítems raramente mejora la fiabilidad o la validez” (Alderson *et al.*, 1998: 146) ya que si una

parte vale más que otra sería más fiable que la parte que más vale tuviera el doble de ítems que la otra “que si se presentara el mismo número de ítems para las dos áreas y los que pertenecen a la parte más importante valen el doble” (Ebel (1979), citado en *ibidem*).

En otras ocasiones se mezclan las escalas analíticas con las holísticas; así, por ejemplo, se evalúa la gramática con una escala analítica y la consigna con un baremo holístico. Incluso, concretamente en la evaluación de las pruebas de expresión oral, algunas instituciones optan por que el evaluador-interlocutor ofrezca una visión general de la actuación del candidato, utilizando para ello una escala holística y el evaluador-examinador lo hace con más detalle, ya que no tiene que interactuar con el aspirante, mediante un baremo analítico.

En definitiva, es necesario decidirse por un sistema de evaluación y experimentarlo con evaluadores y para ello es conveniente la formación de estos.

3.1.3. Obtención y tratamiento de datos.

El registro de las muestras obtenidas a partir de las actuaciones de los candidatos es una forma de experimentar los criterios de evaluación con el fin de garantizar la fiabilidad en las pruebas de expresión oral. La forma más usual de registrarlas es mediante grabaciones de audio; así, tenemos resultados tangibles de las producciones de los alumnos con la finalidad de ser compartidas, observadas y evaluadas, así como para poder otorgar a la producción oral el mismo rango de objetividad y de importancia que se da tradicionalmente a las producciones escritas. Esto también facilita la labor de los evaluadores ante posibles reclamaciones, ya que contamos con un documento de la actuación del candidato, y, por otra parte, puede utilizarse para la formación de estos.

Siguiendo a Bordón (2004a: 999):

La única manera de asignar resultados fiables a pruebas abiertas que precisan calificación subjetiva, es que los evaluadores tengan claro el alcance de cada uno de los parámetros de actuación establecidos en las escalas que se adopten. Un estándar uniforme garantiza la objetividad de la calificación y, consecuentemente, la fiabilidad de la misma. No se contará con una herramienta fiable si no se forma a evaluadores que sepan utilizarla, y eso sólo se consigue con la práctica que lleve a la homogeneización de criterios.

Además, en exámenes de alto impacto, es necesario que se haga una doble corrección del examen, bien por parte del mismo evaluador o por parte de evaluadores diferentes y, en el caso de que hubieran muchas discrepancias en los resultados debe intervenir un tercero o un jefe evaluador para nivelar la calificación. En cuanto a las pruebas orales esto sólo es posible si se graban las producciones de los participantes.

No obstante, como ya hemos comentado con anterioridad, la evaluación sobre pruebas grabadas suelen tener resultados más altos que sobre pruebas directas ya que en las grabaciones de audio se pierden elementos paralingüísticos y extralingüísticos que también intervienen e influyen en la comunicación oral y que son tremendamente importantes para que la comunicación sea efectiva. Una forma de suplir esta “carencia” de las grabaciones de audio es sustituirlas por grabaciones de video donde sí se podrían apreciar todos estos elementos. Sin embargo, la posibilidad de poder grabar las actuaciones también depende de los medios logísticos disponibles y del consentimiento de los participantes.

3.2. La formación de los evaluadores

Todos los estudios consultados inciden en que la formación de los evaluadores es indispensable para que la evaluación se lleve a cabo con fiabilidad, viabilidad y validez. Si esta formación no existe, los evaluadores, al menos, deben ser personas experimentadas en la enseñanza de idiomas y en la evaluación de muestras de capacidad lingüística. Estos últimos, la mayoría profesores, para suplir su falta de formación como profesionales de la evaluación, sobre todo en lo que concierne a exámenes de alto impacto, pueden tomar como referencia el Código de Conducta o de Buena Práctica de ALTE¹¹.

Esta Asociación ha desarrollado un código deontológico que garantiza los derechos de los usuarios y los compromisos que adquieren con ellos los evaluadores, avalando la transparencia y calidad en la evaluación. Se pretende que los usuarios tengan la seguridad de que las pruebas para la obtención de un certificado de capacidad lingüística no sólo aseguran el nivel que determina el resultado de estas con total fiabilidad, sino también su calidad.

Así, ALTE nos ilustra sobre aspectos tales como quiénes forman parte del proceso de evaluación, qué obligaciones tienen los evaluadores y cuáles son las responsabilidades de los usuarios, así como también nos sugiere un proceso de introspección o

¹¹ Hay que recordar que además de ALTE existen otras asociaciones que hemos nombrado anteriormente y que también facilitan la formación de los evaluadores y, por extensión, la de los profesionales de la enseñanza de lenguas como son EALTA o ILTA. Nos hemos centrado en ALTE porque es la más antigua de las tres asociaciones, creada en 1990, y es la que más ha desarrollado su Código de Buena Práctica.

autoevaluación para elaborar y evaluar pruebas con rigor. Seguidamente nos proponemos explicar estos aspectos con más detalle.

En cuanto a las personas que forman parte del proceso de evaluación, distingue tres agentes: los creadores de los exámenes que los elaboran y los que los administran; los usuarios de estos, llamados usuarios secundarios, que son los que utilizan los resultados de los exámenes para tomar decisiones profesionales y/o académicas; y por último, los candidatos o usuarios primarios que son los que se presentan a los exámenes.

Una vez determinados los individuos que intervienen en la evaluación, ALTE nos proporciona un Código de Conducta que trata las obligaciones de los entes certificadores en cuatro ámbitos:

- Utilidad: hay que informar a los usuarios de cómo es el tipo de prueba, qué se evalúa en ella y su utilidad. Para ello, hay que tener en cuenta el contexto y el tipo de alumnado al que va dirigida la prueba, el modelo de lengua a evaluar, los contenidos y los criterios de evaluación que se usan.
- Transparencia: el usuario debe saber cómo se evalúan las pruebas y cómo se expresa el rendimiento de cada candidato.
- Equidad: deben de ser justas para todos los usuarios independientemente de su género, raza, religión, lengua materna o discapacidades que puedan obstaculizar la realización de las mismas. Debemos asegurarnos que las diferencias de rendimiento de los candidatos se deban a la evaluación de la destreza y no a influencias de su raza, género o religión.
- Derechos y responsabilidades: los usuarios deben conocer los resultados de sus exámenes, deben saber que tienen derecho a la revisión de estos, a la reclamación de los resultados, a la protección de sus datos personales, al periodo de tiempo estipulado para la destrucción de los exámenes y a la validez de las certificaciones expedidas.

Además, ALTE también advierte de las responsabilidades de los usuarios en cuatro apartados: selección de exámenes adecuados relacionados con el tipo de candidatos a los que se dirigen, interpretación de los resultados, compromiso de equidad y justicia e información para los usuarios.

Igualmente nos expone un proceso de autoevaluación e introspección que permite una reflexión y concienciación sobre la elaboración, administración y corrección de pruebas con el fin de establecer mejoras. Nos señala cuatro aspectos de la acción introspectiva:

- Diseño y confección de pruebas siguiendo procesos estandarizados de elaboración de pruebas basadas, en su mayoría, en enfoques comunicativos.
- Administración de exámenes.
- Procesado de la documentación mediante la corrección, calificación y comunicación de resultados. En este aspecto ALTE incide en que es fundamental la formación de los evaluadores.
- Análisis y revisión sobre el rendimiento de las pruebas y revisión de contenidos para adaptarlos al MCER.

Todas estas directrices nos sirven de apoyo para tener una idea más clara de lo que se debe hacer en relación con la elaboración de pruebas y cómo se debe evaluar. En este sentido, si nos ceñimos a la evaluación de la expresión oral los aspectos que hemos señalado acrecientan su relevancia en el caso de la formación de los evaluadores de esta destreza, ya que no sólo deben atender a las variables ya mencionadas sino que también tienen que interpretar los descriptores de evaluación de manera coherente y comprender los principios que subyacen en ellos. Así, se hace necesario que los evaluadores de las pruebas de expresión oral graben las actuaciones para que cuando las puntúen y se pongan en común los resultados con otros evaluadores, puedan volver a escuchar la actuación del candidato para exponer las puntuaciones de forma independiente, discutir las notas, modificar la escala si hace falta y dar una valoración consensuada. Para un profesor de idiomas que no esté acostumbrado a este tipo de evaluación de certificación, si no tiene la formación adecuada, puede pasar por alto estas consideraciones; de ahí que sea tan necesario el adiestramiento de esta labor.

Además, en ocasiones, el evaluador debe interactuar con el candidato, aunque si hay más de un evaluador lo conveniente es que uno actúe como interlocutor (evaluador-interlocutor) dedicándose sólo a conversar con el aspirante, y el otro se limita a evaluar la actuación (evaluador-examinador). Esto disminuye la tensión de los candidatos a la vez que quien evalúa se puede centrar en tal fin sin necesidad de mirar o asentir, es decir, interactuar con la persona evaluada. De todas formas, aunque el que asume el papel de interlocutor sólo interactúe con el candidato, también se está formando una valoración de la actuación de éste. Por ello, lo consideramos también evaluador, al igual que al que examina.

La formación de los evaluadores permite que estos sepan cómo actuar con los aspirantes en cada situación, dónde sentarse según su rol en la evaluación de la expresión oral, cómo anotar puntuaciones de forma discreta, etc. Pero si bien la formación de los

evaluadores es necesaria, también lo es la de los administradores de la prueba, que en el caso de las EOI suelen coincidir en una misma persona: el profesor. Se puede considerar administrador de la prueba de expresión oral, en el caso de que hayan dos evaluadores para una misma prueba, al evaluador-interlocutor, que tiene que asegurarse de que los candidatos entiendan la tarea y sepan lo que deben hacer. Asimismo, deberá saber cómo hacer que estos se sientan más cómodos o cómo actuar para introducir en la conversación a los que todavía no han hablado. Todo ello exige una formación al respecto. Según Alderson *et al.* (1998: 116) “se necesitará medio día más de formación para los examinadores que tendrán que interactuar con los candidatos y medio día como mínimo para la formación de los profesores que actuarán como interlocutores mientras que otro hace la función de examinador”. También señalan a otro tipo de administrador de las pruebas de expresión oral como es el “acomodador” (*ibidem*: 116), quien se tiene que asegurar de la correcta entrada y salida del aula o el que los examinados no se comuniquen con los que aún no han hecho la prueba. Por último, además del “acomodador” advierten la figura del “vigilante”, que es el responsable de la distribución y recogida de los exámenes, de la administración del tiempo o de que los candidatos no se ayuden durante el examen.

Como podemos ver, aunque los evaluadores no tengan formación específica en estos aspectos, la existencia de un código deontológico, su propia experiencia y su ética profesional puede llevar tanto a la realización de pruebas de evaluación de calidad como a una corrección fiable de estas. Y es que, además, este último aspecto, la ética, es básico en la evaluación y no lo podemos pasar por alto.

3.3. Dimensión ética de la evaluación

La posibilidad de éxito que tiene un profesor de idiomas de conseguir un proceso de evaluación y unos resultados fiables es totalmente factible si se tienen en cuenta los aspectos mencionados anteriormente, pero esta labor no es sólo del profesor, sino que exige un compromiso por parte de todos. “Se trata de redistribuir y de compartir el poder, la riqueza, las responsabilidades, el acceso a las oportunidades y al conocimiento. Es preciso, en definitiva, incluir la dimensión ética de la evaluación en el orden del día de las sociedades democráticas modernas” (Parrondo, 2004a: 43). Parrondo hace extensiva esta dimensión ética a toda la sociedad.

Desde este punto de vista, muchas instituciones educativas como, por ejemplo, el Instituto Cervantes en su empeño por ser lo más transparente posible en el proceso de

evaluación, ha aplicado el Código de Conducta desarrollado por ALTE en sus exámenes. La información sobre las características de estos es accesible a todos los usuarios a través de su página web¹² en donde se muestra toda esta información e, incluso, existen muestras de exámenes de años anteriores.

El acceso a estos aspectos relevantes de la evaluación de exámenes hace que la cuestión ética esté presente en estas instituciones, ya que permite a los usuarios reclamar el resultado de un examen, es decir, pedir cuentas, (lo que llama McNamara (citado por Bordón 2004b: 24) *accountability*), asegurar dichas instituciones que la repercusión de dicho examen (*washback effect*) sea positiva y tener en cuenta el impacto de éste, es decir su efecto en la sociedad. En este sentido, McNamara (*ibidem*) observa la responsabilidad ética en la evaluación de la lengua como una responsabilidad social.

Otros consideran que la cuestión ética en la evaluación debe estar relacionada con la responsabilidad que conlleva la práctica de una profesión dedicada a la enseñanza. Así Bachman (citado por Bordón, 2004b: 23) consideraba que la ética de los evaluadores conseguida a través de la profesionalización de estos y de la existencia de un código deontológico como el que propone ALTE puede evitar algunos problemas observados en la evaluación de la expresión oral.

Siguiendo esta idea, Stanfield (1993: 190) une la profesionalización de los profesores y la publicación de un código deontológico para estos uniendo profesionalización a ética. Así, los examinadores abordarán con éxito la cuestión ética si tienen un código de ética profesional así como una guía para elaborar, administrar y calificar pruebas para la evaluación de una lengua.

En las EOI de Canarias también se ha intentado cumplir con todos estos aspectos relevantes sobre la claridad de la evaluación y, así, aparte de estar publicada toda la información relevante a los exámenes de certificación en el Boletín Oficial de Canarias (BOC), en concreto, la EOI Santa Cruz de Tenerife ha administrado tanto a los alumnos oficiales como libres un cuadernillo informativo sobre estas pruebas que también ha sido publicado en su página web¹³.

A la vista de todos estos intentos para que la ética esté presente en la evaluación, se puede considerar que hemos avanzado muchísimo a tenor de las investigaciones

¹² Se puede consultar toda la información relativa al Instituto Cervantes y a sus exámenes en www.cervantes.es.

¹³ La página web de la EOI Santa Cruz de Tenerife es www.eoisantacruzdetenerife.com

consultadas sobre este tema aunque todavía nos queda camino por recorrer, ya que aunque es verdad que se han formado muchos expertos en el campo de la evaluación y que, por lo tanto, existen profesionales dedicados a esto, en el caso de las EOI todavía no se ha producido la profesionalización de los evaluadores en pruebas de certificación, por lo que la ética desde el punto de vista de la práctica profesional sí está presente como enseñantes. Sin embargo, a pesar de que ésta se intente, en el ámbito de exámenes de alto impacto, no se puede asegurar ya que no existe una formación específica para el profesorado ni tampoco un código deontológico en las EOI para tal fin relacionado con el tipo de prueba, con el alumnado y con el contexto de realización de los exámenes.

3.4. Conclusiones

Las investigaciones que giran en torno a la evaluación coinciden en la importancia de ésta dentro de la adquisición de una lengua tanto por motivos prácticos como por motivos pedagógicos, aunque el límite que separa la objetividad de la subjetividad en una evaluación sea casi invisible, como ya advierte el MCER (2002:189):

toda evaluación debería ser lo más objetiva posible. Las consecuencias de las valoraciones personales que conllevan las decisiones subjetivas respecto a la selección de contenido y la calidad de la actuación, se deberían reducir tanto como fuera posible, (...). Esto es debido a que los resultados de las pruebas, muy a menudo, se utilizan por terceros para tomar decisiones respecto al futuro de las personas que han sido evaluadas.

En conclusión, para poder garantizar esta objetividad o, por lo menos, una “puntuación objetiva” es necesaria la elaboración rigurosa de pruebas orales comunicativas relacionadas con el objetivo de dichas pruebas y con las características de los alumnos a las que van dirigidas, el desarrollo de ítems adecuados y de exámenes o herramientas de evaluación viables, la definición consensuada del constructo a evaluar por los examinadores así como la formación de estos y su responsabilidad o ética profesional para hacer de la evaluación una herramienta útil y efectiva.

4. ESTUDIO EMPÍRICO

4.1. Planteamiento del estudio empírico

Como ya hemos explicado al principio del presente trabajo, la idea de llevar a cabo esta investigación surge de la comprobación en el aula y en los exámenes de certificación que año tras año hemos administrado en la EOI Santa Cruz de Tenerife de que los resultados de las pruebas de expresión oral, fundamentalmente, están determinados por la subjetividad de los evaluadores y no por criterios de evaluación bien definidos que les puedan otorgar a estas mayor objetividad. Aunque existan actualmente criterios de evaluación para estos exámenes, sus descriptores están formulados de tal forma que llevan a confusión, optando el profesorado, al final, por hacer una interpretación subjetiva de estos adaptándolos a sus necesidades. Todo ello, unido a la falta de preparación por parte del profesorado en relación a la evaluación de exámenes a gran escala, hace que nos cuestionemos la fiabilidad y validez de estas pruebas orales. De esta forma, partiendo de la pregunta ¿es posible elaborar criterios de evaluación mejor definidos para evaluar la expresión oral¹⁴ en las EOI de tal forma que se pueda eliminar o, al menos, disminuir la subjetividad en la evaluación?, es decir ¿cómo podemos evaluar de una manera más objetiva y precisa muestras de lengua que ofrecen unos resultados que deben ser interpretados? se proponen los objetivos de esta investigación ya señalados en el epígrafe 1.2. Hay que destacar que los criterios y las pruebas de evaluación que aquí se proponen están muy relacionados con el contexto de intervención del estudio.

¹⁴ Aunque “la evaluación no debería tener en cuenta principalmente ninguna actuación concreta, sino que debería intentar valorar las competencias generalizables que se muestran en esta actuación” (MCER, 2002: 180) nos hemos centrado sólo en la evaluación de la expresión oral por tres motivos: es la que más problemas presenta al evaluador debido al alto grado de subjetividad que usa en su evaluación cuando tiene que ofrecer un resultado acertado y lo más objetivo posible, la necesidad de acotar el estudio debido a su carácter empírico, y la ausencia de estudios a este respecto, centrados fundamentalmente en las otras destrezas.

4.2. Metodología y diseño

4.2.1. Paradigma de investigación

Debido a que los fundamentos teóricos en los que se basa nuestro trabajo ya se han explicado en los capítulos dos y tres, nos centraremos en explicitar las características de nuestro estudio.

En nuestra investigación nos centramos en las pruebas de certificación de español para extranjeros de las EOI de Canarias¹⁵, fundamentalmente en las tareas de la evaluación de dominio de la expresión oral de nivel avanzado¹⁶ de un grupo de alumnos que estudian el idioma en un contexto formal (EOI) y en situación de inmersión lingüística (Santa Cruz de Tenerife). Se compararon los resultados de estos alumnos, obtenidos con los nuevos criterios de evaluación elaborados a partir de los resultados de la experimentación llevada a cabo en clase durante el curso escolar, con los resultados de estos mismos alumnos en las pruebas de certificación obtenidos mediante los criterios de evaluación establecidos por la Consejería de Educación para determinar si, realmente, con nuevos criterios más claros y precisos se adecuaban los resultados a la actuación oral real de los alumnos y se disminuía la subjetividad que con los de la Consejería de Educación.

Básicamente se analiza lo siguiente:

- Si las tareas y los descriptores de las pruebas de certificación de la expresión oral actuales están bien elaborados y cumplen las reglas de fiabilidad, viabilidad y validez.
- Si esto no es así, mejorar dichas tareas y descriptores diseñando y aplicando criterios de evaluación fiables mediante representaciones cuantitativas o escalas que puedan transformar la actuación de un candidato en una puntuación que

¹⁵ Durante nuestro estudio las pilotamos mediante actividades similares y pruebas de clase de expresión oral, teniendo en cuenta que en la evaluación en clase tenemos más muestras de lengua de los alumnos y que en certificación sólo se tiene que evaluar la actuación desarrollada durante el examen. Sin embargo, los criterios de evaluación y las escalas desarrolladas en nuestra investigación se pueden utilizar también para evaluar actividades comunicativas de expresión oral en clase o las pruebas de progreso y dominio durante el curso.

¹⁶ Nos hemos centrado en el nivel avanzado por tres causas: porque dicho nivel nos proporciona muestras de lengua más amplias en las que basar los criterios de evaluación, porque en la Comunidad Autónoma Canaria no existe aún un currículo de nivel Avanzado, por lo que dificulta aún más la incertidumbre en la evaluación de este, y por la necesidad de acotar el campo de investigación debido al tiempo y recursos disponibles.

determine si éste posee el nivel de dominio de la prueba que se le presenta y en qué grado.

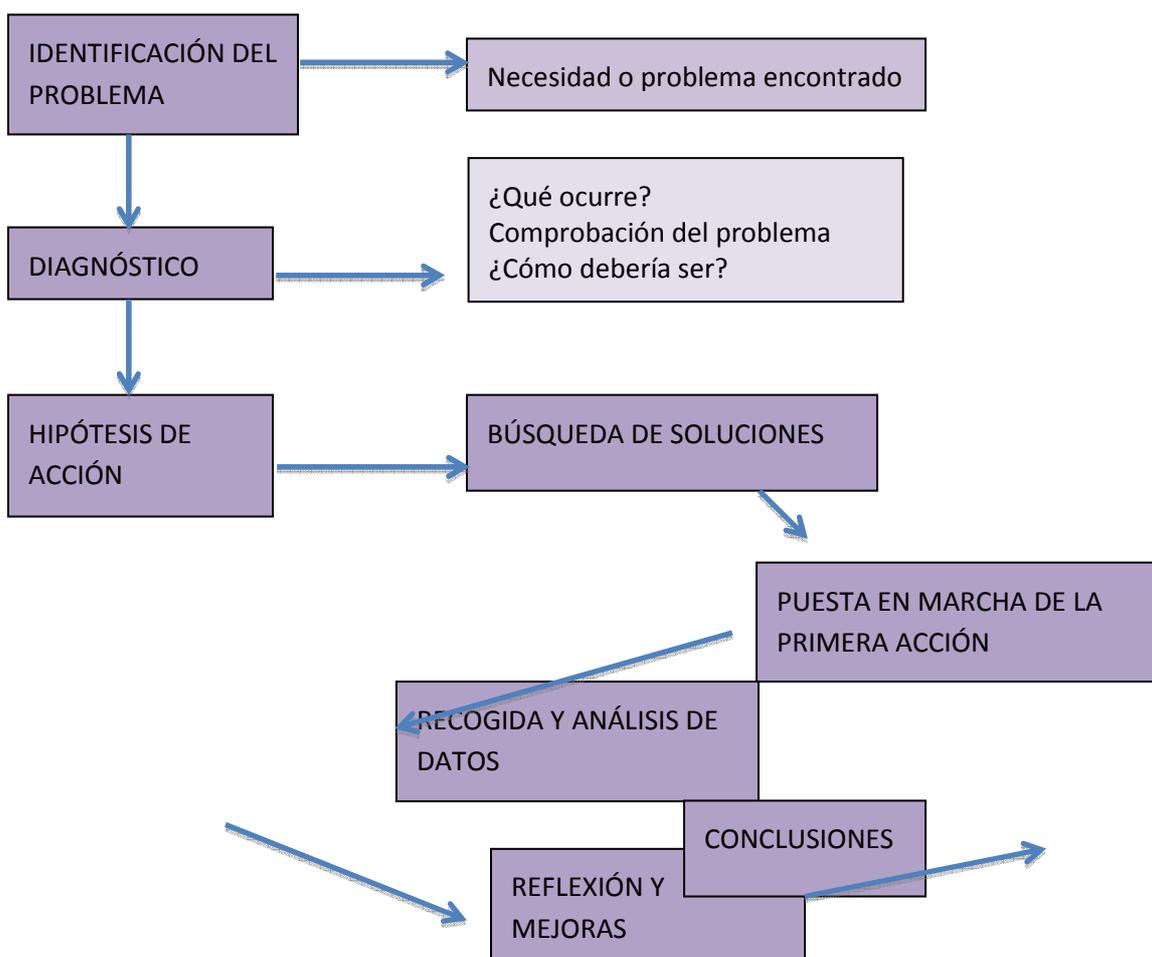
- Determinar cómo influye la corrección de los descriptores en la determinación del nivel de dominio de los candidatos.

La valoración de todos estos datos se llevó a cabo a través de la comprobación, análisis y aplicación de las pruebas de certificación de la expresión oral existentes y de sus criterios de evaluación y su comparación con las directrices del MCER en cuanto a la evaluación, de ALTE en relación con la elaboración de pruebas y criterios de evaluación y de la bibliografía consultada.

Posteriormente se crearon unas nuevas partiendo de estos modelos y se comprobaron las diferencias respecto a las anteriores. Se pretendía comprobar la validez, fiabilidad y viabilidad de los criterios de evaluación creados.

Se utilizó una metodología ecléctica que se adscribe dentro de la investigación acción definida por Elliott (1993: 88) como “el estudio de una situación social para tratar de mejorar la calidad de acción en la misma”, siguiendo un modelo basado en el carácter cíclico de la investigación: *planificar, actuar, observar y reflexionar*. De esta manera, se sigue el modelo planteado por Elliott (1993: 90) ya que se ajusta a las necesidades de nuestro estudio, puesto que permite, siguiendo este carácter cíclico, planificar, actuar, observar y reflexionar. Así, mediante la detección del problema y la observación de lo que ocurre en el aula y, por extensión, en los exámenes de certificación, se establece un plan de acción que se desarrolla con los estudiantes y cuyos resultados nos servirán de observación y reflexión para posibles mejoras, volviéndose a iniciar el ciclo. El esquema a seguir es el siguiente:

Cuadro 14.



Además, se persigue una investigación crítica para conseguir, según Kemmis (1984, citado por Latorre, 2008: 24),

una forma de indagación autorreflexiva realizada por quienes participan (profesorado, alumnado, o dirección por ejemplo) en las situaciones sociales (incluyendo las educativas) para mejorar la racionalidad y la justicia de: a) sus propias prácticas sociales o educativas; b) su comprensión sobre las mismas; y c) las situaciones e instituciones en que estas prácticas se realizan (aulas o escuelas por ejemplo).

Se intenta, por lo tanto, una mejora de la praxis mediante una visión crítica de ésta. Para ello, los datos obtenidos son tanto de tipo cualitativo como cuantitativo. Además, para la elaboración de exámenes de expresión oral, de sus descriptores y la evaluación de estos además de seguir las características propuestas por Bachman y Palmer (1996) de fiabilidad, validez, viabilidad, interactividad, autenticidad e impacto, de las que hemos hablado anteriormente, también se tuvo en cuenta el modelo propuesto por Sari Luoma (2009: 6), quien observa el proceso de elaboración de exámenes y sus criterios en estadios o ciclos en donde las puntuaciones de usos y de necesidades, el desarrollo de

estas puntuaciones y criterios (simbolizado por Luoma en una fábrica en donde la calidad debe estar muy relacionada con la planificación y el desarrollo), la administración de la prueba y, por último, la nota deben estar íntimamente ligadas. Además, la administración y desarrollo de las pruebas debe ser triangular (tareas, candidatos e interlocutores-examinadores) al igual que la nota (rendimiento, calificación y criterios) para garantizar su validez y fiabilidad.

De la misma manera, se intentaron seguir los pasos para elaborar pruebas de examen con rigor siguiendo algunas aportaciones de expertos en la materia como las que propone Bordón (2006) o Yagüe (2010), además de las directrices de ALTE y del MCER y se añadieron cambios y nuevas propuestas a la luz de los resultados, por lo que se pretendió demostrar la validez y fiabilidad de las que existen y se propusieron otras nuevas dependiendo de las necesidades.

Nuestra investigación se llevó a cabo durante el curso académico 2010/2011¹⁷ y se dividió en cuatro ciclos de investigación para permitir la puesta en práctica de la acción y el análisis de las diferentes variables objeto de estudio.

4.2.2. Variables de nuestra investigación

Aunque en esta investigación nos centraremos fundamentalmente en las variables de producto, debido al tema objeto de estudio, es indispensable tener en cuenta otras variables que también intervienen en nuestra investigación, ya que la planificación de la investigación y la experimentación de ésta se ha llevado a cabo durante los tres primeros ciclos de investigación con los alumnos de clase para luego, en el cuarto ciclo, desarrollarlo en las pruebas de certificación de junio del último curso escolar 2010/2011 con estos alumnos oficiales y analizar los resultados a lo largo del curso (enero-mayo) con el fin de establecer correlaciones. Igualmente, como ya se ha dicho anteriormente, se contrastan los resultados con los de las pruebas de certificación mediante los criterios de evaluación establecidos por la Consejería de Educación para comprobar la falta de validez de estos últimos en relación con la producción oral real de los candidatos frente a los elaborados durante este estudio empírico y utilizados en clase. Así, al comparar los resultados entre ellos se puede determinar la fiabilidad y validez de las pruebas y los criterios de evaluación de estas, puesto que si los resultados desprenden que no existe una notable diferencia entre las pruebas realizadas a lo largo del curso y las de

¹⁷ Desde el 14 de setiembre de 2010 hasta el 20 de junio de 2011.

certificación calificadas con los nuevos criterios de evaluación se confirmaría la validez y fiabilidad, significando esto, que a pesar de conocer el profesor-evaluador a los alumnos oficiales y de haberlos “adiestrado” en este tipo de prueba, ésta mediante los criterios de evaluación aplicados mide con objetividad el dominio lingüístico de estos no viéndose el resultado aumentado subjetivamente, ya que no difiere de su habilidad lingüística comunicativa real observada en el aula y, por lo tanto, se disminuye la subjetividad en la evaluación que es lo que nos proponemos.

4.2.2.1. Variables de contexto

Nuestra investigación se desarrolla en la EOI Santa Cruz de Tenerife situada en las Islas Canarias en un grupo de diez alumnos oficiales de nivel Avanzado 2 (B2) entre 18 y 35 años que estudian el español en inmersión lingüística. Los alumnos oficiales son de distinta nacionalidad y sexo: seis mujeres y cuatro hombres de entre 18 a 35 años, hay dos rusos, dos polacos, un lituano, un checo, un marroquí, dos americanos y un italiano. A continuación explicaremos brevemente cómo funcionan la enseñanza y la evaluación en este tipo de institución educativa para poder entender mejor nuestra investigación.

4.2.2.1.1. EOI Santa Cruz de Tenerife

La EOI Santa Cruz de Tenerife es un centro educativo que depende de la Consejería de Educación, Cultura, Universidades y Deportes del Gobierno de Canarias. Por tanto, es un centro público cuyo objeto es la enseñanza de idiomas. Los idiomas ofertados son: alemán, árabe, chino, español para extranjeros, francés, inglés e italiano, en horarios de mañana, tarde y noche. Es el mayor centro de enseñanza de idiomas en Santa Cruz de Tenerife con una matrícula de 4500 alumnos durante el último curso escolar 2010/2011. En lo que se refiere al español para extranjeros, además de la EOI, únicamente alguna organización no gubernamental (Cruz Roja, Cáritas...) ofrece clases de este idioma. También se imparten clases en alguna academia privada de idiomas. No obstante, es la EOI la que cuenta con el mayor número de estudiantes de español como lengua extranjera, y la única que imparte clases y expide títulos oficiales de español en la provincia de Santa Cruz de Tenerife.

Además de los puestos directivos (Dirección, Vicedirección, Jefatura de Estudios y Secretaría de Dirección) y del personal no docente (Secretaría, Conserjería y Mantenimiento), la Escuela se organiza en Departamentos, existiendo tantos Departamentos como idiomas ofertados. Al frente de cada Departamento hay un Jefe de Departamento, a excepción del Departamento de inglés que posee, además, un jefe de

departamento adjunto debido al gran número de profesorado (18) y alumnado (unos 3000) de inglés.

4.2.2.1.2. Organización de las enseñanzas oficiales de idiomas en la Comunidad Autónoma Canaria

Los estudios de las enseñanzas de idiomas en Canarias se estructuran siguiendo los niveles del Consejo de Europa concretados en el MCER (2002) y las directrices de la Consejería de Educación, Cultura, Universidades y Deportes de Canarias (BOC: 2007/205 - Lunes 15 de Octubre de 2007, Decreto 362/2007).

De esta manera, el estudio de los idiomas se distribuye en seis cursos escolares de septiembre a junio, con una carga lectiva mínima de 140 horas por curso escolar. Así, tenemos los siguientes niveles:

Cuadro 15.

NIVELES	CURSOS EN EOI	CORRESPONDENCIA CON EL MCER
Básico	Básico 1	A1
	Básico 2	A2
	Básico 3 (sólo árabe y chino)	-----
Intermedio	Intermedio 1	B1 (se podría decir B1.1)
	Intermedio 2	B1 (se podría decir B1.2)
	Intermedio 3 (sólo árabe y chino)	-----
Avanzado	Avanzado 1	B2 (se podría decir B2.1)
	Avanzado 2	B2 (se podría decir B2.2)

Las clases de cada curso tienen una duración de cinco horas semanales repartidas en dos horas por clase dos veces a la semana y otras dos horas los viernes alternos. A partir del mes de febrero se deja de impartir clase los viernes para dedicarlos a coordinación del profesorado y cumplir así las 140 horas de las que consta cada curso.

Al finalizar cada nivel, los alumnos tienen la opción de presentarse a un examen de certificación del nivel que han cursado (Básico, Intermedio o Avanzado), pero éste no es obligatorio para pasar de nivel. Si los alumnos han aprobado el examen de dominio de final de curso podrán pasar al siguiente nivel aunque no hayan hecho la prueba de

certificación. Además, también se pueden presentar a estos exámenes de certificación alumnos libres, es decir, alumnos que no asisten a clase sino que se preparan por su cuenta, y optar a la certificación de español del nivel correspondiente. Por lo que los profesores de EOI tienen una doble labor que es la de impartir enseñanzas del idioma de su especialidad y, además, certificar las competencias de este idioma.

4.2.2.2. Variables de presagio

4.2.2.2.1. Departamento de Español para extranjeros

En la actualidad, el Departamento de Español para extranjeros consta de dos profesores nativos. No ha habido un crecimiento de profesores proporcional al crecimiento de alumnos. Una de las profesoras tiene cierta permanencia en el Centro puesto que tiene una comisión por cargo directivo; el segundo profesor fluctúa con demasiada frecuencia. Es y ha sido la nota más negativa de este departamento, pues el segundo profesor generalmente viene al centro sin ningún tipo de experiencia en este tipo de enseñanza, lo que revierte en la calidad de ésta.

4.2.2.2.2. Tipo de alumnado que estudia español en la EOI Santa Cruz de Tenerife

La EOI se encuentra enclavada en uno de los lugares más céntricos de Santa Cruz, lo que permite un fácil acceso del alumnado español desde cualquier lugar de donde venga. En cambio, no ocurre lo mismo con los alumnos extranjeros, pues en su mayoría viven en pueblos y zonas de los alrededores de Santa Cruz. Tenemos alumnos que se desplazan desde puntos más lejanos, como Adeje, Garachico o El Puerto de la Cruz. Estas circunstancias han condicionado la organización de nuestro departamento, pues desde un comienzo los horarios lectivos han sido de dos horas alternas, se comienza siempre a las 9:00 horas y se finaliza a las 13:00 horas. De esta manera se ha posibilitado una mayor afluencia a clase de esos alumnos que viven tan retirados de Santa Cruz.

Sin embargo, el que no haya mucha oferta de cursos de tarde ha perjudicado mucho la potenciación de nuestro idioma. Los alumnos que están afincados en Tenerife -la mayoría- necesitan insertarse a través del idioma en nuestra sociedad y en el mundo laboral; el horario de mañana les resulta incompatible y, por eso, hay muchos abandonos a mitad de curso. Para ellos es imprescindible, como mínimo, comunicarse de forma oral con los demás, es decir, entender y ser entendidos. Durante este curso

escolar se imparte clase a dos grupos de Español por la tarde, que eran los que más solicitudes tenían en relación al horario de tarde, la acogida ha sido bastante buena.

En cuanto al número, hay que decir que de los cincuenta alumnos del comienzo de la enseñanza del español en nuestra EOI hace 25 años, hoy contamos con un número cercano a las dos centenas. Suelen ser muy numerosos el Básico 1, el Intermedio 1 y, en algunos casos, el Básico 2 y el Intermedio 2. El Avanzado 1 y 2 fluctúa según los años académicos. Hay que resaltar que el Básico 1 e Intermedio 2 se nutren casi siempre de chicas *au-pair* que no continúan sus estudios en la escuela, pues deben partir a sus países para continuar sus estudios universitarios.

En los últimos años ha habido una gran evolución en cuanto al tipo de alumnos. Hace veinte años Español se nutría de alumnos mayores, sobre todo jubilados y mayormente de nacionalidad alemana e inglesa que vivían en las islas. Hoy el panorama es totalmente diferente. Los alumnos suelen ser mucho más jóvenes, con verdadero interés por estudiar el idioma debido a razones laborales en el futuro, y son estos alumnos los que fundamentalmente se presentan a las pruebas de certificación. Si hace unos años la diversidad del alumnado por razón de su nacionalidad era bastante heterogénea, hoy este panorama está cambiando poco a poco. Hoy las clases de español se nutren de alumnos de Centro-Europa y chinos. No es el caso del Básico 1 del presente curso, donde hay unas diez nacionalidades: una gran parte de alemanes, ingleses, tailandeses, saharauis, hindúes, chinos, islandeses, italianos, franceses, etc. Estos alumnos viven en las islas o, en el caso de los más jóvenes, vienen por un trimestre o un curso escolar como *au-pair* o mediante una beca Erasmus para aprender mejor el idioma para su futuro profesional. Así, tenemos alumnos de muy diversas edades que van desde los 16 hasta los 75 años. Como se puede ver, es un alumnado heterogéneo, tanto en edad y nacionalidad como en intereses, que estudia el español en inmersión lingüística.

Para muchos de estos alumnos, el examen de certificación se convierte en una herramienta fundamental para acceder a un trabajo público o privado e, incluso, para estudiar en la Universidad; de ahí que sea imprescindible que los exámenes de certificación sean lo más válidos, fiables y viables posible.

Sin embargo, y aunque está aprobado por ley, no tenemos un curso específico para alumnos asiáticos y árabes, en el que el nivel Básico esté estructurado en tres cursos escolares (Básico 1, Básico 2 y Básico 3), como ya ocurre en los idiomas árabe y chino para españoles. Esto hace que el seguimiento de la clase para estos alumnos sea muy complicado, al tener que seguir el ritmo del resto, y provoca muchos abandonos;

igualmente influye en la evaluación, ya que son muy pocos los asiáticos que se presentan a los exámenes de certificación.

- Actitud en relación con la expresión oral.

La mayoría de los alumnos de nuestro estudio manifiestan abiertamente que lo que pretenden fundamentalmente en las clases de español es aprender a expresarse mejor de forma oral y prefieren, en relación a la metodología de enseñanza, que todas las actividades sean de expresión oral. Esta petición no se puede llevar a cabo debido a que el currículo de las enseñanzas de EOI exige la integración de destrezas y porque además nuestra metodología está basada en el aprendizaje de la competencia lingüística en su totalidad, lo que incluye, además de la expresión oral, la comprensión oral y la expresión y comprensión escrita. Por otra parte, manifiestan mediante un ejercicio de introspección que una de las causas que dificultan su expresión oral es el miedo al ridículo y a equivocarse.

4.2.2.3. Variables de proceso

4.2.2.3.1. Estructura de los estudios de español para extranjeros y metodología de enseñanza y evaluación.

Los estudios de español para extranjeros se organizan como los del resto de idiomas siguiendo las directrices del apartado anterior.

Normalmente casi todos los grupos tienen clase por la mañana de 9 a 11 o de 11 a 13 horas, y dos grupos reciben clases por las tardes, por demanda expresa del alumnado, de 16 a 18 horas y de 18 a 20 horas. Hay un grupo para cada nivel a excepción de Básico 1 en donde tenemos 2 grupos. Sin embargo, la demanda de los alumnos que quieren cursar estudios de español es mayor y a partir de noviembre, que es cuando llegan la mayoría de extranjeros a las islas, hay muchos alumnos que no obtienen plaza porque los grupos ya se encuentran llenos.

Las clases están formadas por alumnos de distintas nacionalidades que estudian el español en inmersión lingüística y, dado el contexto de enseñanza, se presta atención en la enseñanza del idioma al español meridional.

En relación con las enseñanzas de español por niveles, las enseñanzas del Nivel Básico, Intermedio y Avanzado tienen una duración de 280 horas, distribuidas en dos cursos. No obstante, el Decreto 362/2007 (BOC 2007/205), en su artículo 4 recoge que las

enseñanzas del nivel Básico del idioma español como lengua extranjera impartidas a alumnado de origen asiático o árabe tendrán una duración de 420 horas. Puesto que aún la Consejería de Educación no ha concretado este último término, la programación para el nivel Básico está pensada para ser desarrollada en unas 280 horas. Sin embargo, hay que tener en cuenta que el alumnado de origen asiático y árabe no puede seguir el ritmo de la clase como hacen el resto de las nacionalidades, y es por ello por lo que con este tipo de alumnado nos centramos en el nivel Básico 1 en trabajar la pronunciación, la escritura, la comprensión de textos básicos, así como la interacción en contextos muy sencillos, considerando que si alcanzan estos objetivos aprobarán el nivel Básico 1 a la espera de que la Consejería de Educación se pronuncie al respecto y apruebe las 420 horas de nivel Básico para este tipo de alumnado.

4.2.2.3.2. Metodología en las clases de español

Es necesario que determinemos la metodología a seguir por el Departamento de español, ya que en ella se encuentran las bases que determinan qué modelo de lengua evaluar y qué criterios se van a seguir para la evaluación de ésta.

Se parte de un modelo de lengua entendida como uso de ésta, tal y como aparece definida en el MCER (2002): aprendizaje, enseñanza, evaluación. Asimismo, los niveles previstos para estas enseñanzas se basan en los niveles de referencia del Consejo de Europa.

Los objetivos generales y específicos para cada una de las destrezas, la selección de los distintos tipos de contenidos y los criterios generales de evaluación que se llevan a cabo son igualmente coherentes con este enfoque.

Del concepto de lengua enunciado se deriva un modelo de competencia comunicativa lingüística que tiene una base fundamentalmente práctica, que se equipara con la capacidad de uso del idioma y que comprende tanto la competencia lingüística como las competencias sociolingüística y pragmática. Se trata de que el estudiante de lenguas, como futuro usuario del idioma, llegue a ser capaz de poner en juego las competencias mencionadas para llevar a cabo actividades lingüísticas mediante el procesamiento, receptiva y productivamente, de textos de distinto tipo sobre diversos temas, en ámbitos y contextos específicos, y mediante el desarrollo de las estrategias más adecuadas para llevar a cabo dichas actividades.

De estas competencias, las puramente lingüísticas son un mero vehículo para la comunicación y, en este sentido, el alumno las adquiere a través de actividades y tareas

en las que dichas competencias se ponen en uso. Del mismo modo, las competencias sociolingüística y pragmática se adquieren siempre a través de tareas en las que se utilizan materiales auténticos que les permiten entrar en contacto directo con los usos y características de las culturas en cuestión.

Las tareas de aprendizaje, diseñadas a partir de los objetivos específicos que se pretenden alcanzar, son también comunicativas para ofrecer al alumno adulto la oportunidad de utilizar el idioma tal como lo haría en situaciones reales de comunicación. En el contexto de unas tareas de estas características, los alumnos realizan actividades de expresión, comprensión o mediación o una combinación de ellas. Para llevar a cabo eficazmente las tareas correspondientes, los alumnos son entrenados en estrategias de planificación, ejecución, control y reparación, así como en una serie de procedimientos discursivos generales y específicos para cada destreza, y se les proporcionan unos conocimientos formales que les permiten comprender y producir textos ajustados a las necesidades que generan las situaciones de comunicación.

4.2.2.3.3. Proceso de comprobación, análisis y aprendizaje de la expresión oral de los alumnos

Dado que lo que pretendemos es que los criterios de evaluación y las pruebas de expresión oral de las que dependen estos criterios sean lo más objetivas, válidas y fiables posible en las pruebas de certificación, es indispensable experimentar las pruebas de expresión oral y los criterios de evaluación diseñados con los alumnos. Desde el primer contacto con los alumnos se observó que la capacidad para expresarse oralmente era suficiente, pero limitada para el nivel en el que estaban matriculados (B2), observación que también se había constatado en años anteriores. De los 10 alumnos, cuatro de ellos estaban en dicho nivel porque así lo había determinado el departamento por prueba de clasificación y el resto habían pasado de nivel una vez terminado el curso anterior (Avanzado 1). De los 10 en total, cinco de ellos tenían una producción oral muy limitada para el nivel. Esta producción oral se observó en la interacción en clase e incluso, los que habían promocionado de curso alegaron que en la certificación de intermedio obtuvieron una buena nota en expresión oral, por lo que se vio la necesidad de comprobar las pruebas y los criterios de evaluación de la expresión oral de certificación para comprobar su fiabilidad y validez y elaborar unos nuevos.

De esta manera, se ejecutó un plan de trabajo cíclico que corresponde a cada uno de los ciclos mencionados anteriormente, basados, una vez detectado el problema, en la comprobación, observación, diseño, desarrollo, ejecución y análisis.

De esta forma, en clase, en primer lugar se les pidió a los alumnos que hicieran un ejercicio de introspección para que nos dijeran qué dificultades observaban cuando se expresaban oralmente y casi todos contestaron que el miedo a equivocarse y el tener que traducir mentalmente del idioma materno al español es lo que les producía mayor dificultad. Una minoría expresó dificultad en cuestiones relacionadas con la gramática a la hora de hablar. A partir de estas impresiones se detectó una variable encubierta relacionada con la actitud de los alumnos a la hora de expresarse (miedo y frustración). Posteriormente se realizó un análisis de necesidades para comprobar por qué querían estudiar español y también un 70% de los alumnos alegó que lo que pretendían era expresarse correctamente de forma oral puesto que era muy necesario para ellos en su trabajo, que en su mayoría era de cara al público, es decir, esto determinaba lo que ya se había pensado: que las actividades comunicativas de expresión oral deben ser auténticas puesto que los alumnos quieren expresarse correctamente en su comunicación con los nativos en su vida diaria y que también había sido observado por Bachman y Palmer (1996), Bordón (2006) o el MCER (2002) entre otros. De esta forma, y una vez observadas las dificultades e intereses de los alumnos respecto a la expresión oral se comenzaron a diseñar actividades comunicativas de producción oral que nos servirían de experimentación para crear pruebas de expresión oral para exámenes, y en concreto, para certificación, con el fin de analizar el resultado y actitud de los alumnos en dichas pruebas y compararlos con resultados surgidos a lo largo del curso, así como para practicar dicha destreza en clase a partir de las dificultades y necesidades observadas anteriormente. En las actividades comunicativas diseñadas para clase y para las pruebas de certificación de la expresión oral se siguió la metodología señalada en el epígrafe anterior y se incluyeron los aspectos¹⁸ que numerosos lingüistas (Norris (1998), citado por Figueras, 2004), Vázquez (2000), Bordón (2004a y 2006)) o instituciones (Consejo de Europa (MCER, 2002), ALTE, Instituto Cervantes (PCIC)) han señalado como relevantes para que la prueba sea lo más eficaz posible.

¹⁸ Los aspectos relevantes a incluir en actividades comunicativas de expresión oral ya han sido recogidos en el epígrafe 3.1. del marco teórico.

A partir de la elaboración de estas pruebas se diseñaron criterios de evaluación teniendo en cuenta el constructo que se pretende evaluar¹⁹, las características de estos alumnos²⁰ y el contexto de intervención.

4.2.2.4. Variables de producto

Como ya hemos dicho, nuestra investigación se centra en las variables de producto aunque estas no pueden observarse individualmente de las anteriores ya que dependen de ellas. Nos referimos a que nuestro estudio se basa más en los resultados y no tanto en los procedimientos, es decir, en cómo medir esos resultados teniendo en cuenta las variables de la forma más rigurosa posible sin que pasen inadvertidas las peculiaridades de la capacidad comunicativa que medimos (expresión oral) para concluir con un producto objetivo²¹, fiable y válido.

En primer lugar explicaremos cómo se evalúa en la EOI Santa Cruz de Tenerife tanto en clase como durante las pruebas de certificación y explicaremos cómo se procede para la elaboración de estas pruebas de certificación. Creemos que estos aspectos son importantes porque permiten entender cómo están elaboradas las actuales pruebas de certificación y qué criterios se siguen para evaluarlas, para luego centrarnos en un epígrafe aparte, debido a su importancia y extensión, en los procesos llevados a cabo para la evaluación de la expresión oral en el presente trabajo.

4.2.2.4.1. La evaluación en la EOI Santa Cruz de Tenerife

Se siguen las directrices recogidas en la Orden de 17 de abril de 2009 (BOC 4 de mayo de 2009) por la que se regula la evaluación del alumnado que cursa enseñanzas de idiomas de régimen especial en la Comunidad Autónoma de Canarias. Asimismo se ha tenido en cuenta la clasificación que propone el MCER en su epígrafe 9.3 (2002: 183) y

¹⁹ Ver epígrafe 2.1.1. del estado de la cuestión.

²⁰ Cabe señalar que aunque nos hemos basado en las necesidades, dificultades y características de los alumnos oficiales para la elaboración de las pruebas y los criterios de evaluación de la expresión oral, estas necesidades y dificultades se pueden hacer extensibles -aunque sin llevar a generalizar- a los alumnos libres, puesto que en la práctica docente suelen observarse las mismas impresiones de los alumnos en cuanto a la dificultad de expresarse oralmente independientemente de la nacionalidad. Por ello, se ha comprobado la validez y fiabilidad de estas pruebas con estos alumnos y se han considerado que podrían también ser válidas y fiables con los libres.

²¹ Nos referimos a disminuir la subjetividad de los evaluadores dentro del proceso de evaluación. Sabemos que las pruebas de producción oral son subjetivas y que no se pueden medir por medios mecánicos para hacerlas totalmente objetivas como las de comprensión oral y escrita, pero sí creemos que se puede mitigar esa subjetividad.

se han llevado a cabo las que se han estimado más adecuadas al contexto, tipo de alumnado y necesidades de estos. De esta manera, en la EOI Santa Cruz de Tenerife se llevan a cabo distintos tipos de evaluación:

1- De clasificación, para situar al alumno que es nuevo en la Escuela en el nivel correspondiente. La realizan los profesores del departamento y se evalúan las cuatro destrezas (comprensión de lectura, comprensión auditiva, expresión e interacción escrita y expresión e interacción oral). El examen es cualitativo y no se le asigna puntuación sino sólo el nivel que tiene el candidato según los resultados de la prueba. Esta evaluación se realiza a principio de curso, y a veces, también en el transcurso del mismo durante el desarrollo de éste cuando llega un alumno nuevo.

De esta manera, quien haya superado por evaluación de aprovechamiento un nivel, aunque no haya superado la prueba de certificación correspondiente, podrá acceder al primer curso del nivel inmediato superior, estando exento de la prueba de clasificación.

Esta prueba consta de una autoevaluación del alumno y una prueba de dominio para cada curso, y en ella se miden las cuatro destrezas que se califican del uno al diez sin decimales; se hace la media aritmética y se redondea al valor entero correspondiente. Para hacer la media se tiene que obtener, al menos, un cuatro en tres destrezas. Se necesita como mínimo un cinco para aprobar. El modo de proceder es el siguiente:

- El alumno elige el curso al que quiere acceder y cumplimenta el cuestionario de autoevaluación. De obtener un resultado positivo, realiza la prueba de dominio correspondiente al curso anterior al elegido. Si el resultado es negativo, el alumno deberá replantear su elección.
- Se confecciona un acta para cada uno de los cursos.
- Una vez comenzado el curso, el profesor, en función de las capacidades del alumno y las plazas vacantes, puede reubicar al alumno en un curso superior o inferior, al asignado en la prueba de clasificación.
- Puede realizar la prueba aquel aspirante que no haya cursado estudios de EOI, o, si lo ha hecho, que haya transcurrido un año académico entre el curso en que estuvo matriculado y el de la prueba.
- La clasificación tendrá efecto sólo para el año académico correspondiente y no se certificarán los cursos anteriores.
- Si un alumno oficial del curso pasado quisiera acceder a un nivel superior que el que le corresponde, podrá renunciar a su plaza del nivel que está matriculado en los

plazos establecidos para tal fin, y se presentará por libre a la certificación que él considere del nivel superior.

- Si un alumno quisiera hacer la certificación para un nivel inferior, podrá hacerlo según directrices de la Consejería de Educación.

2.- De progreso, que proporciona datos, a lo largo del curso, sobre los avances en el aprendizaje del alumnado, así como del grado de consecución de los objetivos y contenidos programados.

Se extiende desde principio de curso hasta el momento de la realización de la prueba de dominio (mediados de enero y mayo). Según se recoge en la Orden de 17 de abril de 2009 (BOC 4 de mayo de 2009)

Artículo 7.- Administración de la evaluación.

7.1. La evaluación de progreso será realizada por el profesor de cada grupo y registrará los resultados obtenidos en las cuatro destrezas (comprensión de lectura, comprensión auditiva, expresión e interacción escrita y expresión e interacción oral).

3- De dominio, que mide la dimensión lingüística de la competencia comunicativa del alumnado y permite comprobar el nivel de dominio del idioma logrado al final de cada cuatrimestre. Esta prueba se realiza a finales de enero durante el primer cuatrimestre y a finales de mayo durante el segundo cuatrimestre. Es elaborada por los profesores del departamento y consta de cuatro partes independientes: comprensión de lectura, comprensión auditiva, expresión e interacción escrita y expresión e interacción oral.

3- De aprovechamiento: en la orden citada anteriormente se recoge lo siguiente sobre este tipo de evaluación:

Artículo 5.- Evaluación de aprovechamiento del curso y del nivel.

5.1. La evaluación final de aprovechamiento del alumnado, en cada uno de los cursos, tendrá dos componentes: la evaluación de progreso y la prueba de dominio.

(...)

5.2. La evaluación final de aprovechamiento del Nivel será la obtenida en el último curso de éste.

4- De certificación: los exámenes de certificación son exámenes oficiales de dominio a gran escala donde se determina si un candidato tiene la habilidad lingüística suficiente del nivel que se examina (Básico, Intermedio o Avanzado). Para diseñar, distribuir y evaluar este tipo de examen se siguen las directrices de la Consejería de Educación (BOC N° 167. Miércoles 25 de Agosto de 2010, resolución de 28 de julio de 2010) teniendo en cuenta el MCER y los currículos respectivos de cada nivel. A la vista de estas instrucciones podemos observar cuál es el objetivo de las pruebas de certificación

de español dependiendo del nivel en el que se examine el candidato (el resultado es nuestro):

a) Nivel Básico.

La prueba de certificación de nivel básico tiene como finalidad evaluar la capacidad del alumnado aspirante para usar el idioma de manera suficiente, receptiva y productivamente, tanto en forma hablada como escrita, así como para mediar entre hablantes de distintas lenguas, en situaciones cotidianas y de inmediata necesidad que requieran comprender y producir textos breves, en diversos registros y en lengua estándar, que versen sobre aspectos básicos concretos de temas generales y que contengan expresiones, estructuras y léxico de uso frecuente.

La posesión del Certificado del nivel básico acredita una competencia de idioma del nivel A2 del Marco común europeo de referencia para las lenguas: aprendizaje, enseñanza y evaluación (MCERL), del Consejo de Europa.

b) Nivel Intermedio.

La prueba de certificación de nivel intermedio tiene como finalidad evaluar la capacidad de la persona candidata para utilizar el idioma con cierta seguridad y flexibilidad, receptiva y productivamente, tanto en forma hablada como escrita, así como para mediar en la comunicación entre hablantes de distintas lenguas, en situaciones cotidianas y menos corrientes que requieran comprender y producir textos en lengua estándar, con estructuras habituales y un repertorio léxico común no muy idiomático, y que versen sobre temas generales, cotidianos o en los que se tiene un interés personal.

La posesión del Certificado del nivel intermedio acredita una competencia de idioma del nivel B1 del Marco común europeo de referencia para las lenguas: aprendizaje, enseñanza y evaluación (MCERL), del Consejo de Europa.

c) Nivel Avanzado.

La prueba de certificación de nivel avanzado tiene como finalidad evaluar la capacidad de la persona candidata para utilizar el idioma con soltura y eficacia en situaciones habituales y más específicas que requieran comprender, producir y tratar textos orales y escritos conceptual y lingüísticamente complejos, en una variedad de lengua estándar, con un repertorio léxico amplio aunque no muy idiomático y que versen sobre temas generales actuales o propios del campo de especialización y de interés del hablante.

La posesión del Certificado del nivel avanzado acredita una competencia de idioma de nivel B2 del Marco común europeo de referencia para las lenguas: aprendizaje, enseñanza y evaluación (MCERL), del Consejo de Europa.

(BOC N° 167. Miércoles 25 de Agosto de 2010, resolución de 28 de julio de 2010)

Además, determina las características generales de las técnicas e instrumentos de evaluación.

Las pruebas de certificación han de permitir la obtención de la información necesaria para conocer el dominio alcanzado en el idioma por la persona aspirante. Para ello se utilizarán técnicas e instrumentos viables, eficaces y fiables que han de cumplir, entre otras, las siguientes características:

- Comprender los objetivos y contenidos más significativos del currículo.
- Adecuarse a la naturaleza diversa de los contenidos curriculares.
- Permitir la graduación del dominio alcanzado en cada de las destrezas.
- Permitir la asociación de escalas numéricas a valoraciones cualitativas.
- Adecuarse al carácter propio de cada una de las destrezas.
- Medir el grado de dominio de los distintos aspectos de cada contenido planteado.

- Permitir el conocimiento del grado de transferencia de los aprendizajes adquiridos en cada una de las destrezas a distintos contextos.
- Utilizar los recursos mínimos indispensables para la obtención de la información.

(*Ibidem*)

Las destrezas que se evalúan son cuatro: comprensión de lectura, comprensión auditiva, expresión e interacción escrita y expresión e interacción oral, cada una de ellas calificadas en partes diferentes.

La diferencia entre el examen de dominio que se realiza al final de un curso escolar y la prueba de certificación es que los exámenes de dominio no necesitan ser tan rigurosos como los de certificación, ya que el impacto de los primeros es menor y el conocimiento de la habilidad lingüística que tiene el profesor de sus propios alumnos es mucho mayor, puesto que tiene muestras de lengua generadas durante el curso más que suficientes para determinar su competencia. Sin embargo, en certificación no se conoce a los candidatos, por lo menos a los de matrícula libre, y el impacto del examen es mayor, por lo tanto, la fiabilidad y la validez también deben serlo.

No vamos a extendernos en explicar cómo se estructuran las pruebas de cada una de las destrezas, ya que sobrepasaría los límites de este trabajo; sin embargo, sí nos detendremos más adelante en explicar la estructura de las pruebas de expresión oral en el capítulo cuatro donde se aborda la realización de la investigación, ya que es la capacidad comunicativa en la que nos vamos a centrar en nuestra investigación.

Como se puede ver, uno de los aspectos que señala la Consejería de Educación y que todos los científicos consideran indispensables, como veremos más adelante, es “utilizar técnicas e instrumentos, viables, eficaces y fiables” (*ibidem*), pero al poner en manos de los profesores su elaboración sin tener ninguna formación al respecto, ya se está incumpliendo, al menos, uno de estos principios que es el de la fiabilidad, sin nombrar la viabilidad de la elaboración de las pruebas. ¿Cómo van a cumplir los profesores todas estas características en sus pruebas sin formación y sin tiempo de experimentación para pilotar y determinar su validez? Intentaremos determinar y abordar las dificultades que se plantean en las Escuelas Oficiales de Idiomas no tanto en la elaboración de las pruebas, ya que la experiencia de estos hace que se desenvuelvan con solvencia en estas cuestiones, sino sobre todo en la evaluación de la expresión oral. Para ello, plantearemos posibles soluciones a estas.

- Calificaciones

En relación a cómo se califica en la EOI Santa Cruz de Tenerife, seguimos las directrices de la Orden citada anteriormente de 11 de abril de 2009 (BOC 4 de mayo de 2009), donde se especifica lo siguiente:

Artículo 6.- Obtención de las calificaciones.

6.1. La calificación final de aprovechamiento se obtendrá calculando la media aritmética entre los resultados obtenidos en la evaluación de progreso y la prueba de dominio, con la condición de que en esta última se obtengan, al menos, cuatro puntos. El resultado de esta media se redondeará al valor entero correspondiente.

La calificación final igual o superior a cinco puntos se expresará con el término "Apto". La calificación inferior a cinco se expresará con el término "No Apto".

6.2 Las calificaciones de la evaluación de progreso y de la prueba de dominio se expresarán del uno al diez, sin decimales.

Esta escala numérica, en cada una de las destrezas, se hará corresponder, en sus distintos valores, con los objetivos y contenidos previstos en las programaciones didácticas para cada curso y nivel del idioma, de acuerdo con los criterios establecidos por el centro en su proyecto educativo.

6.3. Las destrezas serán calificadas del uno al diez, sin decimales, en la evaluación de progreso y en la prueba de dominio.

6.4. Tanto en la evaluación de progreso y como en la prueba de dominio la calificación se obtendrá calculando la media aritmética de las calificaciones obtenidas en las cuatro destrezas, redondeada al valor entero correspondiente

6.5. También se podrá superar el curso cuando en la prueba de dominio se obtengan cinco o más puntos, sobre una escala de diez, con independencia de la calificación obtenida en la evaluación de progreso.

A tenor de lo recogido, si un alumno manifiesta un progreso claramente positivo, no se le tienen en cuenta las calificaciones anteriores que pudieran repercutir negativamente en el resultado final de dicha evaluación de progreso. El mismo criterio se aplica si el progreso es negativo: las últimas notas y observaciones son las que prevalecen, siguiendo el criterio de la evaluación continua.

A finales de enero y la primera semana de junio (como mínimo) se informa a los alumnos, a través de una lista con notas por destrezas sobre su evaluación de progreso, de dominio y de aprovechamiento respectivamente.

- Instrumentos de evaluación

Se procura emplear el mayor número de instrumentos de evaluación posibles (esto atañe fundamentalmente a los alumnos a los que se pueda aplicar la evaluación continua), de forma que pueda garantizarse la fiabilidad de la información recogida. A continuación se presenta un listado de los instrumentos de evaluación que pueden emplearse:

-La observación del trabajo diario.

-El cuaderno de clase del alumno.

- Los exámenes y demás pruebas escritas y orales.
- La autoevaluación del alumno, fundamentalmente desarrollada a través de los modelos del Portfolio Europeos de las Lenguas (PEL).
- Las entrevistas.
- El diario de clase.
- Cuestionarios.
- Ficha de corrección de errores.
- Ficha de seguimiento individual.
- Portfolio

- Criterios de calificación y promoción

La evaluación informativa del alumnado oficial se realiza de forma continua con la información que recoge el profesor en la marcha normal de las clases. Además, se realizan al menos dos pruebas globales de seguimiento que evalúan todas las destrezas. Estas dos pruebas se llevan a cabo preferentemente al final de cada uno de los dos cuatrimestres (en torno a finales de enero –primera prueba- y finales de mayo –segunda prueba-).

Los exámenes que se realizan están basados únicamente en las cuatro destrezas. Esto no descarta que a lo largo del curso, si algún profesor desea realizar ejercicios de competencia gramatical o evaluar el vocabulario lo haga libremente, pero la evaluación final debe estar ponderada tal y como se indicó.

4.2.2.4.2. Elaboración y evaluación de las pruebas de certificación de idiomas de la Comunidad Autónoma Canaria.

Los exámenes de certificación se ubican dentro de los exámenes de nivel de dominio que conducen a un certificado oficial (Bordón, 2006). Este tipo de exámenes los deberían realizar expertos que hayan tenido una formación específica en la evaluación y en donde se contara con profesores, pedagogos, psicólogos, evaluadores, etc., “para cubrir todos los aspectos implicados en la realización de un examen de estas características” (Bordón, 2006: 81). Esto además es indispensable para que los criterios de evaluación estén unificados en todas las EOI.

Sin embargo, en las EOI de Canarias, los exámenes de certificación los organizan los propios profesores de las distintas EOI. La Consejería de Educación resuelve “facultar a

los equipos directivos de las EOI para que arbitren el procedimiento más adecuado para que los Departamentos de coordinación didáctica de cada escuela confeccionen los exámenes correspondientes a las pruebas de obtención de los certificados de los idiomas y niveles impartidos en la misma” (BOC N° 167. Miércoles 25 de Agosto de 2010, resolución de 28 de julio de 2010), ocupándose cada escuela de la realización de la prueba de una destreza, lo que produce disparidad de criterios en un examen, ya que no es un equipo de trabajo en común el que las realiza sino los mismos profesores que, además de su carga lectiva, deben realizar y evaluar unas pruebas tan importantes sin ninguna formación al respecto. Además, “los evaluadores, sobre todo los que examinan la actuación oral, tienen que trabajar bajo una considerable presión de tiempo y sólo pueden manejar un número estrictamente limitado de criterios” (MCER 2002: 20). Esto hace que cada vez se haga más necesaria la presencia de un equipo de profesionales elaboradores y evaluadores de pruebas de certificación y de sus criterios de evaluación para que las pruebas puedan ser viables y fiables.

4.2.3. Herramientas utilizadas en nuestro estudio

Para llevar a cabo nuestra investigación se han utilizado diferentes formas de recogida de la información (triangulación de métodos) para alcanzar una mayor veracidad. Nuestro estudio empírico no se ha limitado sólo a la observación o descripción de lo que ocurre sino que se han analizado y se ha actuado en función de los datos recogidos implementando cada paso del ciclo. Por ello, los métodos utilizados han sido tanto cualitativos (entrevistas, observaciones, cuestionarios, autoevaluaciones) como cuantitativos (resultados de las pruebas orales de los candidatos). Los métodos de toma de datos se han seleccionado atendiendo a los objetivos que se pretendían conseguir y al tipo de información que se pretendía obtener de cada uno de ellos. Fundamentalmente hemos empleado los documentos existentes sobre las pruebas de certificación y criterios de evaluación para examinarlos y modificarlos a partir de las actuaciones de los candidatos observadas en clase, de sus necesidades, del contexto de intervención y de sus impresiones recogidas mediante cuestionarios, debates y autoevaluaciones. Además, hemos realizado autoevaluaciones y reflexiones introspectivas de los alumnos sobre su propia producción oral, ya que mediante ellas el evaluador “podrá entender por qué el aprendiz ha alcanzado el resultado correcto o ha cometido un error. Igualmente, el alumno tendrá conciencia de lo que puede y no puede hacer, y, en este último caso, ver claramente en qué parte del proceso se encuentra el fallo” (Bordón, 2004a: 999). A

continuación señalamos los métodos utilizados en función del objetivo que se pretendía conseguir con la utilización de estos:

- a) Para comprobar y analizar la validez y fiabilidad de las pruebas y de los criterios de evaluación utilizados hasta el momento en las pruebas de certificación:
 - Análisis de documentos: escalas de criterios de evaluación y descriptores de los cursos 2007/2008, 2008/09 y 2010/2011; pruebas de expresión oral de los cursos 2007/08 y 2008/2009.
 - Entrevistas con los profesores de la EOI Santa Cruz de Tenerife y con los profesores de español de otras EOI.

- b) Para la realización de unas pruebas y criterios de evaluación más válidos, fiables y objetivos.
 - Observación en clase.
 - Cuaderno del profesor y ficha del alumno: donde se recogieron los datos relativos a debates y reflexiones de los alumnos respecto a su capacidad lingüística, fundamentalmente en lo referido a la expresión oral y apreciaciones de estos sobre los nuevos criterios de evaluación y las nuevas pruebas.
 - Cuestionarios relacionados con las impresiones de los alumnos de los resultados obtenidos.
 - Autoevaluación de los alumnos sobre su propia actuación en cuanto a las producciones orales.
 - Entrevistas con los profesores de español de otras EOI de Canarias.

- c) Para la recogida de las muestras de lengua de los alumnos:
 - Grabaciones de audio de las actuaciones de estos.
 - Pruebas de exámenes de progreso, dominio y certificación.
 - Criterios de evaluación de exámenes de certificación y exámenes de progreso y dominio elaborados estos últimos como resultado de nuestra investigación.
 - Actividades comunicativas de clase.

- d) Para comprobar la validez y fiabilidad de las pruebas de expresión oral y los criterios de evaluación elaborados:
 - Grabaciones de audio de la actuación de los participantes en las pruebas.

- Análisis de las tareas de las pruebas mediante fichas que contenían las características de estas siguiendo a Bachman y Palmer (1996:108)²².
- Pruebas y criterios de evaluación de los exámenes de progreso y dominio y comparación de sus resultados, así como los de dominio de final de curso y certificación mediante el cálculo de coeficientes de correlación de acuerdo al orden de la escala.
- Actividades comunicativas de clase.
- Autoevaluación por parte de los alumnos después de las pruebas de certificación: en este caso se les preguntó qué les había parecido la prueba, si estaban de acuerdo con su nota o si creían que su nota era adecuada a su producción oral real.
- Observación en clase.

4.2.4. Desarrollo del estudio empírico: procesos para la evaluación de la expresión oral atendiendo a las variables.

Teniendo en consideración las anteriores variables, se desarrolló nuestra investigación, una vez detectado el problema, en cuatro ciclos:

4.2.4.1. Ciclo 1.

a) Comprobación de las pruebas de expresión oral de certificación de las EOI de Canarias.

Se examinaron las pruebas de expresión oral de nivel avanzado de la última certificación anterior a la de nuestro estudio²³ y se comprobó que cumplían algunas características que avalaban su efectividad, aunque algunas de ellas no parecían muy claras para los estudiantes, la mayoría sí lo era. Estas pruebas siguen criterios como la autenticidad en la tarea y el rigor en las instrucciones (Bordón, 2006). Autenticidad porque la tarea de examen corresponde a un uso de la lengua real, entendido esto como situación de uso real de la lengua, no en contextos preparados o artificiales sino que se recrean situaciones de la vida real en donde el candidato debe demostrar sus conocimientos y habilidades en el manejo de la lengua hablada que le confirme al evaluador que podrá desenvolverse correctamente en una acción futura. Es verdad que el contexto en el que se produce la actuación es de examen

²² Se puede ver una muestra en el anexo 6.

²³ Para ver un ejemplo de estas pruebas de certificación orales de nivel avanzado correspondientes al curso 2008/2009, ver anexo 1.

pero se pretende “que la tarea del examen responda a un uso de lengua en situación de no examen” (Bordón, 2006:167).

Además también se observó rigor en las instrucciones ya que son “claras, concretas y concisas” (*ibídem*: 170). Aparte de explicitarnos el contexto en el que se dará la actuación, explica lo que se pretende que debe expresar el candidato sin dar excesiva información sobre la gramática o el vocabulario que debe usar. Se pretende proporcionar cierta ayuda al candidato, además de que se percate de algunas funciones comunicativas o exponentes que debe utilizar. Así, por ejemplo, si se le pide que debe dar un consejo al compañero sobre algo, el candidato deberá, o por lo menos podrá interpretar que debe usar las formas de dar consejo con sus respectivos tiempos verbales; sin embargo, no se le ha dicho qué verbos debe utilizar, qué vocabulario o qué registro, por ejemplo.

Igualmente pudimos constatar que las pruebas contenían diversos tipos de actividades comunicativas que iban desde las monologadas, las cuales constituían una parte de la prueba, y que consistían en dar una opinión, defender o argumentar una idea determinada, hasta las dialogadas, que suponían la segunda parte de la prueba, y que se basaban fundamentalmente en conversaciones con el compañero donde se debía dar una información, llegar a un acuerdo o actuar según un rol determinado.

En las escasas tareas en las que los alumnos de la certificación anterior a la de esta investigación tuvieron problemas, pudimos constatar que las instrucciones no estaban muy bien elaboradas, que el contexto era muy improbable que se diera en la realidad o que, en el caso de las monologadas, el discurso u opinión que tenía que producir era de un tema poco actual y desconocido para el alumno.

- Comprobación de los criterios de evaluación de las pruebas de expresión oral de certificación de español.

Durante nuestra experiencia y concretamente en este estudio hemos observado que existen algunos problemas en relación con la fiabilidad de la corrección. La expresión oral, debido a su naturaleza de inmediatez y espontaneidad exige para su evaluación tareas de producción abierta, lo cual supone, a su vez, disponer de evaluadores que califiquen las actuaciones de los candidatos, lo que puede llevar consigo el riesgo de subjetividad. Además, en los criterios de evaluación, algunos

rasgos como la fluidez, la adecuación o el tono que deberían tener igual importancia, no siempre se especifican de la misma forma.

Se pudo comprobar mediante entrevistas con otros profesores no sólo de español sino también de otros idiomas de la EOI Santa Cruz de Tenerife, ya que los criterios de evaluación de certificación son los mismos para todos los idiomas, que estos discrepaban en la calificación de un mismo candidato ya que o bien no seguían las indicaciones del descriptor y lo modificaban según sus necesidades o bien cada uno hacía diferentes interpretaciones del descriptor. Esto nos llevó a suponer que los criterios de evaluación y la formulación de los descriptores no estaban bien definidos, ya que los resultados basados en estos criterios variaba de un evaluador a otro.

También cabe destacar que los profesores-evaluadores de EOI no han sido formados en este tipo de evaluación, sólo disponen de unas directrices dadas por la Consejería de Educación y publicadas en el Boletín Oficial de Canarias, que en algunos apartados, es muy poco clara al respecto. Bordón (2006: 257-258) ya advertía que el evaluador de exámenes de alto impacto, que tiene que evaluar actuaciones de personas que no conoce (alumnos libres de EOI) “no podrá llevar a cabo con rigor y honradez su labor, si no ha sido entrenado para hacerlo”. Sin embargo, y aunque lo idóneo sería disponer de profesionales formados y que hubiese la aplicación de un código deontológico para el ejercicio de la profesión de evaluador (Stansfield, 1993), si se pudiera “disponer de buenos criterios de evaluación que proporcionen descripciones claras y operativas de niveles de actuación y que asignen puntuaciones, que podrían ser computados a los distintos niveles especificados”, y si además pudiéramos conseguir que “quienes vayan a evaluar hayan experimentado los criterios previamente (Bordón, 2006: 172), creemos que la subjetividad podría disminuirse y, por lo tanto, se puede conseguir una mayor objetivación de los criterios de evaluación para obtener la fiabilidad de la prueba, así como de su corrección y calificación más acorde con la producción oral del alumno. Esto es lo que pretendemos demostrar en este trabajo.

De esta forma, se examinaron con detenimiento los criterios de evaluación de las pruebas de expresión oral de certificación de los cursos 2007/08, 2008/2009 y

2010/2011²⁴. Cabe destacar que las pruebas de expresión oral del curso 2007/08 pertenecen a las antiguas enseñanzas de EOI, por lo que los criterios de evaluación de ese curso pertenecen al Ciclo superior de español que en las enseñanzas actuales correspondería con un nivel un poco superior al nivel avanzado actual (podríamos decir un B2+/C1). No obstante, se quiso también comprobar los criterios y descriptores de evaluación de ese curso porque eran diferentes a los de los cursos posteriores. Por otra parte, hay que mencionar que los criterios de evaluación están en relación con las pruebas de examen y con el constructo que se pretende evaluar que en el caso de estos tres cursos escolares, siempre ha sido el mismo: el de evaluar la competencia comunicativa del candidato observando cómo ha asimilado los conocimientos para aplicarlos en la práctica (habilidad). Sin embargo, existen algunas diferencias entre los exámenes de ciclo superior y los de nivel avanzado actuales en cuanto a qué evaluar, a las subcompetencias a evaluar o a las puntuaciones. Las diferencias y semejanzas observadas en el análisis de las tablas utilizadas para llevar a cabo los criterios de evaluación y para puntuarlos de estos tres cursos escolares se pueden observar en el siguiente cuadro:

²⁴ Para ver los criterios de evaluación en su totalidad y las escalas de evaluación de estos tres cursos escolares, consultar anexo 2 y 3.

Cuadro 16: Comparativa de los criterios de evaluación de las pruebas orales de certificación de las EOI de Canarias.

CURSO	2007/2008	2008/2009	2010/2011
Nivel	Ciclo Superior (B2-C1)	Avanzado (B2)	Avanzado (B2)
Constructo ¿Qué evalúa?	Ver anexo 2a	Se expone en general. Se remite al MCER	Se expone en general para el nivel. Hace alusión al currículo de nivel avanzado aunque todavía no existe en Canarias. Se remite al MCER
¿Se especifican los criterios de evaluación?	Sí, para cada destreza. Ver anexo 2b	Sí, pero son iguales para todas las destrezas. Lo especifica por competencias	
Competencias que incluye	Pragmática, lingüística y sociolingüística		
Enunciación de los descriptores	Se usan palabras como “ampliamente”, “adecuada” o “suficiente” que llevan a confusión, ya que no todos los profesores creen lo mismo sobre lo que es adecuado, amplio o suficiente.		
Partes de la prueba	Presentación, monólogo e interacción.	Monólogo e interacción	
¿Hay una puntuación por tarea?	No, la puntuación es total	Sí Monólogo: 4 puntos Interacción: 6 puntos	Sí Monólogo: 40 puntos Interacción: 60 puntos
Duración de las tareas	Presentación: 1-3 minutos Monólogo: 3-5 minutos Interacción: 6 – 10 min. pareja 8 – 12 min. tríos	No especifica por partes, en total 25 minutos.	Monólogo: 5 minutos Interacción: 6 – 7 min. pareja 8 – 9 min. tríos
¿Tiempo de preparación?	Monólogo: 20 minutos Interacción: 2-3 minutos	No especifica	Monólogo: 5 minutos Interacción: 3 minutos
Puntuación parcial por destreza	Sí		
Puntuación total	25 puntos	10 puntos	100 puntos

Hay que destacar que en relación con las competencias y subcompetencias que se evalúan en la expresión oral, en las pruebas del curso 2007/2008 se incluye la subcompetencia “formato”, entendida como la tipología de texto utilizado en el discurso, dentro de la competencia pragmática y, sin embargo, en los cursos siguientes se elimina. Además, en el curso 2008/2009 no se incluye la competencia sociolingüística en el monólogo.

Por otra parte, relacionado con la puntuación, en el curso 2008/2009 se observa que se le da menos importancia en la interacción a la competencia sociolingüística aunque las subcompetencias de la competencia lingüística están menos puntuadas (de 0,8 a 0 en la interacción) que las de las otras competencias (de 1,2 a 0 en la interacción)²⁵. Esto se ha hecho para que la puntuación total de la competencia lingüística sea la misma que la de la competencia pragmática (2,4 puntos). Así se le da más puntuación a la consigna o el registro, por ejemplo, subcompetencias que los candidatos superan normalmente sin dificultad, que al léxico, pronunciación y fluidez, y gramática, de mayor complejidad. En el monólogo, en cambio, se le da la misma puntuación a cada subcompetencia (de 0,8 a 0) para obtener en total 4 puntos en esta tarea.

En los criterios de las pruebas de expresión oral de certificación del pasado curso 2010/2011 se mantienen en general las mismas competencias y subcompetencias que las del año 2008/2009, aunque antes la pronunciación y la fluidez se puntuaba a la vez y ahora se evalúa cada una por separado con sus propios descriptores. Además, en las dos tareas se evalúan las tres competencias señaladas (pragmática, lingüística y sociolingüística) y se le da más importancia a la interacción que al monólogo, igual que las del año 2008/2009. Además, la puntuación mayor se le da al léxico y a la gramática seguidas (de mayor a menor puntuación) de la consigna, organización y cohesión, pronunciación, fluidez y registro. Lo mismo ocurre en el monólogo. Se indica, también, que la puntuación total se divide por 10 y se redondea al valor entero correspondiente.

- Conclusiones de las comprobaciones

Se pudo constatar que los criterios de evaluación no correspondían con el constructo a evaluar, y que, incluso, muchas de ellas no lo tenían especificado, demostrando la poca

²⁵ Lo mismo ocurre con las puntuaciones del monólogo aunque son diferentes porque se calculan sobre 4 puntos.

fiabilidad de la prueba. En general todas las pruebas y criterios de evaluación siguen las directrices del MCER, cuyo planteamiento de la lengua se basa en el enfoque comunicativo de la lengua, y, por lo tanto, en la integración de las competencias pragmática, lingüística y sociolingüística en cada competencia comunicativa, en este caso: la expresión oral. Sin embargo, al proporcionar en la puntuación mayor importancia a la gramática y al léxico, por ejemplo, como ocurre en las pruebas de los últimos años, se sigue teniendo una concepción formal de la lengua donde lo importante es lo lingüístico y no tanto lo pragmático o sociolingüístico como apuntaba el estructuralismo. De este modo, aunque se tiene en cuenta en la puntuación las competencias pragmática y sociolingüística, estas se valoran en menor medida, sobre todo la sociolingüística, haciéndose palpable una preferencia hacia lo lingüístico. Así cuando un alumno aprueba aunque no utilice correctamente del todo el subjuntivo, aún se escuchan comentarios entre los profesores-evaluadores asombrándose de cómo pudo aprobar sin dominar perfectamente este modo verbal, desdeñando los otros aspectos. De ahí que muchas correcciones observadas de años anteriores sean listas de errores gramaticales sin ninguna otra aclaración.

Como se puede apreciar, no se tiene en cuenta la competencia estratégica en ninguna de las certificaciones que, aunque como advertía Canale y Swain (1980) no se diferencia de manera independiente, sí está presente en todas las competencias formando parte de la habilidad lingüística comunicativa. Pero entonces ¿cómo medirla en las actuaciones de los candidatos?, ¿cómo puntuarla?, ¿no debería formar parte de las escalas y criterios de evaluación de la expresión oral?. El MCER (2002: 67) nos sirve de guía recogiendo dentro de la competencia estratégica los siguientes aspectos: planificación, compensación, control y corrección que se traducen en los descriptores generales para el nivel Avanzado 2 (B2) que mostramos a continuación:

Cuadro 17.

Planificación	Planifica lo que hay que decir y los medios para decirlo considerando el efecto que puede producir en el oyente.
Compensación	Utiliza circunloquios y paráfrasis para suplir carencias de vocabulario y de estructura.
Control y corrección	Corrige deslices y errores si se da cuenta de ellos, o si dan lugar a malentendidos. Toma nota de los errores más frecuentes y controla conscientemente el discurso en esas ocasiones.

En la expresión oral la competencia estratégica se puede ver en las habilidades que tiene el alumno, por ejemplo en interacción, para tomar la palabra (turnos de habla), cooperar

o pedir aclaraciones con el fin de mantener la fluidez en la comunicación para que ésta no se rompa. Así tenemos los descriptores del MCER para el nivel B2 en actividades de interacción (2002: 84-85) que nos lo indica:

Cuadro 18.

Turnos de habla	Interviene adecuadamente en la discusión, utilizando el repertorio lingüístico apropiado para hacerlo. Inicia, mantiene y termina el discurso adecuadamente, haciendo un uso eficaz de los turnos de palabra. Inicia el discurso, toma la palabra adecuadamente, y termina la conversación cuando es necesario, aunque puede que no lo haga siempre con elegancia. Utiliza frases típicas (por ejemplo, «Es una pregunta de difícil respuesta») para ganar tiempo y mantener su turno mientras formula lo que va a decir.
Cooperar	Puede reaccionar adecuadamente y seguir las aportaciones e inferencias, de manera que contribuya al desarrollo de la discusión. Contribuye a la marcha de la discusión sobre temas cotidianos, confirmando la comprensión, invitando a otras personas a que intervengan, etc.
Pedir aclaraciones	Es capaz de plantear preguntas para comprobar que ha comprendido lo que el interlocutor ha querido decir y consigue aclaración de los aspectos ambiguos.

Por otra parte, en relación con la puntuación total, en las especificaciones de todas las pruebas de certificación se permite el redondeo no admitiéndose los decimales. Se considera que a partir de 0,5 se redondea al valor superior, y si es inferior a 0,5 el redondeo es a la baja. Esto hace que, por ejemplo, un alumno que esté suspendido con un 4,5 porque el mínimo es 5, realmente está aprobado ya que se redondea a la alza dando lugar a un 5, comparándose la competencia lingüística de este candidato de manera igualitaria con el que obtiene un 5, cuando hay una diferencia de medio punto.

Por último, las distintas partes de la prueba (monólogo e interacción) tienen puntuaciones diferentes. A excepción del curso 2007/08 que no tiene escalas diferenciadas para cada parte de la prueba, en el resto se le da más importancia a la parte de interacción. Creemos que es por la consideración de que la conversación se considera la forma más auténtica de la expresión oral (Vázquez, 2000, Bordón 2006, MCER, 2002). Sin embargo, las tareas monologadas también se producen a menudo en la vida real cuando por ejemplo, se tiene que defender una postura u opinión ante varias personas y, por otra parte, también son interactivas (emisor-audiencia), por lo que creemos que no se le debería dar una puntuación diferenciada. A esto hemos de agregar que tampoco podemos escudarnos en que la interacción es una actividad más compleja, porque hablar sobre un tema con escaso tiempo de preparación, como ocurre en las

tareas monologadas de nivel avanzado, no es una labor nada fácil; incluso es difícil para los propios nativos.

Por ello, se hace necesaria la mejora o reelaboración de los criterios de evaluación, descriptores y puntuaciones consultadas para luego experimentarlos con el grupo-clase y ver hasta qué punto se aumenta la fiabilidad en la corrección y la mejor comprensión de los descriptores por parte de los profesores.

- Actividad de introspección con el alumnado y análisis de sus necesidades:

Después de comprobar las pruebas y los criterios de evaluación, se pretendió recoger datos del alumnado mediante un debate en clase y actividades de comprensión y expresión oral para ver qué impresión tenían de su expresión oral en español, qué problemas observaban en el desarrollo de esta capacidad lingüística, así como el análisis de sus necesidades. La mayoría aseguró no tener demasiados problemas con la expresión oral, ya tenían cierto nivel, se entendían con hablantes nativos e incluso, muchos de ellos, vivían hacía tiempo en las Islas Canarias, por lo que siempre estaban practicando el idioma español. Sin embargo, como ya hemos señalado con anterioridad las dificultades planteadas se basaban más en criterios personales o afectivos (miedo a equivocarse, nerviosismo, frustración) que lingüísticos, pragmáticos o culturales. Algunos también advirtieron lo absurdo de muchas situaciones planteadas en los exámenes y lo artificial y forzado que resultaba interactuar de manera natural en una situación planteada de repente; otros consideraban que algunas tareas monologadas versaban no tanto sobre temas de actualidad sino sobre temas específicos o muy técnicos de los que apenas tenían información.

Para el análisis de necesidades se realizó una actividad con vídeo en el que distintos estudiantes como ellos decían por qué querían estudiar español, y posteriormente tenían que expresar su opinión.

Los datos obtenidos nos sirvieron para plantear actividades orales comunicativas para luego experimentarlas en clase y validarlas con el fin de elaborarlas y aplicarlas en las pruebas de certificación de junio²⁶.

²⁶ Hay que recordar que durante este curso escolar la Consejería de Educación de Canarias encomendó la realización de las pruebas de certificación de español a las EOI de Canarias, por lo que nos repartimos la elaboración de estas por destrezas. A la EOI Santa Cruz de Tenerife le correspondieron las pruebas de expresión oral. No se pudieron emplear los criterios elaborados ya que estos ya estaban impuestos por dicho organismo.

4.2.4.2. Ciclo 2

- Elaboración de actividades comunicativas y pruebas orales.

A la vista de las consideraciones de los alumnos sobre las actividades de expresión oral y según sus intereses se comenzaron a elaborar actividades de expresión oral para practicar en el aula, así como las pruebas de expresión oral del examen de dominio de enero. Los resultados y observaciones surgidos en el aula nos servirían posteriormente para la realización de las pruebas de expresión oral de certificación que teníamos que elaborar y que se fueron perfilando desde este momento.

De esta forma, para la elaboración de las actividades para el aula intentamos seguir la misma estructura observada en los exámenes de certificación de monólogo e interacción con el fin de experimentarlas con los alumnos para los futuros exámenes. Como de todas maneras era una actividad de aula, se tuvieron en cuenta algunas características que se pasan por alto en los exámenes debido a que en estos últimos el tiempo es limitado. Estas actividades también servían a los alumnos como instrucción para la prueba de dominio de enero y para los exámenes de certificación y al final de la actividad se incluían otras actividades para practicar los contenidos gramaticales o la integración de otras destrezas. Se siguieron las siguientes fases en clase según Pinilla (2004: 890):

- Preactividad: para introducir la actividad, motivar al alumnado o conocer los conocimientos previos de estos.
- Exposición del objeto de la actividad y los pasos que deben seguir para llevarla a cabo con éxito.
- Desarrollo de la actividad.
- Evaluación de ésta.
- Postactividad: deducción de reglas gramaticales, integración de destrezas, etc.

Además, se fueron realizando las pruebas de los exámenes de dominio de enero y, a su vez, siguiendo el mismo esquema de dos tareas (monólogo e interacción), las pruebas de expresión oral de certificación. Así, siguiendo las fases que propone Teresa Bordón (2006: 67) para la elaboración de un examen de L2, se fueron elaborando para las pruebas de certificación 20 modelos distintos de examen de expresión oral (diez para tareas monologadas y 10 para interacción); siguiendo el MCER en donde las actividades relativas a la lengua oral que debe realizar el candidato son: expresión oral, en las que el

candidato produce un texto oral (llamadas por nosotros monologadas siguiendo a Calsamiglia y Tusón, 1999) y de interacción oral en la que el candidato debe actuar con uno o más interlocutores para construir una conversación (discurso dialógico). En la fase previa decidimos que el constructo a evaluar de lengua oral era el que combinaba los conocimientos con la habilidad lingüística tal como lo concibe el MCER, como ya se especificó en el epígrafe 2.1.1. En la fase de diseño, se decidió que seguiríamos la misma estructura que los años anteriores de dos partes: una parte para monólogo y otra para interacción, aunque también se observó la necesidad de rescatar la presentación inicial con el candidato como tarea introductoria que sí aparecía en las pruebas del 2007/2008, aunque ésta no se evaluaría, solamente se utilizaría para intentar relajar al candidato y liberarlo de la tensión que produce hacer un examen, lo cual contribuye a la aceptabilidad de la prueba y a su fiabilidad. Por otra parte, para saber qué dificultad tendrían que tener las actividades y qué contenidos son los que deberían incluir se consultó, además del MCER (2002) el PCIC (2007) para constatar con mayor seguridad²⁷ los contenidos del nivel avanzado. Posteriormente, y una vez establecidos los criterios de evaluación relacionados con las pruebas, se pasó a la fase de experimentación administrando las pruebas a los candidatos en el mes de enero y se interpretaron los resultados.

- Modificación de los criterios de evaluación y de su puntuación.

Los niveles de lengua que nos propone el MCER son fundamentales para redactar los criterios de evaluación. Sin embargo, hemos observado que existe cierta confusión con los niveles que propone el Marco y los que se imparten en las EOI. Así, al pasar del esquema tradicional de las enseñanzas de idiomas (niveles de primero a quinto) al esquema de los niveles del MCER²⁸ se ha producido cierto desconcierto entre el profesorado, las editoriales y las propias instituciones docentes. Para las EOI el nivel B2 es un nivel Avanzado pero se tiene en mente el antiguo 5º que se impartía como un nivel superior (C1) cuando realmente, según el MCER el nivel avanzado es un intermedio-alto. Esto hace que algunos profesores al evaluar estén calificando un

²⁷ En Canarias todavía no ha sido publicado el currículo de nivel Avanzado por lo que los contenidos para este nivel que se recogen en la programación del Departamento de español se han determinado siguiendo el MCER (2002) y el PCIC (2007).

²⁸ A1 plataforma-A2 acceso (nivel principiante), B1 umbral-B2 avanzado (nivel intermedio), C1 dominio-C2 maestría (nivel avanzado).

examen que no corresponde al nivel real a evaluar sino al que ellos tienen en mente según las antiguas enseñanzas. Así, confunden la nomenclatura del MCER (Básico, Intermedio o Umbral, Avanzado, Superior) con las de los cursos de la EOI (Básico, Intermedio y Avanzado) y así, si el MCER dice que el B2 es un intermedio alto, algunos profesores lo identifican con un B1.2 que es el intermedio alto en la EOI y el B2 con un C1 que era el antiguo superior. Esto lleva a evaluaciones erróneas y de ahí que se haya generado a partir de la Asociación de Profesores de EOI un movimiento de protesta para que se reconozca el nivel avanzado de EOI como un C1 y no como un B2. Hemos de añadir que se considera que el MCER señala seis niveles y seis son los cursos de un idioma en la EOI, por lo que se podría corresponder cada curso con un nivel, a que se subdivida un mismo nivel en dos cursos. Muchos profesores de EOI solicitan este cambio y creen que, en tal caso, se podrían hacer subdivisiones teniendo en cuenta las características del alumnado como, por ejemplo, si poseen una lengua materna muy alejada de la lengua de estudio, como ya ocurre con los españoles que estudian árabe y chino que para conseguir un nivel básico o intermedio deben pasar tres cursos. Lo que queremos decir es que algunos profesores-evaluadores tienden a enseñar y examinar en sus aulas por encima del nivel que corresponde, sobre todo en el nivel avanzado por esa confusión existente, y otros siguen los contenidos que establece el MCER y el PCIC. Esto afecta gravemente a la evaluación y de ahí que se haga necesario en los exámenes unos criterios de evaluación claros en este sentido, así como que los profesores tengan claro el constructo y los contenidos y habilidades que lo integran.

Por otra parte, esta confusión a la hora de evaluar se mantiene porque los criterios de evaluación de la expresión oral en las pruebas de certificación son los mismos para todos los niveles e idiomas sin tener en cuenta las tareas con las que se van a utilizar dichos criterios; son criterios muy generales y nada específicos, lo que contribuye aún más a la confusión y subjetividad. Un alumno de nivel básico no puede hablar igual que uno de nivel avanzado o intermedio y criterios como “cumple con las pausas adecuadas para el nivel” o “tiene la fluidez adecuada o suficiente” llevan a error, sobre todo si no se tienen claros los contenidos o si no se tiene experiencia en este tipo de enseñanzas y en la evaluación de estas pruebas de dominio del idioma, aún más si son de alto impacto.

Por todo ello, modificamos en principio la puntuación de las escalas haciéndolas más homogéneas, dándoles igual importancia a todas siguiendo a Alderson *et al.* (1998: 146), ya que la ponderación diferenciada raramente mejora la fiabilidad o la validez, y

teniendo en cuenta el constructo donde la competencia comunicativa del alumno está integrada por distintas subcompetencias como apuntaban Hymes (1971) o Bachman y Palmer (1996) y como recoge el MCER (2002).

Posteriormente, según las observaciones recogidas sobre las actividades de expresión oral desarrolladas en clase y con la puntuación cambiada ponderando todas las subcompetencias que las integran por igual, se suministraron las pruebas elaboradas en clase en la prueba de dominio de enero y se utilizaron para evaluarlas los criterios de evaluación modificados en su puntuación²⁹. Se quiso sólo modificar la puntuación para comprobar si el desfase experimentado en años anteriores y observado a principio de clase entre la puntuación resultante de la actuación del candidato y la capacidad lingüística observada por el profesor en clase era resultado de la puntuación ponderada de los descriptores o no. Es decir, en ocasiones los estudiantes obtenían como calificación de final de curso para la expresión oral, por ejemplo, un nueve (sobresaliente), y cuando se presentaban a la prueba de certificación su actuación era calificada con un cinco o al contrario (menor nota en examen final de curso y mayor nota en certificación), por lo que existía un problema en la evaluación. Se podría pensar que el profesor que lo evaluara en clase fuera menos exigente que el de certificación o viceversa, pero si eso era así también podría ser que existiera demasiada subjetividad en la nota de los examinadores; por lo tanto, en ambos casos no existiría fiabilidad. Por todo ello, merecía la pena investigarlo.

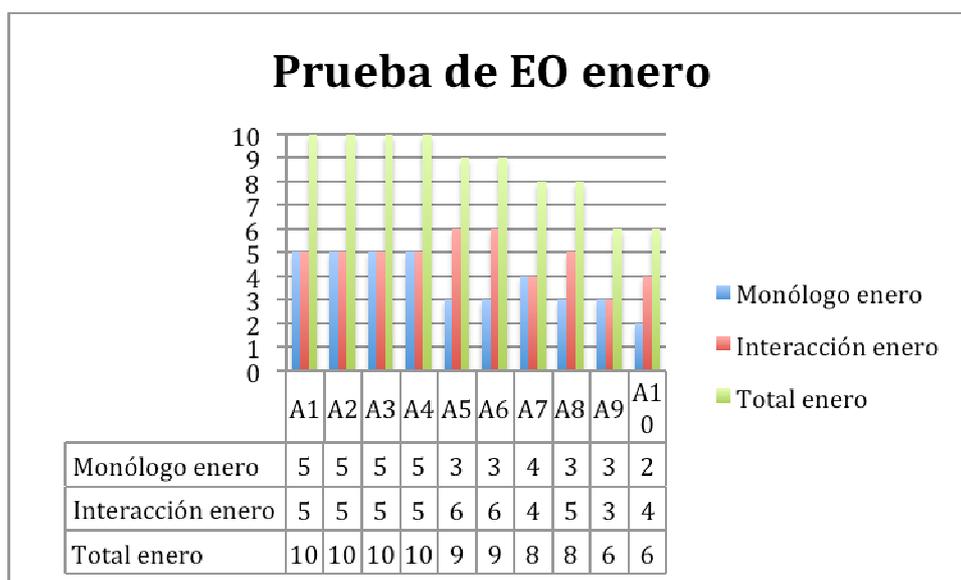
- Resultados de las pruebas de dominio de expresión oral de enero

Con las nuevas pruebas elaboradas y con los criterios de evaluación modificados en su puntuación se administraron las pruebas de expresión oral de enero y los resultados fueron los siguientes: de los 10 alumnos todos mostraron una competencia lingüística correcta para el nivel, pero mientras que el resultado de muchos era equiparable con su capacidad observada en clase, otros obtuvieron una puntuación por encima de su capacidad lingüística a juicio de la profesora que les impartía clase, y al igual que habíamos detectado sin la puntuación cambiada. Incluso se buscó la opinión de una

²⁹ Se modificó la puntuación en las subcompetencias haciendo que cada una tuviera la misma puntuación sin discriminación entre ellas, siendo fieles al constructo a evaluar que da igual importancia a todas las subcompetencias integradas en cada una de sus competencias (pragmática, lingüística o sociolingüística) y que conforman la competencia comunicativa del candidato.

segunda profesora, concretamente la que les dio clase el curso pasado que conocía bastante bien la producción de estos alumnos y coincidió con las impresiones de la primera profesora. Por ahora estas conclusiones y apreciaciones se basaron en la experiencia y en las observaciones en clase con estos alumnos, así como en las comparaciones entre las actuaciones en clase y esta prueba de dominio. A continuación se muestran las puntuaciones obtenidas divididas por tareas para luego poder compararlas con las de las pruebas de progreso de marzo y las de dominio finales de junio:

Cuadro 19.



Como se puede ver, la puntuación muestra una actuación de los alumnos muy buena: cuatro alumnos obtienen un 10, dos alumnos un 9, otros dos un 8 y dos más un 6. Cabe destacar que de los alumnos que obtienen un 10 sólo dos de ellos se merecían tal puntuación y los otros la obtuvieron porque los evaluadores tuvieron que ser fieles a los descriptores sin modificarlos. De esta forma, por ejemplo, uno de estos alumnos de “falso 10” tuvo una organización muy buena en su discurso pero no usó apenas conectores, por lo que la puntuación de la organización y cohesión no debería ser ni muy alta ni muy baja sino un término medio; sin embargo, el descriptor de puntuación más alto establece: “produce un monólogo bien organizado y estructurado, y utiliza con corrección una gran variedad de conectores y marcadores del discurso”. La primera parte sí la cumple el candidato pero la segunda no, por lo que este descriptor no lo podemos utilizar para evaluar al candidato: tenemos que encontrar otro que se ajuste al resultado de la prueba. Sin embargo, para encontrar un descriptor que incluya una

ausencia de conectores nos encontramos con lo siguiente: “produce un monólogo desorganizado y mal estructurado, y hace un uso muy limitado de los conectores y marcadores del discurso” pero esto tampoco obedece a la actuación del candidato. Al final el profesor-evaluador no sabe cómo evaluar este aspecto, guiándose, al final, por su propia impresión e incluyendo un alto grado de subjetividad en la evaluación.

- Conclusiones del ciclo 2

Estos resultados hacen ver que aunque las notas de los estudiantes sean positivas, no reflejan en algunos de ellos su verdadera capacidad, no tanto porque la puntuación de la escala fuera errónea debido a que la ponderación no se ajustaba al constructo, sino porque los descriptores no cubren todos los posibles casos que puedan darse en las muestras de lengua de los alumnos, como pudimos ver en el caso de la medición de la coherencia y de la cohesión. Este aspecto de la competencia pragmática aparece puntuado de manera conjunta en la escala, pero hemos observado que para que se pueda medir de la forma más fiable posible es necesario separarlo y desarrollar descriptores para cada parte, esto son descriptores de organización y descriptores de cohesión.

Además, en la competencia sociolingüística sólo se incluye el registro haciendo referencia a los marcadores de uso social. Consideramos que dicha competencia se debería ampliar con descriptores referidos a las expresiones idiomáticas o de la sabiduría popular propias de este nivel, como se recoge en el PCIC.

Por otra parte, palabras como “ampliamente”, “adecuado” o “suficiente”, descriptores como “el diálogo se realiza con fluidez y soltura, y las pausas apropiadas” o “el diálogo se realiza con la fluidez y pausas apropiadas” llevan a confusión y, cuanto menos, son difíciles de medir, ya que ¿quién determina lo que es fluidez y soltura y lo que sólo es fluidez? En este caso los profesores-evaluadores pueden tener distintas opiniones, por lo que este tipo de descriptores y palabras que invitan a la subjetividad deben evitarse, a no ser que se deje claro o que tengan claro los evaluadores lo que es adecuado, suficiente o amplio en el nivel que se evalúa. Ya el MCER (2002: 198) indicaba la necesaria precisión en la redacción de los descriptores de evaluación y Bordón (2006: 71) nos advertía que no sólo deben estar los criterios de evaluación bien descritos, sino que además siempre tiene que haber una experimentación para garantizar la objetividad de la calificación ya que no todos los evaluadores

van a entender de igual manera apreciaciones como “vocabulario *adecuado* a la situación”, o “errores *frecuentes* que *interrumpen* la comunicación”. El alcance de

adecuado, frecuente o interrumpe, puede variar de un evaluador a otro, en función de diversos factores –subjetivos– como, por ejemplo, su familiaridad o desconocimiento de la interlengua del candidato, su propio nivel de exigencia en cuanto al vocabulario, o, incluso, su estado de ánimo.

Además, también hemos observado que los alumnos utilizan muchas estrategias para llevar a cabo la comunicación que son indicadoras de su mayor o menor habilidad lingüística comunicativa. Al no estar recogido este aspecto de manera específica en las escalas utilizadas es conveniente introducir, además de las tres competencias (pragmática, lingüística y sociolingüística) la competencia estratégica, también señalada por el MCER dentro del uso de la lengua. De esta forma, se tendrían en cuenta aspectos como las habilidades que tienen los estudiantes para tomar la palabra (turnos de palabra), si respeta el de los otros o no, las negociaciones que hace con sus interlocutores (cooperación), o las estrategias que usa para solicitar aclaraciones o hacer circunloquios en el caso que no recuerde alguna palabra específica, por ejemplo. Creemos que este aspecto también debe tenerse en cuenta en las escalas de evaluación de la expresión oral.

También, en relación con la pronunciación, los profesores-evaluadores tienen dificultades para decidir si considerar negativo o no la influencia de la lengua materna del candidato en la pronunciación, esto es debido a que algunos alumnos tienen una pronunciación excelente sin “acento extranjero” y otros, en cambio, se expresan bastante bien aunque con algunos sonidos de su lengua materna. Por ello, se considera que en estos descriptores se debe dejar claro lo que debe hacer el alumno relacionado con el nivel y siguiendo el PCIC.

Por último, en relación con las pruebas elaboradas, los alumnos observaron, mediante un cuestionario que se les administró para recoger tal información, que en algunos temas no sabían lo que decir, bien porque no tenían conocimientos sobre ellos o bien porque requerían creatividad por parte de los candidatos. En relación con la primera dificultad, los temas que se tratan en las pruebas son aquellos que ya se han abordado previamente en clase y sobre los que se supone que el alumno debe tener ciertos conocimientos, aunque también es verdad que en los exámenes de certificación hay alumnos libres que no asisten a clase y que no han “estudiado” estos temas. Por ello, en los temas de las pruebas, fundamentalmente en las monologadas, se deben seguir, a falta de un currículo de nivel avanzado de español en Canarias, los actuales documentos que nos pueden servir de guía como los del MCER y más específicamente para español los

del Instituto Cervantes (PCIC). Estos temas son los que se corresponden con el nivel; los alumnos libres también deberían saberlos aunque no hayan recibido clases en la EOI. El segundo problema de las actividades que requieren creatividad se podría subsanar dándoles a los candidatos unas directrices a seguir, unos puntos a tratar para que puedan guiarse en su producción, directrices con información que ellos luego deben desarrollar.

Por otra parte, hemos observado en relación con las tareas de examen que algunos aspectos de la validez se cumplen y otros no. Así, la validez interna se puede considerar que sí se cumple, puesto que los contenidos de las tareas están relacionados con lo que se pretende evaluar y el efecto de estas en cuanto a las respuestas de los candidatos es positivo. Sin embargo, la validez externa no se da debido a que la actuación de los alumnos en clase no se corresponde con el resultado de la prueba por la ambigüedad de los descriptores, lo que conlleva falta de fiabilidad. Creemos, así, que modificando los descriptores y los criterios de evaluación, además de la puntuación, se podrían conseguir unos resultados de las muestras de lengua de los candidatos generadas en la prueba más adecuados y ajustados a su capacidad lingüística comunicativa real.

En suma, después de esta fase de experimentación y de interpretación de los resultados de las pruebas de dominio orales de enero, se hace necesario modificar los criterios de evaluación, incluyendo junto a las competencias pragmática, lingüística y sociolingüística, la estratégica; ampliar los prescriptores de la competencia sociolingüística no sólo limitándonos al registro sino también a las expresiones idiomáticas y a las relaciones sociales, separar la organización y la cohesión en descriptores diferentes, precisar los descriptores referentes a la pronunciación y, en el caso de la temática de las tareas, ajustarlas a lo especificado en el MCER y el PCIC.

4.2.4.3. Ciclo 3

- Implementación del ciclo 2 en relación a las pruebas y a los criterios de evaluación.

A partir de las conclusiones observadas en el ciclo 2, se procede a modificar las pruebas y los criterios de evaluación que se utilizarán nuevamente como experimentación en clase y que luego servirán como modelos para ir elaborando y modificando las pruebas de expresión oral de certificación³⁰. Así, como ya hemos visto, en las tareas

³⁰ Hay que tener en cuenta que los criterios de evaluación están modificados de acuerdo con el tipo de alumnado con el que se experimentan (alumnos adultos de B2 de distinta nacionalidad

monologadas se limitan los temas sobre los que los candidatos tienen que producir muestras de lengua a los que se señalan como propios del nivel B2 en el PCIC (2007: 446) y se intenta que se relacionen con temas de actualidad. Además, al ser también la producción de un monólogo una tarea difícil de desarrollar pues, en general, los candidatos deben producir un discurso oral en donde se argumente su opinión, se le facilitan una serie de puntos o ideas que pueden utilizar en su producción, además de ofrecerles pistas para que utilicen determinados contenidos pero sin ser demasiado explícitos en dichos puntos, de este modo se evita que los candidatos se “queden en blanco” ante actividades que exigen cierta creatividad. Esto también se hace en las tareas de interacción que constituyen la segunda parte de la prueba.

En cuanto a los criterios de evaluación y sus escalas, se implementan también según lo especificado en las conclusiones del ciclo 2: se le da la misma puntuación a las distintas subcompetencias, inclusión de la competencia estratégica³¹, ampliación de las subcompetencias de la competencia sociolingüística³², separación en la evaluación de la organización y la cohesión del discurso y redacción de descriptores más claros en relación con las habilidades y contenidos del nivel B2³³.

- Realización de pruebas de expresión oral de progreso durante el mes de marzo y resultados.

Siguiendo la misma estructura planteada para las pruebas de certificación y utilizando un modelo de pruebas similares a estas ya elaboradas según lo especificado en los anteriores ciclos, se procedió a realizar una prueba de progreso durante el mes de marzo para demostrar la validez y fiabilidad de estas. Así, los resultados de esta prueba nos servirían para compararlos con los de enero y posteriormente con los de junio.

en inmersión lingüística) y según lo observado en las actuaciones de estos en clase siguiendo lo que Luoma (2009: 84) llama “muestras de desempeño”, en donde a partir de la escala existente se escogieron las muestras de los alumnos del nivel que se consideraban más correctas, se compararon con la redacción de las escalas, se observaron las discrepancias y acuerdos y se modificaron. Se siguen, también, las directrices del MCER (2002) y de ALTE (1995, updated July 2005).

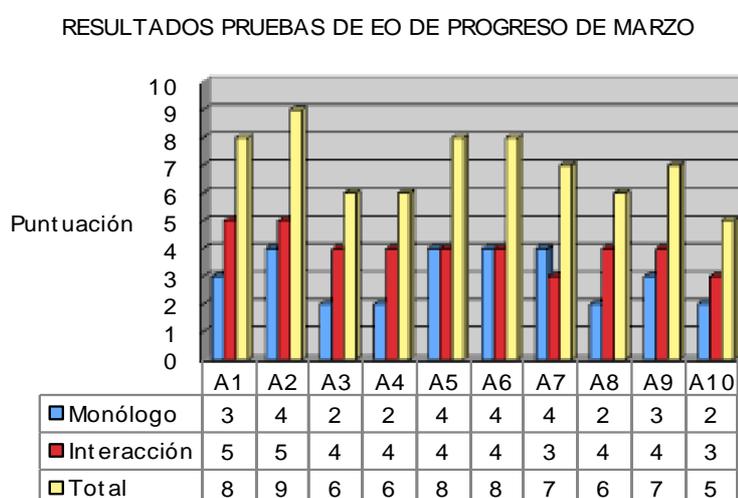
³¹ Se incluyeron en ellas las habilidades más frecuentes observadas en la actuación de los alumnos en clase y se diferenciaron según el tipo de tarea (monólogo o interacción).

³² Se añadieron descriptores relativos a las expresiones idiomáticas debido a que forman parte del currículo de B2 (PCIC) y se utilizan frecuentemente en las conversaciones y opiniones. Además, los elementos fraseológicos y paremiológicos se trabajaron durante todo el curso en clase.

³³ Para ver los nuevos criterios de evaluación modificados y las escalas aplicados a estos, consultar anexo 4.

También cabe destacar que durante toda la investigación y con el fin de garantizar la calidad de las tareas de expresión oral, se utilizó un cuadro de análisis de tareas siguiendo a Bachman y Palmer (1996, 49-50 y 108) que facilitó la comprobación de estas, analizando los elementos que se utilizaban en la tarea teniendo en cuenta cinco características: ambiente (condiciones físicas y tiempo de realización), formato (organización e instrucciones generales), características del input de las tareas (instrucciones, formato, lengua y naturaleza de las instrucciones usadas en las actividades), características de la respuesta esperada (formato, lengua y naturaleza de la respuesta esperada) y las relaciones entre las instrucciones de la prueba y la respuesta dada³⁴. Los resultados de estas pruebas de expresión oral de progreso realizadas en marzo aplicando los criterios de evaluación modificados fueron los siguientes: un alumno obtuvo un nueve, tres alumnos un ocho, dos alumnos un siete, tres alumnos un seis y un alumno un cinco. En el siguiente gráfico se pueden observar mejor los resultados divididos por tareas:

Cuadro 20.



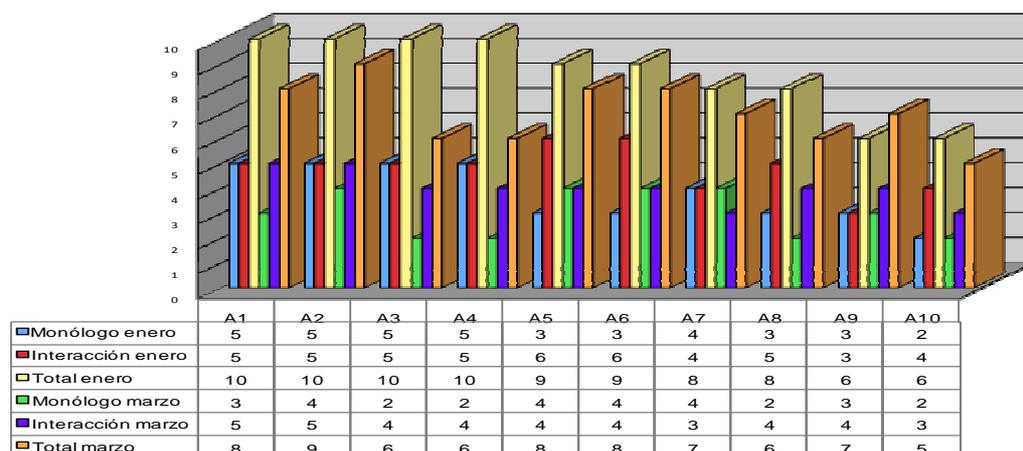
Con los nuevos descriptores los resultados se muestran más fiables en relación con la capacidad de los candidatos observada en clase, además de que hay que tener en cuenta que los alumnos habrán aprendido más durante el tiempo transcurrido y que si se han aplicado en sus estudios o, simplemente, por estudiar el idioma en inmersión lingüística

³⁴ Para ver el cuadro de análisis de tareas utilizado con un ejemplo de dicho análisis aplicado a una prueba, consultar anexo 6. El cuadro ha sido traducido al español.

deberían haber mejorado. Sin embargo, la mayoría de los alumnos respecto a las notas obtenidas en enero han obtenido una calificación inferior; recordemos que los criterios utilizados anteriormente no eran válidos ni fiables y, en cambio, ahora podemos considerar que sí lo son si comparamos los resultados respecto a las pruebas y tareas hechas en clase y evaluadas también con estos nuevos criterios.

- Comparación de los resultados obtenidos en enero y los resultados de marzo

Cuadro 21.



Para poder comprobar la escasez de validez y fiabilidad de los resultados obtenidos en enero y sospechados desde un principio por la diferencia existente entre el resultado evaluado y la capacidad de expresarse oralmente de los alumnos, se establecieron correlaciones por orden de escala entre las pruebas de enero y marzo. De esta forma, se podría constatar la diferencia de puntuación existente entre ellas y, por lo tanto, la falta de fiabilidad de los resultados de enero. Las correlaciones que se hicieron fueron sobre las puntuaciones totales tal como mostramos a continuación³⁵:

³⁵ En el análisis de estas correlaciones los alumnos están ordenados siempre de forma idéntica, de tal manera que A1 siempre corresponderá con ese alumno, A2 con el siguiente y así sucesivamente. Se sigue la fórmula para calcular la correlación de acuerdo al orden de la escala que aparece en Alderson *et al.* (1998: 263-264)

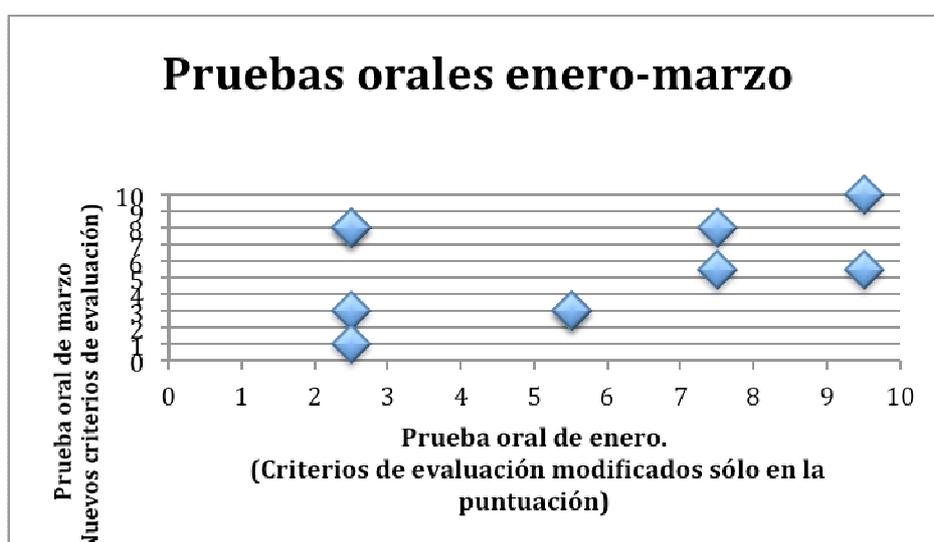
Cuadro 22. **Correlación entre las pruebas de enero y marzo por orden de escala**

Alumno	Resultados enero	Resultados marzo	Escala enero	Escala marzo	d	d ²
A1	10	8	2,5	3	0,5	0,25
A2	10	9	2,5	1	1,5	2,25
A3	10	6	2,5	8	5,5	30,25
A4	10	6	2,5	8	5,5	30,25
A5	9	8	5,5	3	2,5	6,25
A6	9	8	5,5	3	2,5	6,25
A7	8	7	7,5	5,5	2	4
A8	8	6	7,7	8	0,5	0,25
A9	6	7	9,5	5,5	4	16
A10	6	5	9,5	10	0,5	0,25

d= diferencia; d²= diferencia al cuadrado

Con estos datos y observando las diferencias de puntuación, podemos observar que la correlación entre las dos pruebas es de +0,4. Es decir, existe cierta similitud entre las notas ya que se evalúan aspectos del mismo rasgo (competencia de la expresión oral) pero no hay coincidencia total y con algunos estudiantes hay mucha distancia entre la puntuación de enero y la de marzo (A1, A3, A4, A8), lo que nos hace suponer que la puntuación de enero no es fiable y que la de marzo tiene mayor fiabilidad porque coincide con la de la actuación de los candidatos observada en clase por la profesora. Por lo tanto, existe una correlación baja entre las dos pruebas. Esto se puede apreciar mejor en el siguiente gráfico de dispersión:

Cuadro 23.



- Evaluación por parte de los alumnos de las pruebas administradas y de los resultados obtenidos en marzo según los nuevos criterios de evaluación

En las clases siguientes a la administración de las pruebas de marzo se debatió con los alumnos sobre dichas pruebas y resultados. El 95% de los alumnos manifestaron que estaban de acuerdo con estas y que las instrucciones para su desarrollo eran más claras, además de reducir, gracias a los puntos incluidos en ellas, la falta de creatividad y la incertidumbre sobre qué decir ante determinados temas.

En cuanto a los resultados, los que habían obtenido menor puntuación en marzo que en enero se manifestaron descontentos y pusieron en duda los criterios de evaluación aplicados aún cuando se les explicó cómo se habían utilizado y los criterios que se habían incluido. Además, consideraban que había pasado más tiempo durante sus estudios y que, por lo tanto, deberían hablar mejor y que eso se debería haber reflejado en el resultado de la prueba. Sin embargo, después de reflexionar sobre su propia actuación en el examen reproduciéndosela en clase gracias a las grabaciones efectuadas de cada prueba, observando sus dificultades y tras realizar una reflexión sobre su expresión oral actual a partir de un cuestionario, comprendieron que la nota actual estaba más acorde con su competencia lingüística en lugar que la de enero.

Por lo tanto, se estimó que las pruebas y criterios de evaluación actuales desarrollados según lo expuesto, al menos, eran más válidos y fiables que los de enero.

- Resultados de las pruebas de expresión oral de dominio de junio

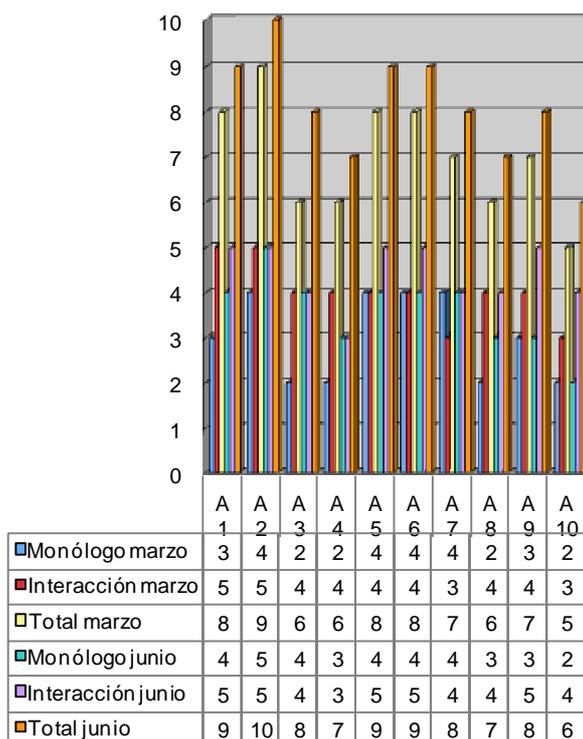
Siguiendo el mismo formato y los mismos criterios de evaluación de marzo, ya que los alumnos se mostraban bastante receptivos ante estos y que se cumplía la validez interna (aparente, constructo, contenido y respuesta), se procedió a administrar las pruebas de fin de curso de expresión oral en junio, con la finalidad, además, de demostrar su validez externa mediante la comparación de estos resultados de junio con los de marzo y teniendo en cuenta variables como el tiempo transcurrido y, por consiguiente, el mayor aprendizaje de los alumnos. Como los resultados de las pruebas de marzo diferían de los de enero, obteniendo los de marzo una puntuación más baja, se podría pensar que la puntuación de marzo quizá estaba influida por la subjetividad del profesor-evaluador o por su exigencia o flexibilidad al puntuar las pruebas³⁶. Así, si se

³⁶ Aunque esto resultaba improbable ya que los descriptores estaban elaborados para evitar esa subjetividad y los resultados obtenidos en marzo eran más acordes con la actuación lingüística de los alumnos observada en clase que los de enero (validez concurrente), se precisa constatar

constataba que los resultados de marzo guardaban correlación con los de junio evaluados con los mismos descriptores es que la fiabilidad y validez de estas se cumplía y que no estaban influidos por estos aspectos.

Los resultados de la prueba de expresión oral de junio se muestran en el siguiente gráfico:

Cuadro 24.



Se muestra así, la validez de la prueba ya que tanto lo referido a su relación con el contenido y respuesta de los alumnos como a los resultados de dicha prueba y las medidas de la capacidad de expresión oral de estos obtenidas fuera de esta prueba (en clase o prueba de marzo) coinciden.

Aún así para demostrar la validez concurrente y la fiabilidad interna tanto de las tareas de las pruebas como de sus criterios de evaluación mostramos a continuación la correlación por orden de escala entre estas dos pruebas (marzo y junio):

una vez más la fiabilidad en la puntuación estableciendo una correlación entre las pruebas de marzo y junio.

Cuadro 25.

Correlación entre las pruebas de marzo y junio por orden de escala

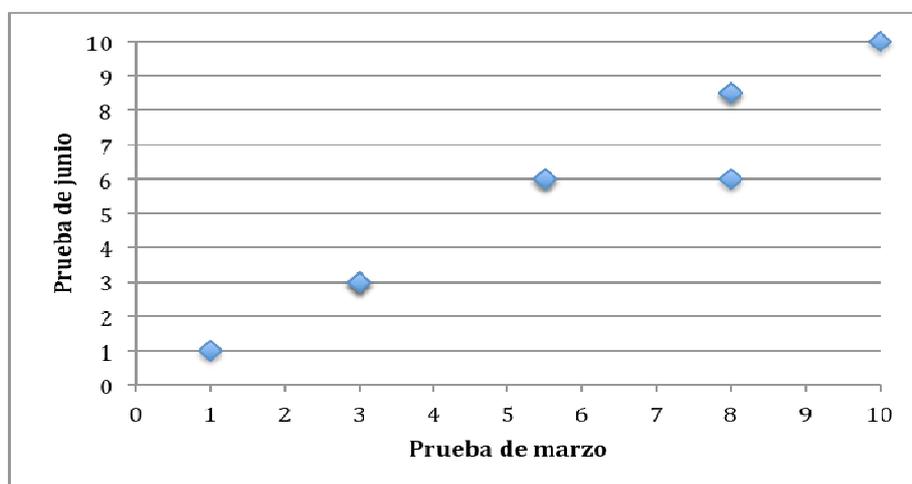
Alumno	Resultados marzo	Resultados junio	Escala marzo	Escala junio	d	d ²
A1	8	9	3	3	0	0
A2	9	10	1	1	0	0
A3	6	8	8	6	2	4
A4	6	7	8	8,5	0,5	0,25
A5	8	9	3	3	0	0
A6	8	9	3	3	0	0
A7	7	8	5,5	6	0,5	0,25
A8	6	7	8	8,5	0,5	0,25
A9	7	8	5,5	6	0,5	0,25
A10	5	6	10	10	0	0

d= diferencia; d²= diferencia al cuadrado

La correlación entre estas dos pruebas es de +0,96. Es bastante alta por lo que la fiabilidad y la validez se cumplen. Esto quiere decir que con los nuevos criterios de evaluación y escalas modificados no existe tanta diferencia entre las notas y, por lo tanto, se logra disminuir la subjetividad del profesor-evaluador, lo cual se demuestra en la estabilidad de los resultados entre estas dos pruebas frente a los resultados de enero.

El diagrama de dispersión en relación con las correlaciones de estas pruebas es el siguiente:

Cuadro 26.



Como se puede comprobar, los puntos tienden a progresar desde abajo a la izquierda hacia arriba a la derecha, y aunque no es posible unir todos los puntos con una línea recta debido a los resultados del alumno A3 con una diferencia entre las notas de dos puntos (8-6), la coincidencia es importante entre los dos grupos de notas.

- Conclusiones del ciclo 3

A la vista de los resultados, los nuevos criterios de evaluación son más válidos y fiables que los propuestos para las pruebas de certificación³⁷. Lamentablemente, durante estos tres ciclos de investigación no se ha podido comprobar la fiabilidad de los correctores debido a la escasez de recursos humanos y a la falta de tiempo de otros profesores para que se implicaran en nuestra investigación como evaluadores con el fin de hacer una doble corrección de las pruebas. Sin embargo, creemos que esta fiabilidad se puede demostrar a través de las correlaciones hechas entre las pruebas vistas anteriormente en donde se verifica que dicha fiabilidad sí existe.

De todas formas, esto sí se pudo hacer en el siguiente ciclo donde habían dos evaluadores para calificar las pruebas de expresión oral de certificación aunque con los criterios de la Consejería de Educación.

Por otra parte, en cuanto a la fiabilidad y validez de las tareas diseñadas para las pruebas orales, tanto de clase y pruebas de progreso y dominio como de certificación, se fue comprobando mientras estas se iban elaborando a partir de unas preguntas que se planteaban antes o después de diseñar la prueba³⁸. Así, para comprobar la fiabilidad de estas se tuvo en cuenta si se cambiaban las características del lugar en el que se realizaban las pruebas según los diferentes días en los que se efectuaban, si habían diferentes formatos de pruebas o todas guardaban una cierta relación, en qué medida se cambiaba el input de una tarea a otra, en qué medida cambiaban las características de la respuesta esperada de una tarea a otra o hasta qué punto se modificaba la relación entre el input y la respuesta de una forma injustificada según las distintas pruebas. En la experimentación de estas pruebas similares a las de certificación se contestaron todas

³⁷ Hay que recordar que los criterios de evaluación y escalas de las pruebas de certificación los proporciona la Consejería de Educación sin posibilidad de cambiarlos por los nuestros, y que con esos criterios se evaluaron las pruebas de enero aunque modificando sólo la puntuación de los descriptores, con el fin de comprobar si era esta la que influía en la fiabilidad de la nota o el descriptor. Estos descriptores de la Consejería de Educación son uno de los motivos por los que se plantea el problema observado en la práctica docente y que genera esta investigación.

³⁸ A partir de las características propuestas por Bachman y Palmer (1996) para elaborar pruebas de calidad, se comprobó la fiabilidad y validez de las pruebas siguiendo algunos de los pasos propuestos por Martínez (2008b: 363-371).

estas preguntas de manera positiva, no observándose ningún problema en el desarrollo de estas. Además, para la validez de contenido y de constructo se comprobó la relación de estas con el constructo a evaluar, planteado al principio de este trabajo y en qué medida tanto estas como el sistema de corrección (escalas) reflejaba la definición del constructo, diseñándose cada tarea teniendo en cuenta unas habilidades concretas que debería provocar, o qué elementos del input, de la respuesta esperada o de la relación de ambos podían hacer variar la actuación de los candidatos, en relación a este último posiblemente la experiencia personal de los candidatos, podría variar la respuesta, pero en este caso el evaluador tendría que intervenir para conseguir el constructo definido.

En cuanto al análisis de la validez aparente o llamada por Bachman y Palmer (1996) “autenticidad”, la cual a su vez está relacionada con la validez de constructo, de contenido y de respuesta³⁹, se ha considerado en qué grado la descripción de las tareas en uso en contextos reales incluye información sobre el contexto, el tipo de input utilizado, la respuesta esperada o las relaciones de estas últimas. Asimismo para analizar la interactividad (Bachman y Palmer, 1996) de la prueba en su conjunto se ha considerado al candidato y sus características para la realización de la prueba así como el grado en que debe utilizar el conocimiento lingüístico para desarrollar las tareas teniendo en cuenta el nivel (Avanzado, B2) o cuánto tiempo se les otorgará para la planificación de estas. Además, de una manera razonable se analizó el impacto de la prueba atendiendo a las demandas de los estudiantes y a las sugerencias de los candidatos en el diseño de las tareas, se les ofreció al final de las tareas una retroalimentación clara que les sirviera como mejora de su actuación, en donde todos los candidatos fueron informados del sistema utilizado en la evaluación, se hicieron tareas similares a las que se evalúan en certificación y se aseguró que los criterios y los procedimientos se aplicaran igualmente a todos los candidatos, que las decisiones tomadas a partir de estas pruebas eran las adecuadas para cada tipo de examen o tarea (clase, progreso, dominio, certificación) y se estimó la repercusión de este tipo de pruebas en la actualidad y qué consecuencias tendría una falta de fiabilidad y validez de estas.

Por último, se estudió la viabilidad en el diseño y desarrollo de estas, para lo que se tuvo que calcular aproximadamente qué materiales se necesitaban y cuáles eran los

³⁹ Estos conceptos ya han sido explicados en los epígrafes 2.1.5. y 2.2.

recursos disponibles. De este último aspecto hablaremos en las conclusiones de nuestra investigación.

Después de todos estos análisis a lo largo de estos tres ciclos, se evidenció que ciertas actividades se habían formulado mejor que otras, por lo que se procedió a la modificación de estas. En general estaban bien elaboradas ya que cumplían las características de fiabilidad, validez, viabilidad, autenticidad, interactividad e impacto que señalaban Bachman y Palmer (1996) y estaban realizadas siguiendo el enfoque comunicativo de la lengua. También se advirtió que los criterios de evaluación analizados en un principio (los de la Consejería de Educación) no cumplían estas características por lo que hemos explicado con anterioridad. Así, a partir de estas observaciones se terminaron de elaborar las pruebas de exámenes orales de certificación junto a sus criterios de evaluación, modificados y mejorados respecto a los analizados en un principio.

4.2.4.4 Ciclo 4

- Administración de las pruebas de expresión oral de certificación de español de nivel B2 y evaluación de estas

A principio de junio se administraron las pruebas de certificación en las EOI de Canarias. Como ya hemos dicho, las EOI se encargaron de la elaboración de las pruebas repartiéndose entre ellas las distintas partes de las que consta bien por destrezas, bien por niveles. El departamento de español de esta EOI se encargó de la realización de las pruebas de expresión oral de certificación de todos los niveles. Lamentablemente, repetimos una vez más, no pudimos aplicar los criterios de evaluación elaborados que como resultado de nuestra investigación ya habíamos experimentado y comprobado en clase con los alumnos oficiales, sino que por directrices de la Consejería de Educación tuvimos que utilizar las escalas establecidas por ellos.

En cuanto a las tareas de expresión oral, se administraron al igual que como habíamos hecho en clase siguiendo su mismo modelo y secuenciación. La temática de las pruebas fueron las recogidas en el PCIC y el MCER para el nivel B2 que se limitaron a las nuevas tecnologías, la juventud, el alcohol, la familia y las relaciones personales, los tópicos, el arte, la cultura, la salud, la piratería musical, la historia, el cine, los deportes,

los recuerdos de la infancia y las asociaciones sin ánimo de lucro (ONG).⁴⁰ El procedimiento a seguir en la administración de las pruebas fue el siguiente:

Los alumnos pasaban a la clase donde iban a ser evaluados por parejas o tríos (dependiendo del número total de candidatos) y mantenían una breve presentación con el examinador para superar la tensión del primer momento. Posteriormente se les daba a cada uno una ficha que contenía la primera tarea del examen y que se correspondía con la producción de un monólogo. Se consideró que era conveniente empezar por tareas que sólo requerían expresión oral, sin interacción y con un breve tiempo de planificación para ordenar sus ideas. “De este modo se pretende rebajar la tensión y la ansiedad que suele crear este tipo de pruebas” (Bordón, 2006: 213). Se les advertía que poseían un tiempo de preparación de unos cinco minutos en donde podían incluso tomar notas para ordenar sus ideas aunque luego no podían leerlas sino utilizarlas como guión. Después de esos cinco minutos los candidatos debían producir un discurso oral monologado en relación con el tema que le había tocado en su ficha, primero actuaba un candidato y luego otro. A continuación se les administraba la segunda tarea de la prueba que se correspondía con una interacción entre los candidatos como si estuvieran en una situación real, ya que según la mayoría de los investigadores (Bachman y Palmer (1996), Vázquez (2000), Figueras (2004), Parrondo (2004), MCER (2002), Bordón (2006), PCIC (2007), Yagüe (2010)), “la materialización más “auténtica” de la expresión oral es la que reviste la forma de una conversación” (Bordón, 2006: 204). Para ello se les dejaba tres minutos escasos de preparación, ya que en el desarrollo del discurso podían llegar al acuerdo que generalmente les pedía la tarea. Después de esos tres minutos de planificación comenzaba la interacción que duraba unos 6 o 7 minutos para las parejas o unos 8 o 9 minutos para los tríos. Al finalizar la prueba se les recordaba cómo se iba a proceder para comunicarles sus resultados y abandonaban el aula.

El tribunal de evaluación estaba formado por dos profesores, uno de ellos hacía de interlocutor con los candidatos y grababa las intervenciones⁴¹, y el otro actuaba como examinador (siguiendo a Alderson *et al.*, 1998) pudiéndose centrar sólo en el análisis de la actuación de estos. Sin embargo, ambos profesores evaluaban independientemente a los alumnos para luego poner sus resultados en común y comprobar la fiabilidad de

⁴⁰ Para ver un modelo de estas tareas, consultar el anexo 5.

⁴¹ Sólo en audio a través de un ordenador.

estos. Si existían muchas diferencias entre los resultados establecidos por cada profesor-evaluador se hacía una nueva corrección gracias a la grabación de las actuaciones. El mismo proceso se repetía con cada pareja o trío de candidatos.

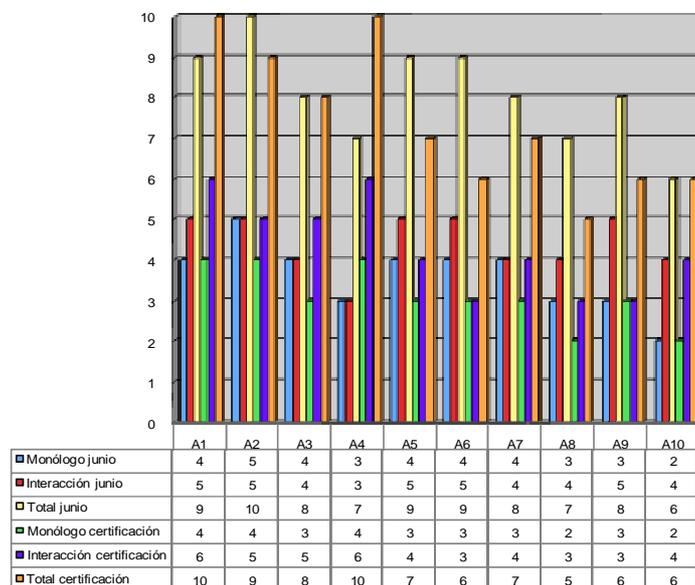
En cuanto a la agrupación de los participantes en las pruebas, a los alumnos oficiales se les permitió que se agruparan libremente según las afinidades entre ellos, ya que se conocían, y a los alumnos libres, como no se conocían ni los conocían los profesores-evaluadores, se agruparon para formar parejas o tríos de manera alfabética según la primera letra de su apellido. Esto quizá pudo hacer que los alumnos oficiales tuvieran cierta “ventaja” sobre los libres ya que pudieron sentirse más cómodos al actuar con personas conocidas.

Por último, cabe destacar que no fue posible que la profesora que les impartía clase no los evaluara, porque el departamento de español de la EOI Santa Cruz de Tenerife sólo está formado por dos profesores y el tribunal evaluador tenía que estar compuesto, también como mínimo, por dos profesores, por lo que la escasez de recursos humanos hizo que se produjera tal situación. Sin embargo, a partir de los resultados obtenidos y comparándolos con los de junio podremos determinar si la subjetividad influyó en los resultados debido, en mayor parte, a la imprecisión de los descriptores y en la poca relación del constructo con la puntuación de la prueba, ya que los criterios de evaluación aplicados a las pruebas de certificación no fueron los que habíamos elaborado y que resultaban más adecuados en relación con el constructo y la capacidad a evaluar sino los impuestos por la Consejería. Esto nos sirvió para comparar los resultados obtenidos en expresión oral en las pruebas de certificación de los alumnos oficiales mediante los criterios de evaluación de la Consejería y los obtenidos con los nuevos criterios de evaluación de nuestro estudio.

- Resultados de las pruebas de expresión oral de certificación de los alumnos oficiales y comparación con las pruebas de dominio de junio.

Los resultados obtenidos de los alumnos oficiales de nivel B2 en las pruebas de expresión oral de certificación aplicando los criterios de evaluación de la Consejería de Educación y comparados con los obtenidos en junio en las pruebas de dominio se pueden apreciar en el siguiente gráfico:

Cuadro 27.



A través del gráfico se pueden comprobar que algunas notas de las pruebas de certificación son superiores o inferiores a las obtenidas por los alumnos en las pruebas de junio hechas días antes a las de certificación, por lo que se comprueba de nuevo que los descriptores y las escalas que se utilizan para la puntuación de las muestras de lengua no son válidos, ya que contienen imprecisiones que llevan a puntuaciones poco acertadas. Esto se puede ver de una forma más clara evidenciando las correlaciones que existen entre los resultados de las dos pruebas:

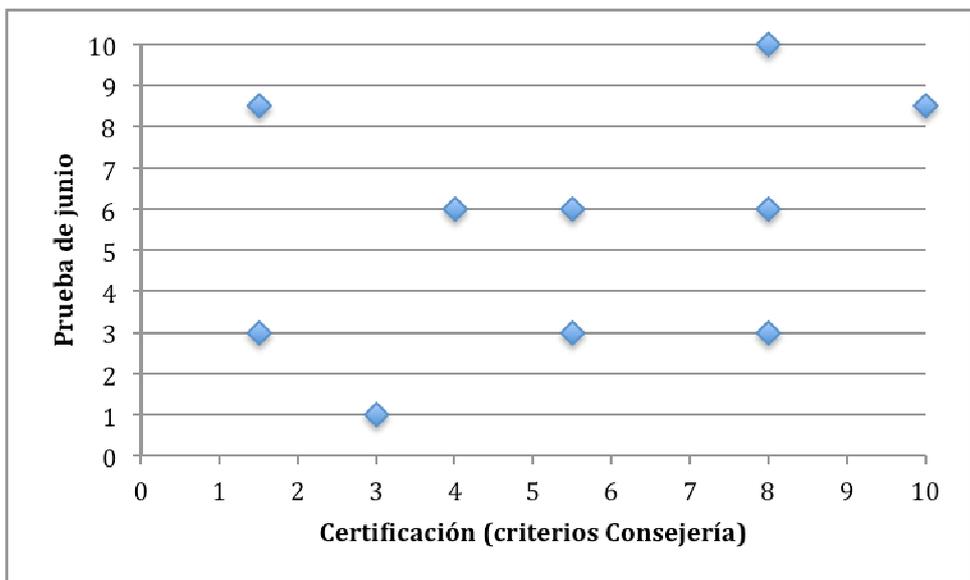
Cuadro 28: **Correlación entre las pruebas de junio y certificación (criterios de la Consejería) por orden de escala**

Alumno	Resultados junio	Resultados certificación	Escala junio	Escala certificación	d	d ²
A1	9	10	3	1,5	1,5	2,25
A2	10	9	1	3	1	1
A3	8	8	6	4	2	4
A4	7	10	8,5	1,5	7	49
A5	9	7	3	5,5	2,5	6,25
A6	9	6	3	8	5	25

A7	8	7	6	5,5	0,5	0,25
A8	7	5	8,5	10	1,5	2,25
A9	8	6	6	8	2	4
A10	6	6	10	8	2	4

d= diferencia; d²= diferencia al cuadrado

Cuadro 29.



La correlación entre las dos pruebas es de +0,4. Existe cierta relación pero ésta es muy baja, por lo que los resultados de certificación obtenidos por los candidatos no corresponden exactamente con su capacidad para expresarse oralmente evaluada y comprobada en clase. Como podemos observar en el gráfico de dispersión los puntos aparecen diseminados por toda el área del gráfico lo que indica una correlación bastante escasa.

- Doble corrección de los resultados de las pruebas de certificación con los resultados de las mismas pruebas evaluadas con los nuevos criterios de evaluación.

Debido a la escasa correlación con los resultados de las pruebas de certificación y las de clase, se volvieron a corregir las mismas pruebas de certificación pero ahora con los criterios de evaluación elaborados a partir de este trabajo, gracias a las grabaciones de audio efectuadas mientras se examinaban los candidatos y para comprobar también las correlaciones entre estas (certificación (nuevos criterios)/certificación (criterios de la Consejería)) y las de junio (junio/certificación(nuevos criterios)). Los resultados totales se muestran a continuación:

Cuadro 30:

ERROR: syntaxerror
OFFENDING COMMAND: --nostringval--

STACK:

-mark-
/sfnts