

9. La evaluación

9.1. Introducción

El término “evaluación” se utiliza en este capítulo con el sentido concreto de valoración del grado de dominio lingüístico que tiene el usuario. Todas las pruebas son una forma de evaluación, si bien hay otras muchas formas de evaluar (por ejemplo, las listas de control utilizadas en la evaluación continua, la observación cotidiana del profesor) que no podrían considerarse como “pruebas”. Evaluar es un concepto más amplio que medir o valorar la competencia o el dominio de la lengua. Toda medición o valoración es una forma de evaluación, pero en un programa de lenguas se evalúan aspectos, no propiamente del dominio lingüístico, como la eficacia de métodos y materiales concretos, el tipo y la calidad del discurso producido a lo largo del programa, la satisfacción del alumno y del profesor, la eficacia de la enseñanza, etc. Este capítulo se ocupa concretamente de la evaluación del dominio lingüístico y no de la evaluación en su sentido más amplio.

Hay tres conceptos que tradicionalmente se consideran fundamentales en cualquier análisis de la evaluación: la validez, la fiabilidad y la viabilidad. Resulta útil respecto al análisis de este capítulo tener una idea general de lo que quieren decir estos términos, cómo se relacionan entre sí y cómo se adecuan al *Marco de referencia*.

La *validez* es el concepto del que se ocupa el *Marco*. 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.

La *fiabilidad*, por otro lado, es un término técnico. Es básicamente el grado en que se repite el mismo orden de los candidatos en cuanto a las calificaciones obtenidas en dos convocatorias distintas (reales o simuladas) de la misma prueba de evaluación.

Pero más importante que la fiabilidad es la *precisión de las decisiones* que se adoptan en relación con un determinado nivel de exigencia. Si el informe de la evaluación da resultados tales como aprobado o suspenso, o Niveles A2+, B1 o B1+, ¿qué exactitud tienen estas decisiones? La exactitud de las decisiones depende de la validez que tenga el nivel concreto (por ejemplo, el nivel B1) para el contexto. También depende de la validez de los criterios utilizados para tomar la decisión y de la validez de los procedimientos mediante los cuales se aplicaron esos criterios.

Si dos organizaciones o regiones distintas utilizan criterios referidos a los mismos niveles de exigencia para determinar sus decisiones de evaluación con respecto a la misma destreza y si los niveles mismos son válidos y apropiados para los dos entornos implicados y se interpretan de forma consistente en el diseño de las tareas de evaluación y en la interpretación de las actuaciones, entonces los resultados de los dos sistemas guardarán correlación. Tradicionalmente, la correlación entre dos pruebas que se supone que evalúan el mismo constructo se conoce como «*validez concurrente*»; un concepto que se relaciona evidentemente

con la fiabilidad, pues las pruebas que no son fiables no guardan correlación. Sin embargo, lo fundamental es el grado de coincidencia entre las dos pruebas respecto a *lo que se evalúa* y a *cómo se interpreta la actuación*.

De estas dos cuestiones se ocupa el *Marco de referencia*. La siguiente sección describe tres formas principales de utilizar el *Marco*:

1. Para especificar el contenido de las pruebas y de los exámenes:	<i>lo que se evalúa</i>
2. Para establecer los criterios con los que se determina la consecución de un objetivo de aprendizaje:	<i>cómo se interpreta la actuación</i>
3. Para describir los niveles de dominio lingüístico en pruebas y exámenes existentes, permitiendo así realizar comparaciones entre distintos sistemas de certificados:	<i>cómo se pueden realizar las comparaciones</i>

Hay muchos tipos diferentes de evaluación y estos asuntos tienen que ver con ellos. Es un error suponer que un enfoque (por ejemplo, un examen público) es necesariamente superior en sus efectos educativos que (por ejemplo, la evaluación que realiza el profesor). Efectivamente, constituye una ventaja importante el hecho de que una serie de niveles comunes -como son los niveles comunes de referencia del *Marco*- posibiliten la relación entre sí de formas distintas de evaluación.

La tercera sección del capítulo presenta las opciones entre distintos tipos de evaluación. Estas opciones se presentan en forma de pares opuestos; en cada caso se definen los términos utilizados y se analizan las relativas ventajas y desventajas con respecto al propósito de la evaluación en su entorno educativo. También se exponen las consecuencias que conlleva el ejercicio de una u otra de las opciones alternativas, para después indicar la adecuación del *Marco de referencia* con el tipo de evaluación concreta.

Un procedimiento de evaluación también tiene que ser práctico, *viable*; la viabilidad tiene que ver concretamente con la evaluación de la actuación. Los examinadores trabajan bajo la presión del tiempo; sólo ven una muestra limitada de la actuación y existen límites definidos del tipo y del número de categorías que pueden manejar como criterios. El *Marco de referencia* pretende proporcionar un punto de referencia, no un instrumento práctico de evaluación; tiene que ser integrador, pero todos sus usuarios deben ser selectivos, lo que supondría el uso de un esquema operativo más simple que sintetice las categorías separadas del *Marco*. Por ejemplo, las categorías utilizadas en las escalas ilustrativas de descriptores que están yuxtapuestas al texto en los capítulos 4 y 5 son generalmente mucho más sencillas que las categorías y exponentes que se desarrollan en el texto mismo. Para ello, la última sección de este capítulo analiza este tema con ejemplos.

9.2. El Marco de referencia como recurso de evaluación

9.2.1. La especificación del contenido de las pruebas y los exámenes

Se puede consultar la descripción de «El uso de la lengua y el usuario o alumno» realizada en el capítulo 4, en concreto en la sección 4.4, que se ocupa de las actividades lingüísticas comunicativas, cuando se vaya a preparar la especificación de una tarea para una evaluación comunicativa. Se va aceptando paulatinamente que la evaluación auténtica requiere el muestreo de una serie de tipos de discurso relevantes; por ejemplo, respecto a los exámenes de expresión oral, una prueba elaborada recientemente es ilustrativa al respecto. En primer lugar, hay una *conversación* simulada que funciona como introducción y después se da un *debate informal* sobre temas por los que el examinando declara su interés. A esto le sigue una fase de *transacción* en forma de actividad -ya sea cara a cara o simulada- de búsqueda de información por teléfono. A continuación, se desarrolla una fase de *expresión* basada en un *informe* escrito en el que el examinando ofrece una *descripción* de su área de especialidad académica y de sus planes. Por último, hay una *cooperación centrada en el objetivo*, una tarea de consenso entre los examinandos.

En resumen, las categorías que el *Marco de referencia* utiliza para las actividades comunicativas son las siguientes:

	Interacción (Espontánea, turnos breves)	Expresión (Preparada, turnos largos)
Orales:	<i>Conversación</i> <i>Debate informal</i> <i>Cooperación centrada en el objetivo</i>	<i>Descripción</i> de su especialidad académica
Escritas		<i>Informe o descripción</i> de su especialidad académica

Para elaborar los detalles de las especificaciones de la tarea, el usuario puede consultar la sección 4.1, «el contexto del uso de la lengua» (ámbitos, condiciones y restricciones, contexto mental), la sección 4.6, que se ocupa de los «textos» y el capítulo 7, que analiza «las tareas y su papel en la enseñanza de la lengua», sobre todo la sección 7.3, que estudia «la dificultad de la tarea».

La sección 5.2, que se ocupa de «las competencias lingüísticas comunicativas» perfecciona la elaboración de los ítems de la prueba o de las fases de una prueba oral con el fin de conseguir evidencia de las competencias lingüística, sociolingüística y pragmática relevantes. El conjunto de especificaciones de contenidos de *Threshold Level*, desarrolladas por el Consejo de Europa para más de veinte lenguas europeas (véase la bibliografía para el capítulo 2 en el apartado de Bibliografía general), así como de los niveles *Waystage* y *Vantage* para el inglés, más sus equivalentes en el caso de haberse desarrollado para otras lenguas y niveles, pueden considerarse como auxiliares del documento principal del *Marco de referencia*. En estas especificaciones se incluyen ejemplos, en un nivel más detallado, que ofrecen orientaciones útiles para elaborar pruebas y exámenes de los niveles A1, A2, B1 y B2.

9.2.2. Criterios para el logro de los objetivos de aprendizaje

Las escalas constituyen una fuente para el desarrollo de escalas de valoración que evalúan el logro de un objetivo concreto de aprendizaje, y los descriptores pueden contribuir a la formulación de los criterios. El objetivo puede ser un nivel amplio de dominio lingüístico general, expresado en un nivel común de referencia (por ejemplo, B1). Puede ser, por el contrario, un conjunto de actividades, destrezas y competencias, como se analizó en la sección 6.1.2, que se ocupa de «las competencias parciales y la variedad de objetivos con relación al Marco de referencia». Un objetivo modular de este tipo se podría perfilar en una “parrilla” de categorías por niveles, como la que se ha presentado en el Cuadro 2.

Al analizar la utilización de descriptores, resulta esencial distinguir entre:

1. Descriptores de actividades comunicativas, que se encuentran en el capítulo 4.
2. Descriptores de aspectos del dominio de la lengua relacionados con competencias concretas, que se encuentran en el capítulo 5.

Los primeros son muy apropiados para la evaluación que realiza el profesor o para la autoevaluación respecto a tareas de la vida real. Estos dos tipos de evaluación se realizan sobre la base de una representación muy pormenorizada de la capacidad lingüística del alumno desarrollada durante el curso en cuestión y son atractivos porque pueden ayudar tanto a alumnos como a profesores a centrarse en un enfoque orientado a la acción.

Sin embargo, *no* suele ser aconsejable incluir descriptores de actividades comunicativas en los criterios que utiliza un examinador para medir o valorar la actuación en una prueba concreta de comprensión oral o escrita si lo que interesa es informar de resultados en cuanto al nivel de dominio lingüístico alcanzado. Esto es así porque, para informar sobre el dominio lingüístico, la evaluación no debería tener en cuenta principalmente ninguna actuación concreta, sino que debería intentar valorar las competencias generalizables que se muestran en esa actuación. Naturalmente, tiene que haber sólidas razones de carácter educativo para centrarse en lograr el éxito en la realización de una actividad dada, sobre todo en el caso de los usuarios básicos más jóvenes (niveles A1 y A2). Dichos resultados serán menos generalizables, pero el carácter generalizable de los resultados no suele ser el centro de atención en las primeras etapas del aprendizaje de la lengua.

Esto refuerza el hecho de que las valoraciones pueden tener muchas funciones diferentes; lo que resulta apropiado para una finalidad puede no serlo para otra.

9.2.2.1. Descriptores de actividades comunicativas

Los descriptores de actividades comunicativas (capítulo 4) se pueden utilizar de tres formas distintas respecto al logro de los objetivos.

1. **Construcción:** Como se vio en la sección 9.2.1, las escalas de las actividades comunicativas ayudan a definir las especificaciones para el diseño de tareas de evaluación.

2. **Informe:** Las escalas de las actividades comunicativas también pueden ser muy útiles para informar de los resultados. A las personas interesadas en los resultados del sistema educativo como, por ejemplo, los empresarios, suelen interesarles los resultados de conjunto más que un perfil detallado de competencia.
3. **Autoevaluación del alumno y evaluación por el profesor:** Por último, los descriptores de las actividades comunicativas se pueden utilizar para la autoevaluación que realiza el propio alumno y para la evaluación que realiza el profesor de varias maneras, de las cuales ofrecemos algunos ejemplos:
 - **Lista de control:** Para la evaluación continua o para la evaluación sumativa al final de un curso. Los descriptores de un nivel concreto se pueden recoger en una lista. Como alternativa, el contenido de los descriptores se puede «fragmentar». Por ejemplo, el descriptor *Pide y ofrece información personal* puede ser fragmentado en componentes explícitos como *Sé presentarme; digo dónde vivo; digo mi dirección en francés; digo la edad que tengo*, etc., y *Pregunto a alguien cómo se llama; pregunto a alguien dónde vive; pregunto a alguien qué edad tiene*, etc.
 - **“Parrilla”:** Para la evaluación continua o la sumativa, definiendo un perfil en una “parrilla” de categorías seleccionadas (por ejemplo: *Conversación; Debate; Intercambio de información*) y definidas en distintos niveles (B1+, B2, B2+).

Este uso de los descriptores ha venido siendo cada vez más habitual durante los últimos diez años. La experiencia ha demostrado que la consistencia con que profesores y alumnos pueden interpretar los descriptores se mejora si estos describen no sólo LO QUE el alumno sabe hacer sino también EN QUÉ MEDIDA lo hace bien o mal.

9.2.2.2. Descriptores de aspectos del dominio lingüístico referidos a competencias concretas

Los descriptores de aspectos del dominio lingüístico se pueden utilizar de dos formas en relación con el logro de los objetivos.

- **Autoevaluación del alumno y evaluación por el profesor:** Siempre que los descriptores sean enunciados positivos e independientes, se pueden incluir en listas de control para la autoevaluación del alumno o para la evaluación por el profesor. Sin embargo, la mayoría de las escalas existentes tienen el problema de que los descriptores a menudo están redactados de forma negativa en los niveles inferiores y con referencia a la norma hacia la mitad de la escala. También establecen, a menudo, distinciones puramente verbales entre niveles sustituyendo una o dos palabras de descripciones contiguas que tienen poco significado fuera del texto de la escala. En el anejo A se analizan algunas formas de desarrollar descriptores que eviten estos problemas.
- **Valoración de la actuación:** Un uso más evidente de las escalas de descriptores del capítulo 5 que tratan aspectos de las competencias es ofrecer puntos de partida para el desarrollo de criterios de evaluación. Transformando las impresiones personales -no sistemáticas- en juicios ponderados, dichos descriptores pueden contribuir al desarrollo de un marco compartido de referencia entre un grupo de examinadores.

Existen básicamente tres formas de presentar los descriptores para que se puedan utilizar como criterios de evaluación:

- En primer lugar, se pueden presentar los descriptores como una *escala*, a menudo combinando descriptores de distintas categorías en un párrafo holístico por cada nivel. Éste enfoque es muy frecuente.
- En segundo lugar, se pueden presentar como una *lista de control*, generalmente una lista por nivel, a menudo con descriptores agrupados bajo encabezamientos, es decir, en categorías. Las listas de control son menos corrientes para la evaluación que se realiza en el acto.
- En tercer lugar, se pueden presentar como una “*parrilla*” de categorías seleccionadas, en realidad como un conjunto de escalas paralelas para categorías separadas. Este enfoque posibilita un perfil de diagnóstico. Sin embargo, existen límites al número de categorías que los examinadores pueden manejar.

Hay dos formas diferenciadas de proporcionar un cuadro de subescalas:

- *Escala de dominio*: Se presenta un cuadro descriptivo que defina los niveles adecuados para determinadas categorías; por ejemplo, desde el nivel A2 al B2. La evaluación entonces se realiza directamente en esos niveles, utilizando posiblemente otras mejoras como, por ejemplo, un segundo dígito o signos de suma para conseguir una mayor diferenciación si se desea. De ese modo, aunque la prueba de actuación fuera dirigida al nivel B1, y aunque ninguno de los alumnos alcanzase el nivel B2, todavía sería posible que los mejores alumnos consiguieran el nivel B1+, B1++ o B1.8.
- *Escala de valoración*: Se selecciona o define un descriptor para cada una de las categorías que describa el nivel de exigencia requerido para aprobar un módulo o un examen concreto de esa categoría. A ese descriptor se le denomina entonces «Aprobado» o «3» y la escala es referida a la norma respecto a ese nivel (una actuación muy deficiente = «1»; una excelente actuación = «5»). La formulación de «1» y «5» podrían constituirlos otros descriptores extraídos o adaptados de los niveles contiguos de la escala que aparece en la sección apropiada del capítulo 5; también el descriptor puede ser formulado en relación con la redacción del descriptor definido como «3».

9.2.3. Descripción de los niveles de dominio de la lengua para facilitar la comparación de pruebas y exámenes

Las escalas de los niveles comunes de referencia pretenden facilitar la descripción del nivel de dominio lingüístico alcanzado en los diplomas y certificados existentes y de esa forma permitir la comparación entre sistemas. Los estudios de medición reconocen cinco formas clásicas de relacionar evaluaciones separadas: (1) equiparación; (2) calibración; (3) moderación estadística; (4) punto de referencia, y (5) moderación social.

Los tres primeros métodos son tradicionales: (1) producción de versiones alternativas de la misma prueba (equiparación), (2) relación de los resultados de distintas pruebas en una escala común (calibración), y (3) corrección según la dificultad de las pruebas o la exigencia de los examinadores (moderación estadística).

Los dos últimos métodos suponen conseguir un cierto entendimiento mediante el debate (moderación social) y la comparación de muestras de trabajo en relación con definiciones normalizadas y ejemplos (punto de referencia). El apoyo a este proceso para la consecución de un entendimiento es uno de los objetivos del *Marco de referencia*. Por este motivo, las escalas de descriptores que se utilizan para este propósito han sido normalizadas con una metodología rigurosa de desarrollo. En el campo de la educación este enfoque se está describiendo cada vez más como evaluación centrada en niveles; se da generalmente por sentado que el desarrollo del enfoque centrado en niveles lleva tiempo, pues los participantes adquieren un sentido del significado de los niveles mediante el proceso de ejemplificación e intercambio de opiniones.

Se puede aducir que este enfoque es potencialmente el método más consistente de relación porque supone el desarrollo y la validación de una visión común del constructo. El motivo fundamental por el que resulta difícil relacionar evaluaciones lingüísticas, a pesar de la hechicería estadística de las técnicas tradicionales, es que las evaluaciones generalmente valoran elementos radicalmente distintos, incluso cuando pretenden abarcar los mismos ámbitos. Esto es, en parte, debido a (a) una escasa conceptualización y activación del constructo, y en parte a (b) una interferencia derivada del método de evaluación.

El *Marco de referencia* ofrece unos principios para solucionar el primer problema en relación con el aprendizaje de lenguas modernas en un contexto europeo. Los capítulos del 4 al 7 elaboran un esquema descriptivo que intenta categorizar de forma práctica el uso de la lengua, las competencias y el proceso de enseñanza y aprendizaje, de manera que ayude a los alumnos a poner en funcionamiento la capacidad lingüística comunicativa que queremos fomentar.

Las escalas de descriptores configuran un cuadro conceptual que se puede utilizar para:

- Relacionar sistemas nacionales e institucionales entre sí mediante el *Marco de referencia*.
- Planear los objetivos de exámenes concretos y de módulos de curso utilizando las categorías y los niveles de las escalas.

El anejo A ofrece a los lectores una visión general de los métodos utilizados para desarrollar escalas de descriptores y relacionarlos con la escala del *Marco*.

La Guía del usuario para examinadores elaborada por ALTE (Documento CC-Lang (96) 10 revisado) proporciona unos consejos pormenorizados sobre el modo de hacer operativos los constructos de las pruebas, así como sobre la forma de evitar distorsiones innecesarias derivadas de los efectos del método de la prueba.

9.3. Tipos de evaluación

Se pueden realizar varias distinciones importantes respecto a la evaluación. La lista siguiente no es de ninguna manera exhaustiva y no existe ninguna distinción relativa a la colocación de un término en la columna de la izquierda o en la de la derecha.

Cuadro 7. Tipos de evaluación

1	Evaluación del aprovechamiento	Evaluación del dominio
2	Con referencia a la norma (RN)	Con referencia a un criterio (RC)
3	Maestría RC	<i>Continuum RC</i>
4	Evaluación continua	Evaluación en un momento concreto
5	Evaluación formativa	Evaluación sumativa
6	Evaluación directa	Evaluación indirecta
7	Evaluación de la actuación	Evaluación de los conocimientos
8	Evaluación subjetiva	Evaluación objetiva
9	Valoración mediante lista de control	Valoración mediante escala
10	Impresión	Valoración guiada
11	Evaluación global	Evaluación analítica
12	Evaluación en serie	Evaluación por categorías
13	Evaluación realizada por otras personas	Autoevaluación

9.3.1. Evaluación del aprovechamiento / evaluación del dominio

La *evaluación del aprovechamiento* es la evaluación del grado en que se han alcanzado objetivos específicos, es decir, la evaluación de lo que se ha enseñado. Se relaciona, por tanto, con el trabajo semanal o trimestral, con el manual, con el programa. y está orientada al curso, representa una perspectiva interna.

La *evaluación del dominio*, por otro lado, es la evaluación de lo que alguien sabe o es capaz de hacer en cuanto a la aplicación en el mundo real de lo que ha aprendido; representa, por tanto, una perspectiva externa.

Los profesores tienen una tendencia natural a interesarse más por la evaluación del aprovechamiento, con el fin de conseguir retroalimentación para la enseñanza. A los empresarios, los administradores educativos y a los alumnos adultos les suele interesar más la evaluación del dominio lingüístico, esto es, la evaluación de los resultados, lo que la persona sabe hacer ahora. La ventaja de la evaluación del aprovechamiento es que se encuentra cerca de la experiencia del alumno. La ventaja de la evaluación del dominio es que ayuda a todos a ver dónde están; los resultados, así, son transparentes.

En la evaluación de carácter comunicativo dentro de un enfoque de enseñanza y aprendizaje centrado en las necesidades, se puede argumentar que la distinción entre el aprovechamiento (centrado en el contenido del curso) y el dominio lingüístico (centrado en la capacidad para desenvolverse en el mundo real) debería ser teóricamente pequeña. La evaluación del aprovechamiento tiene un componente de dominio en la medida en que considera el uso práctico de la lengua en situaciones adecuadas y pretende ofrecer una imagen equilibrada de la

competencia emergente. Según una evaluación de dominio se compone de elementos lingüísticos y de tareas comunicativas basadas en un programa transparente y adecuado y ofrece al alumno la oportunidad de mostrar lo que ha logrado, tiene un elemento de aprovechamiento.

Las escalas de descriptores ilustrativos se relacionan con la evaluación del dominio, esto es, con la capacidad para desenvolverse en el mundo real. En el capítulo 6 se analiza la importancia de la evaluación del aprovechamiento como un refuerzo del aprendizaje.

9.3.2. *Con referencia a la norma (RN) / con referencia a un criterio (RC)*

La clasificación de los alumnos por grado o calidad, eso es *con referencia a la norma*, supone que la valoración y la posición de cada uno se establece con respecto a los compañeros.

Como una reacción contra la referencia a la norma está la clasificación *con referencia a un criterio*, de este modo el alumno es evaluado meramente en función de su capacidad en la asignatura de que se trate, sin tener en cuenta la capacidad de sus compañeros.

La *referencia a la norma* se puede entender en relación con la clase ("tú haces el número 18") o en relación demográfica ("tú eres el 21.567"; "estás entre el 14 por ciento superior") o en relación con el grupo de alumnos que realizan la prueba. En este último caso, hay que adaptar las puntuaciones de las pruebas para ofrecer un resultado «justo» comparando el trazado de la curva de distribución de los resultados de la prueba con la curva de los años anteriores para mantener un patrón y asegurarse de que el mismo porcentaje de alumnos consigue todos los años calificaciones de «sobresaliente», sin tener en cuenta la dificultad de la prueba o la capacidad de los alumnos. Un uso habitual de la evaluación referida a la norma ocurre en las pruebas de clasificación para formar los grupos de los distintos niveles.

La *referencia a un criterio* supone partir de un cuadro en cuyo eje vertical se sitúa el grado de dominio lingüístico (como en una línea continua) y en cuyo eje horizontal se recoge la serie de ámbitos relevantes, de manera que los resultados individuales de una prueba puedan situarse en relación con el mapa total de criterios. Esto supone: (a) la definición del ámbito o ámbitos adecuados que son el objeto de la prueba o el módulo concreto, y (b) la determinación de los «puntos de corte»: la puntuación o puntuaciones de la prueba que se consideran necesarias para alcanzar el nivel de dominio establecido de la prueba.

Las escalas de descriptores ilustrativos se componen de especificaciones de criterio de las categorías que conforman el esquema descriptivo y los niveles comunes de referencia presentan un conjunto de patrones comunes.

9.3.3. *Maestría RC / Continuum RC*

El enfoque de la *maestría con referencia a un criterio* es un enfoque en el que se establece un solo «nivel mínimo de competencia» o «punto de corte» para dividir a los alumnos entre los que han alcanzado la maestría y los que no, sin ningún tipo de gradación de calidad respecto al logro del objetivo establecido.

El enfoque del *continuum con referencia a un criterio* es un enfoque en el que una capacidad individual se sitúa en referencia a una línea continua con todos los grados de capacidad pertinentes para evaluar un área determinada.

Existen en realidad muchos enfoques con referencia a un criterio, pero se puede definir la mayoría de ellos principalmente desde una interpretación bien de «maestría» o bien de «continuum». Se produce mucha confusión cuando se identifica erróneamente y de modo exclusivo la referencia a un criterio con el enfoque de maestría; el enfoque de maestría es un enfoque de aprovechamiento relacionado con el contenido de un curso o de un módulo determinado y pone menor énfasis en asignarle a ese módulo (y, por tanto, a su aprovechamiento) un grado de dominio lingüístico.

La alternativa al enfoque de maestría consiste en establecer una relación entre los resultados de cada prueba con el correspondiente grado de dominio lingüístico, normalmente una serie de calificaciones. Desde este enfoque, ese *continuum* que detalla los grados de dominio representa el «criterio», es decir, la realidad externa que asegura que los resultados de la prueba significan algo. La referencia a este criterio externo se puede llevar a cabo con un análisis escalar (por ejemplo, el modelo de Rasch) para relacionar los resultados de todas las pruebas entre sí y, de esa manera, contrastar los resultados directamente con referencia a una escala común.

El *Marco de referencia* se puede utilizar con un enfoque de *maestría* o de *continuum*. La escala de niveles utilizados en el enfoque de *continuum* se puede equiparar con los niveles comunes de referencia y el objetivo que hay que alcanzar en el enfoque de *maestría* se puede describir en el cuadro conceptual de las categorías y los niveles ofrecidos por el *Marco de referencia*.

9.3.4. *Evaluación continua / evaluación en un momento concreto*

La *evaluación continua* es la evaluación que realiza el profesor, y puede que también el alumno, respecto a las actuaciones en clase, los trabajos y los proyectos realizados a lo largo del curso. La calificación final, por tanto, refleja todo el curso o semestre.

La *evaluación en un momento concreto* consiste en dar calificaciones y tomar decisiones sobre la base de un examen o de otro tipo de evaluación, que tiene lugar un día concreto, generalmente al final de un curso o antes de su comienzo. Lo que ha ocurrido con anterioridad, sin embargo, es irrelevante; lo decisivo es lo que la persona sabe hacer ahora, en ese momento preciso.

A menudo se considera la evaluación como algo exterior al curso, algo que tiene lugar en momentos determinados con el fin de tomar decisiones; la evaluación continua es una evaluación que está integrada en el curso y que contribuye de forma acumulativa a la evaluación final del curso. Aparte de calificar los deberes de casa y las pruebas breves de aprovechamiento, que pueden ser esporádicas o habituales, la evaluación continua se puede realizar de las siguientes formas: mediante listas de control o “parrillas” cumplimentadas por los profesores o por los alumnos, mediante la evaluación de una serie de tareas específicas, mediante la evaluación formal sobre la base del libro de texto o manual, y mediante una carpeta de muestras de trabajo, en fases distintas de acabado y en diferentes momentos del curso.

Ambos enfoques tienen ventajas e inconvenientes. La evaluación en un momento concreto asegura que las personas todavía saben hacer cosas que podrían haber aprendido dos años antes. Sin embargo, provoca traumas por el examen y favorece a determinados tipos de alumnos. La evaluación continua, por su parte, permite que se tenga más en cuenta la creatividad y las distintas cualidades, pero depende mucho más de la capacidad que tenga el profesor de ser objetivo; si se lleva a un extremo, puede convertir la vida en una prueba interminable para el alumno y en una pesadilla burocrática para el profesor.

Las listas de especificaciones de criterios que describen la capacidad respecto a las actividades comunicativas (capítulo 4) pueden resultar útiles para la evaluación continua. Por otra parte, las escalas de valoración desarrolladas a partir de los descriptores de aspectos de la competencia (capítulo 5) se pueden utilizar para dar calificaciones en la evaluación realizada en momentos concretos.

9.3.5. *Evaluación formativa /evaluación sumativa*

La *evaluación formativa* es un proceso continuo de acopio de información sobre el alcance del aprendizaje, así como sobre los puntos fuertes y débiles, que el profesor puede introducir en su planificación del curso y en la retroalimentación efectiva que da a los alumnos. La evaluación formativa se utiliza a menudo en un sentido amplio para incluir información no cuantificable proveniente de cuestionarios y consultas.

La *evaluación sumativa* resume el aprovechamiento al final del curso con una calificación. No es necesariamente una evaluación de dominio lingüístico; de hecho, gran parte de la evaluación sumativa es una evaluación referida a la norma, realizada en un momento concreto y de aprovechamiento.

La cualidad que tiene la evaluación formativa es que pretende mejorar el aprendizaje; su debilidad, sin embargo, es inherente a la idea misma de la retroalimentación. La retroalimentación sólo funciona si el receptor está en situación (a) *de fijarse*, es decir, está atento, motivado y acostumbrado a la forma en que llega la información; (b) *de recibir*, es decir, no está inundado de información y tiene una forma de registrarla, organizarla y personalizarla.; (c) *de interpretar*, es decir, tiene suficientes conocimientos previos y es suficientemente consciente para comprender el asunto cuestionado y no realizar acciones contraproducentes y (d) *de integrar* la información, es decir, tiene tiempo, orientación y recursos adecuados para reflexionar, integrar y, de este modo, recordar la información nueva. Esto supone cierta capacidad de dirigirse a uno mismo, lo que requiere la preparación para conseguir tal dirección autónoma, el control del propio aprendizaje y el desarrollo de formas de actuar según la retroalimentación recibida.

Esta preparación o formación del alumno, esto es, la toma de conciencia, se ha venido denominando "evaluación formativa". Se puede utilizar una variedad de técnicas que ayudan a prepararse para la toma de conciencia; un principio básico consiste en comparar la impresión (por ejemplo, lo que dice que sabe hacer de una lista) con la realidad (por ejemplo, escuchar realmente el material del tipo mencionado que aparece en la lista de control y comprobar si lo entiende de verdad). *DIALANG* relaciona de esta forma la autoevaluación con la actuación en un examen. Otra técnica importante es analizar muestras de trabajo —tanto ejemplos neutros como

muestras provenientes de alumnos— y fomentar en los alumnos el desarrollo de un metalenguaje personalizado sobre el nivel de calidad deseable; en ese caso, pueden utilizar este metalenguaje para hacer un seguimiento de su trabajo en busca de puntos fuertes y débiles y formular un contrato de aprendizaje autodirigido.

Gran parte de la evaluación formativa o de diagnóstico opera en un nivel muy pormenorizado de los aspectos lingüísticos o destrezas concretas que se han enseñado recientemente o que se van a enseñar pronto. Para la evaluación de diagnóstico las listas de exponentes presentadas en la sección 5.2. son todavía demasiado generales como para que sean prácticas; sería necesario referirse a la especificación concreta que fuera pertinente (Plataforma, Umbral, etc.). Sin embargo, las escalas de descriptores que definen aspectos distintos de la competencia en diferentes niveles (capítulo 4) pueden resultar útiles para ofrecer retroalimentación formativa procedente de la evaluación de la expresión oral.

Aunque podría parecer que los niveles comunes de referencia serían más adecuados para la evaluación sumativa, la retroalimentación procedente incluso de la evaluación sumativa puede ser de diagnóstico y, por ello, formativa, tal y como demuestra el Proyecto *DIALANG*.

9.3.6. Evaluación directa / evaluación indirecta

La *evaluación directa* es la evaluación de lo que está haciendo realmente el alumno que se somete a examen. Por ejemplo, un grupo reducido está debatiendo algo, el examinador observa, compara con una "parrilla" de criterios, relaciona las actuaciones con las categorías más adecuadas de la parrilla y evalúa.

La *evaluación indirecta*, por el contrario, utiliza una prueba, habitualmente en papel, que a menudo evalúa las destrezas.

La evaluación directa está limitada en la práctica a la expresión oral, la expresión escrita y la comprensión oral en interacción, pues nunca se puede observar directamente la actividad de comprensión. La comprensión escrita, por ejemplo, sólo puede ser evaluada indirectamente pidiendo a los alumnos que evidencien la comprensión marcando casillas, completando frases, contestando preguntas, etc. El alcance y el control lingüísticos se pueden evaluar, bien directamente mediante la valoración de su equiparación con los criterios, o bien indirectamente mediante la interpretación y la generalización, a partir de las respuestas a las preguntas de una prueba. Una prueba directa clásica es una entrevista; una prueba indirecta clásica es un ejercicio de tipo *cloze* o de rellenar huecos.

Los descriptores del capítulo 5 que definen distintos aspectos de la competencia en niveles diferentes pueden ser utilizados para desarrollar criterios de evaluación en pruebas directas. Los parámetros del capítulo 4 pueden servir de base para la selección de temas, textos y tareas para pruebas de evaluación directa de las destrezas de expresión y pruebas de evaluación indirecta de destrezas de comprensión oral y escrita. Los parámetros del capítulo 5 pueden además permitir la identificación de competencias lingüísticas clave que pueden incluirse en una prueba indirecta de conocimientos lingüísticos, así como de competencias clave de carácter pragmático, sociolingüístico y lingüístico en las que centrarse a la hora de formular las preguntas para realizar pruebas de las cuatro destrezas basadas en ítems.

9.3.7. Evaluación de la actuación / evaluación de los conocimientos

La *evaluación de la actuación* requiere que el alumno proporcione una muestra lingüística de forma hablada o escrita por medio de una prueba directa.

La *evaluación de los conocimientos* requiere que el alumno conteste preguntas que pueden ser de una serie de diferentes tipos de ítem para proporcionar evidencia sobre el alcance de sus conocimientos y su grado de control lingüístico.

Por desgracia, nunca se pueden evaluar las competencias directamente. Hay que basarse siempre en una serie de actuaciones a partir de las cuales se intenta generalizar una idea del dominio lingüístico, que se puede considerar como competencia “en uso”, llevada a la práctica. En este sentido, por tanto, todas las pruebas evalúan sólo la actuación, aunque partiendo de esa evidencia se puede intentar extraer inferencias respecto a las competencias que subyacen.

Sin embargo, una entrevista requiere más «actuación» que completar los huecos de frases, y, a su vez, completar huecos requiere más «actuación» que elegir entre varias opciones. En este sentido, la palabra «actuación» se viene utilizando para referirse a la expresión lingüística, pero se utiliza con un sentido más limitado en la expresión «pruebas de actuación»; aquí la palabra se refiere a una actuación adecuada en una situación (relativamente) auténtica y a menudo relacionada con el trabajo o el estudio. En un sentido ligeramente más impreciso de este término que en «evaluación de la actuación», se podría decir que los procedimientos de evaluación oral son pruebas de actuación en cuanto que hacen generalizaciones respecto al dominio lingüístico a partir de actuaciones de una serie de estilos de discurso que se consideran adecuados al contexto y a las necesidades de aprendizaje de los alumnos. Algunas pruebas equilibran la evaluación de la actuación con una evaluación de los conocimientos de la lengua como sistema, mientras que otras no lo hacen.

Esta distinción es muy parecida a la que existe entre pruebas directas e indirectas. El *Marco de referencia* se puede utilizar de forma parecida; las especificaciones del Consejo de Europa de distintos niveles (Plataforma, Umbral, Avanzado) ofrecen además, en las lenguas en las que están disponibles, una descripción detallada y adecuada del conocimiento de la lengua meta.

9.3.8. Evaluación subjetiva / evaluación objetiva

La *evaluación subjetiva* es una valoración que realiza un examinador. Lo que normalmente se entiende por esto es la valoración de la calidad de una determinada actuación.

La *evaluación objetiva* es la evaluación a la que se le ha despojado de la subjetividad. Lo que normalmente se entiende por esto es una prueba indirecta en la que los ítems tienen sólo una respuesta correcta, como, por ejemplo, los ejercicios para elegir una opción correcta entre varias.

Sin embargo, el asunto de la subjetividad y la objetividad es bastante más complejo. A menudo, se describe la prueba indirecta como «prueba objetiva» cuando el corrector consulta una clave predefinida para decidir si acepta o rechaza una respuesta y después cuenta las respuestas correctas para dar un resultado. Algunos tipos de pruebas profundizan en este proceso

ofreciendo sólo una respuesta posible para cada pregunta (por ejemplo: elegir una opción entre varias, o bien las pruebas de tipo *c*, que se derivaron de los ejercicios de tipo *cloze* por este motivo) y, a menudo, adoptan la corrección automática para eliminar los errores del corrector. En realidad, la objetividad de las pruebas descritas de esta forma como «objetivas» es algo exagerada, pues fue una persona quien decidió restringir la evaluación a técnicas que ofrecen más control sobre la situación de la prueba (en sí una decisión subjetiva que puede que otros no admitan). Después, alguien escribió la especificación de la prueba, y puede que otra persona escribiera el ítem intentando hacer operativo un punto concreto de la especificación. Por último, alguien eligió el ítem de entre todos los ítems posibles para esta prueba. Puesto que todas estas decisiones suponen un elemento de subjetividad, es mejor describir dichas pruebas como pruebas puntuadas objetivamente.

En la evaluación directa de la actuación generalmente se otorgan calificaciones según un juicio o valoración. Eso significa que la decisión respecto a la calidad de la actuación del alumno se toma subjetivamente teniendo en cuenta factores concretos y haciendo referencia a líneas generales o bien a determinados criterios o a la propia experiencia. La ventaja de un enfoque subjetivo es que la lengua y la comunicación son muy complejas, no permiten su atomización y son mayores que la suma de sus partes. Muy a menudo resulta difícil establecer qué es lo que está evaluando realmente un ítem; por lo tanto, dirigir los ítems de las pruebas hacia aspectos específicos de competencia o actuación no es tan sencillo como parece.

No obstante, y para ser justos, 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, sobre todo cuando se trata de una evaluación sumativa. 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.

Se siguen los siguientes pasos, se puede reducir la subjetividad de la evaluación, para aumentar así la validez y la fiabilidad, si:

- 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**.

Como se vio al principio de este capítulo, el primer paso que hay que dar para reducir la subjetividad de las valoraciones realizadas en todas las etapas del proceso de evaluación es desarrollar una comprensión común del constructo implicado, un marco común de referencia. El *Marco de referencia* pretende ofrecer dicha base para la **especificación del contenido** y quiere ser una fuente de desarrollo de **criterios específicos definidos** para las pruebas directas.

9.3.9. Valoración mediante escala /valoración mediante lista de control

Valoración mediante escala consiste en determinar que una persona está en un nivel o banda concreta de una escala compuesta por algunos de estos niveles o bandas.

Valoración mediante lista de control consiste en evaluar a una persona en relación con una lista de aspectos que se consideran adecuados para un nivel o módulo concreto.

En la «valoración mediante escala» se trata de colocar a la persona clasificada según una serie de bandas. El énfasis es, por tanto, vertical: ¿a qué altura de la escala se encuentra? El sentido de las distintas bandas o niveles debería aclararse mediante descriptores de escala. Puede haber varias escalas para distintas categorías, y se pueden presentar en la misma página en forma de "parrilla" o en páginas distintas. Puede haber una definición de cada banda o nivel o de bandas y niveles alternos, o de los niveles superior, medio e inferior.

La alternativa es una lista de control que muestra el camino que se ha recorrido; el énfasis, en este caso, es horizontal: ¿qué cantidad del contenido del módulo se ha realizado con éxito? La lista de control se puede presentar en forma de lista de elementos, igual que un cuestionario; se puede presentar, por otro lado, como si fuera una rueda, o con cualquier otra forma. La respuesta puede ser *sí* o *no*, y se puede diferenciar más con una serie de pasos (por ejemplo: de 0 a 4) preferiblemente identificados con epígrafes y con definiciones que expliquen cómo se deberían interpretar los epígrafes.

Puesto que los descriptores ilustrativos constituyen especificaciones de criterio independientes que han sido ajustadas a los niveles requeridos, se pueden utilizar como una fuente para producir tanto una lista para un nivel concreto —y así ocurre en algunas versiones del *Portfolio*— como escalas o cuadros clasificatorios que abarcan todos los niveles adecuados, según se presentan en el capítulo 3, para la autoevaluación en el cuadro 2 y para la evaluación realizada por el examinador en el cuadro 3.

9.3.10. Impresión / valoración guiada

Impresión: valoración totalmente subjetiva realizada en función de la experiencia de la actuación del alumno en clase, sin hacer referencia a criterios específicos relativos a una evaluación específica.

Valoración guiada: valoración en la que se reduce la subjetividad del examinador al complementar la impresión con una evaluación consciente relativa a criterios específicos.

El término «impresión» se utiliza aquí para referirse al momento en que un profesor o un alumno realiza una valoración simplemente según su experiencia de la actuación en clase, de los trabajos realizados en casa, etc. Muchas formas de valoración subjetiva, sobre todo las utilizadas en la evaluación continua, suponen valorar una impresión sobre la base de la reflexión o de la memoria centradas posiblemente en la observación consciente de una persona determinada durante un periodo de tiempo. Muchos sistemas escolares funcionan sobre esta base.

La expresión «valoración guiada» se utiliza aquí para describir la situación en la que esa impresión se orienta hacia una valoración meditada mediante un enfoque de evaluación. Dicho enfoque supone: (a) una actividad de evaluación que sigue algún tipo de procedimiento, y, en su caso, (b) un conjunto de criterios definidos que distinguen entre las distintas puntuaciones o calificaciones, y (c) algún tipo de formación dirigida a la normalización. La ventaja del enfoque guiado de la valoración es que, si se establece de esta forma un marco común de referencia para el grupo de examinadores, la consistencia de las valoraciones puede aumentar considerablemente. Esto ocurre sobre todo si se proporcionan «puntos de referencia» en forma de muestras de actuación y de vínculos fijos con otros sistemas. La importancia de dichas orientaciones aumenta por el hecho de que la investigación en varias disciplinas ha demostrado repetidas veces que, si no hay un trabajo de formación para unificar criterios de evaluación, las diferencias en el rigor de los examinadores pueden explicar casi tantas de las diferencias que existen en la evaluación de los alumnos como las que explica su capacidad real, dejando los resultados prácticamente al azar.

Las escalas de descriptores de los niveles comunes de referencia se pueden utilizar para proporcionar un conjunto de criterios definidos, como se explicó anteriormente en (b), o para describir los niveles de exigencia representados por criterios existentes en función de los niveles comunes. En el futuro, quizá se puedan proporcionar muestras o ejemplos de actuaciones típicas en distintos niveles comunes de referencia para contribuir a la creación de unos parámetros o criterios estándar.

9.3.11. Global / analítica

La *evaluación global* consiste en realizar una valoración sintética tomada en conjunto. En este caso, el examinador pondera distintos aspectos de forma intuitiva.

La *evaluación analítica* consiste, por su parte, en analizar distintos aspectos de forma separada.

Esta distinción se puede realizar de dos formas: (a) en función de lo que se busca; (b) en función de cómo se alcanza una banda, una calificación o una puntuación. Los sistemas a veces combinan un enfoque analítico en un nivel con un enfoque holístico en otro.

- *Qué se evalúa*: Algunos enfoques evalúan una categoría global como, por ejemplo, «la expresión oral» o «la interacción», asignando una puntuación o una calificación. Otros, más analíticos, exigen que el examinador asigne resultados separados para varios aspectos independientes de la actuación. Sin embargo, en otros enfoques, el examinador tiene que partir de una impresión global, analizar mediante distintas categorías y realizar después una valoración holística meditada. La ventaja que tienen las categorías separadas de un enfoque analítico es que animan al examinador a observar detenidamente; proporcionan, además, un metalenguaje para la negociación entre examinadores y para la retroalimentación que se da a los alumnos. El inconveniente es que existen abundantes testimonios que sugieren que los examinadores no pueden mantener fácilmente separadas las categorías de una valoración holística, y, también, sufren una sobrecarga cognitiva cuando se les presentan más de cuatro o cinco categorías.

- *Cálculo del resultado*: Algunos enfoques equiparan de forma holística la actuación observada con descriptores de una escala de valoración, ya sea la escala holística (de carácter global) o analítica (de tres a seis categorías de una “parrilla”). Dichos enfoques no suponen ninguna aritmética; se informa de los resultados o bien con un solo número o bien con un «número de teléfono» que abarca todas las diferentes categorías. Otros enfoques más analíticos exigen dar una determinada nota para varios elementos distintos y después sumarlas para dar una puntuación, que posteriormente se puede convertir en una calificación. Una de las características de este enfoque es que se ponderan las categorías, es decir, a cada una de las categorías no le corresponde igual número de elementos.

Los cuadros 2 y 3 del capítulo 3 proporcionan ejemplos de la autoevaluación y de la evaluación realizada por el examinador respectivamente de escalas *analíticas* de criterios (es decir, “parrillas”) utilizadas con una estrategia de valoración de carácter *global* (es decir, el examinador correlaciona lo que puede deducir de la actuación con las definiciones y realiza una valoración).

9.3.12. Evaluación en serie / evaluación por categorías

La *evaluación por categorías* supone una sola tarea de evaluación (que puede tener distintas fases para crear discursos diferentes, como ya se vio en la sección 9.2.1) en la que se valora la actuación en relación con las categorías de una “parrilla” de puntuación: el enfoque analítico esbozado en 9.3.11.

La *evaluación en serie* supone una serie de tareas aisladas de evaluación (a menudo juegos de roles con otros alumnos o con el profesor) que se evalúan con una simple calificación global según una escala definida de puntos, por ejemplo: de 0 a 3 o de 1 a 4.

La evaluación en serie intenta corregir la tendencia existente en las evaluaciones por categorías a que los resultados de una categoría afecten a los de otra. En niveles inferiores el énfasis suele recaer en el logro del objetivo de la tarea; la finalidad es completar una lista de control de lo que el alumno sabe hacer sobre la base de la evaluación realizada por el profesor o el alumno de las actuaciones efectivamente realizadas, más que de una simple impresión. En niveles superiores, sin embargo, las tareas pueden estar diseñadas para mostrar aspectos concretos del dominio lingüístico en la actuación. Se informa de los resultados en forma de perfil.

Las escalas de distintas categorías de competencia lingüística yuxtapuestas al texto del capítulo 5 ofrecen una fuente para el desarrollo de los criterios de una evaluación por categorías. Como los examinadores sólo pueden abordar un número pequeño de categorías, hay que llegar a acuerdos en el proceso. La elaboración de tipos de actividades comunicativas de la sección 4.4 y la lista de distintos tipos de competencia funcional recogida en la sección 5.2.3.2 pueden contribuir a la determinación de tareas apropiadas para la evaluación en serie.

9.3.13. Evaluación realizada por otras personas / autoevaluación

Evaluación realizada por otras personas: valoraciones realizadas por el profesor o por el examinador.

Autoevaluación: valoraciones de una persona respecto al dominio lingüístico propio.

Los alumnos pueden participar en muchas de las técnicas de evaluación descritas anteriormente. La investigación sugiere que siempre que no haya riesgos (por ejemplo, si alguien va a ser aceptado para un curso), la autoevaluación puede ser un complemento eficaz de las pruebas y de la evaluación que realiza el profesor. La precisión aumenta en la autoevaluación: a) cuando la evaluación se hace en relación con descriptores claros que definen patrones de dominio lingüístico y, en su caso, b) cuando la evaluación se relaciona con una experiencia concreta. Esta experiencia puede ser en sí misma incluso una actividad de examen; probablemente también se realiza con mayor precisión cuando los alumnos reciben alguna formación específica. Dicha autoevaluación estructurada puede llegar a guardar correlación con las evaluaciones que realizan los profesores y con las pruebas, igualando a la correlación (nivel de validez concurrente) que habitualmente existe entre los profesores mismos, entre las pruebas y entre la evaluación que realiza el profesor y las pruebas.

Sin embargo, el potencial más importante de la autoevaluación está en su utilización como herramienta para la motivación y para la toma de conciencia, pues ayuda a los alumnos a apreciar sus cualidades, a reconocer sus insuficiencias y a orientar su aprendizaje de una forma más eficaz.

Los usuarios del Marco de referencia pueden tener presente y, en su caso, determinar:

- *Qué tipos de evaluación, de los enumerados en este capítulo, son:*
 - *Más adecuados a las necesidades de los alumnos en su sistema.*
 - *Más apropiados y viables en la cultura pedagógica de su sistema.*
 - *Más gratificante para los profesores en cuanto a su implicación profesional a partir de la formación recibida.*
- *La forma en que la evaluación del aprovechamiento (orientada al centro escolar; orientada al aprendizaje) y la evaluación del dominio lingüístico (orientada al mundo real; orientada al resultado) se equilibran y se complementan en su sistema, y hasta qué punto se evalúa tanto la actuación comunicativa como los conocimientos lingüísticos.*
- *Hasta qué punto se evalúan los resultados del aprendizaje en relación con niveles y criterios definidos (referencia a un criterio) y hasta qué punto se asignan las calificaciones y las evaluaciones según la clase en que se encuentra el alumno (referencia a la norma).*
- *Hasta qué punto a los profesores:*
 - *Se les informa de los niveles de exigencia (por ejemplo: descriptores comunes, muestras de actuaciones).*
 - *Se les anima a ser conscientes de la existencia de una serie de técnicas de evaluación.*
 - *Se les forma en técnicas y en interpretación.*
- *Hasta qué punto es deseable y viable desarrollar un enfoque integrado de la evaluación continua del manual y de la evaluación en un momento concreto en relación con niveles relacionados entre sí y definiciones de criterios.*
- *Hasta qué punto es deseable y viable implicar a los alumnos en su autoevaluación referida a descriptores definidos de tareas y a aspectos de dominio en distintos niveles, en la aplicación de esos descriptores en — por ejemplo— la evaluación en serie.*
- *La adecuación a la situación de los alumnos de las especificaciones y de las escalas proporcionadas en el Marco de referencia, y de la forma en que podrían ser complementadas o desarrolladas.*

Las versiones de las escalas de valoración para la autoevaluación y para el examinador se presentan en los cuadros 2 y 3 del capítulo 3. La distinción más llamativa entre las dos -aparte de la formulación en términos de lo que es capaz de hacer el alumno- es que, mientras que el cuadro 2 se centra en actividades comunicativas, el cuadro 3 lo hace en aspectos genéricos de la competencia visibles en cualquier actuación oral. Sin embargo, cabe imaginar sin dificultad una versión adaptada del cuadro 3 para la autoevaluación. La experiencia indica que, al menos, los alumnos adultos son capaces de realizar tales valoraciones cualitativas respecto a su propio nivel de competencia lingüística.

9.4. La evaluación viable y un metasistema

Las escalas intercaladas en los capítulos 4 y 5 presentan un ejemplo de una serie de categorías relacionadas con el esquema descriptivo más integrador que aparece en tales capítulos, y del que se ha extraído de forma simplificada. No se pretende que cualquiera utilice, en un enfoque práctico de evaluación, todas las escalas en todos los niveles; a los examinadores les resulta difícil abordar un gran número de categorías, y además la serie total de niveles presentada puede que no sea adecuada al entorno concreto. Más bien, el conjunto de escalas pretende ser una herramienta de consulta.

Sea cual sea el enfoque adoptado, cualquier sistema práctico de evaluación tiene que reducir el número de categorías posibles a un número viable. La experiencia demuestra que más de cuatro o cinco categorías comienzan a provocar una sobrecarga cognitiva, y que siete categorías es psicológicamente un límite máximo, lo que hay que seleccionar. Respecto a la evaluación oral, si se considera que las estrategias de interacción son un aspecto cualitativo de la comunicación adecuado para la evaluación oral, las escalas ilustrativas contienen catorce categorías cualitativas pertinentes:

Estrategias de turnos de palabra
Estrategias de colaboración
Petición de aclaración
Fluidez
Flexibilidad
Coherencia
Desarrollo temático
Precisión
Competencia sociolingüística
Alcance general
Riqueza de vocabulario
Corrección gramatical
Control del vocabulario
Control fonológico

Es evidente que, aunque los descriptores de muchas de estas características podrían ser incluidos en una lista general, catorce categorías son demasiadas para la evaluación de cualquier actuación. En todo enfoque práctico, por tanto, dicha lista de categorías sería planteada de forma selectiva. Las características tienen que ser combinadas, reformuladas y reducidas a un conjunto

más pequeño de criterios de evaluación que sea adecuado para las necesidades de los alumnos implicados, para los requisitos de la tarea de evaluación concreta y para el estilo de la cultura pedagógica específica. Se podrían ponderar equitativamente los criterios resultantes o, como alternativa, podría darse mayor peso específico a determinados factores que son cruciales para la tarea concreta.

Los siguientes cuatro ejemplos demuestran cómo se puede llevar a cabo todo esto. Los tres primeros ejemplos son breves anotaciones sobre las formas de utilizar las categorías como criterios de pruebas en los enfoques de evaluación existentes; el cuarto ejemplo, por su parte, muestra cómo se unieron y se volvieron a formular los descriptores de las escalas del *Marco de referencia* con el fin de ofrecer una "parrilla" de evaluación para una finalidad concreta en una ocasión determinada.

Ejemplo 1:

Cambridge Certificate in Advanced English (CAE), Parte 5: Criterios para la evaluación (1991).

Criterios para la prueba	Escalas ilustrativas	Otras categorías
Fluidez	Fluidez	
Corrección y alcance	Alcance general Riqueza de vocabulario Corrección gramatical Control de vocabulario	
Pronunciación	Control fonológico	
Aprovechamiento de la tarea	Coherencia Adecuación sociolingüística	Éxito en la tarea Necesidad del apoyo del interlocutor
Comunicación interactiva	Estrategias de turnos de palabra Estrategias de colaboración Desarrollo temático	Alcance y facilidad del mantenimiento de la intervención

Nota sobre otras categorías: En las escalas ilustrativas, las especificaciones sobre el éxito en la tarea están en relación con la clase de actividad implicada bajo el título de *Actividades comunicativas*. El *Alcance y facilidad de la intervención* se incluye en el apartado *Fluidez* de esas escalas. No tuvo éxito el intento de escribir y calibrar los descriptores sobre *Necesidad de apoyo del interlocutor* para incluirlos en el conjunto de escalas ilustrativas.

Ejemplo 2:

International Certificate Conference (ICC): Certificado de inglés para los negocios, Prueba 2: Conversación de negocios (1987).

Criterios de la prueba	Escalas ilustrativas	Otras categorías
Escala 1 (sin denominación)	Adecuación sociolingüística Corrección gramatical Control de vocabulario	Éxito en la tarea
Escala 2 (Uso de las características del discurso para iniciar y mantener la fluidez de la conversación)	Estrategias de turnos de palabra Estrategias de colaboración Adecuación sociolingüística	

Ejemplo 3:

Eurocentres: Evaluación de la interacción en grupo reducido (RADIO) (1987)

Criterios de la prueba	Escalas ilustrativas	Otras categorías
Alcance	Alcance general Riqueza de vocabulario	
Corrección	Corrección gramatical Control de vocabulario Adecuación sociolingüística	
Pronunciación	Fluidez Control fonológico	
Interacción	Estrategias de turnos de palabra Estrategias de colaboración	

Ejemplo 4:

Consejo Nacional Suizo de Investigación: Evaluación de actuaciones en vídeo.

Contexto: En el anejo A se explica cómo fueron clasificados por escalas los descriptores ilustrativos en un proyecto de investigación realizado en Suiza. Al final del proyecto de investigación, los profesores que habían participado fueron invitados a un congreso para presentar los resultados y fomentar la experimentación realizada en Suiza con el *Portfolio europeo de las lenguas (European Language Portfolio)*. En el congreso, dos de los temas de debate fueron (a) la necesidad de relacionar la evaluación continua y las listas de autoevaluación con un marco general, y (b) las distintas formas de utilizar en la evaluación los descriptores clasificados por escalas del proyecto. Como parte de este proceso de análisis, se clasificaron vídeos de algunos de los alumnos investigados con referencia a la “parrilla” de valoración presentada como cuadro 3 en el capítulo 3. En ella se presenta una selección de descriptores ilustrativos refundida y reinventada.

Criterios de la prueba	Escalas ilustrativas	Otras categorías
Alcance	Alcance general Riqueza de vocabulario	
Corrección	Corrección gramatical Control de vocabulario	
Fluidez	Fluidez	
Interacción	Interacción global Turnos de palabra Colaboración	
Coherencia	Coherencia	

Los sistemas distintos con alumnos distintos en contextos distintos simplifican, seleccionan y combinan las características de formas distintas para distintos tipos de evaluación. En realidad, en vez de ser muy larga, la lista de catorce categorías quizá sea incapaz de incorporar *todas* las variantes que eligen las personas y tenga que ser ampliada para poder abarcar la totalidad.

Los usuarios del Marco de referencia pueden tener presente y, en su caso, determinar:

- *La forma en que se simplifican las categorías teóricas para constituir enfoques operativos en su sistema.*
- *Hasta qué punto los principales factores utilizados como criterios de evaluación en su sistema pueden ser situados en la serie de categorías presentadas en el capítulo 5 para el que se ejemplifican escalas en el Anejo B, Documento B4, siempre que se desarrollen más detalladamente para abarcar adecuadamente ámbitos específicos de uso.*