

## 9. Avaluació

### 9.1. Introducció

En aquest capítol, el terme “avaluació” s'utilitza en el sentit de valoració del grau de domini de la llengua que té l'usuari. Totes les proves són una forma d'avaluació, però també hi ha moltes altres maneres d'avaluar (per exemple, les llistes de control utilitzades en l'avaluació contínua; l'observació informal del professor) que no es poden descriure com a proves. Avaluar és un concepte que va més enllà de la simple valoració. Tota valoració és una forma d'avaluació, però en un programa de llengua no tan sols s'avalua el domini de la llengua que té l'aprenent sinó també altres aspectes, com ara l'eficàcia d'alguns materials pedagògics o mètodes concrets, el tipus i la qualitat del discurs produït realment en el programa, la satisfacció de l'aprenent i del docent, l'eficàcia de l'ensenyament, etc. Aquest capítol se centra només en l'avaluació de l'aprenentatge, i no en aspectes més generals de l'avaluació.

Hi ha tres conceptes que tradicionalment s'han considerat fonamentals en qualsevol debat sobre l'avaluació: la validesa, la fiabilitat i la viabilitat. És útil, pel que fa al contingut d'aquest capítol, tenir una idea general del significat d'aquests termes, de com es relacionen entre ells i de quina importància tenen en el *Marc de referència*.

La *validesa* és un dels conceptes clau del *Marc*. Es pot dir que un prova o un procediment d'avaluació té validesa si es pot demostrar que allò que s'avalua realment (el constructe) és el que, en el context en qüestió, s'hauria d'avaluar, i que la informació obtinguda és una representació precisa del domini de la llengua que posseeixen els alumnes o aprenents que s'examinen.

La *fiabilitat* és un terme estadístic que fa referència al grau en què es repeteix el mateix ordre dels candidats pel que fa a les qualificacions obtingudes en dues administracions diferents (reals o simulades) de la mateixa prova d'avaluació.

Però més important que la fiabilitat és la *precisió de les decisions* que s'adopten en relació amb un determinat nivell d'exigència. Si l'informe de l'avaluació dona resultats com ara “aprovat” o “suspens” o nivell A2+, B1, B1+, quina precisió tenen aquestes decisions? La precisió de les decisions depèn de la validesa que tingui el nivell concret (per exemple, el Nivell B1) per al context. També depèn de la validesa dels criteris utilitzats per prendre la decisió i de la validesa dels procediments amb què es van desenvolupar aquests criteris.

Si dos organismes o territoris diferents utilitzen criteris referits als mateixos nivells d'exigència per determinar les seves decisions d'avaluació per a la mateixa habilitat, si els mateixos nivells són vàlids i apropiats per als dos contextos implicats i si s'interpreten de la mateixa manera en el disseny de les tasques d'avaluació i en la interpretació de les actuacions dels aprenents, aleshores els resultats dels dos sistemes estan correlacionats. Tradicionalment, la correlació entre dues proves que se suposa que avaluen el mateix constructe es coneix com a “*validesa concurrent*”. Òbvia-

ment, aquest concepte es relaciona amb la fiabilitat, ja que les proves que no són fiables no es correlacionen. No obstant això, el més important és el grau de coincidència entre les dues proves pel que fa a *allò que s'avalua* i a *com s'interpreta l'actuació*.

El *Marc* s'ocupa d'aquestes dues qüestions. La secció següent descriu les tres maneres principals d'utilitzar el *Marc*.

1. Per especificar el contingut de les proves i dels exàmens: *Què s'avalua?*
2. Per establir els criteris que determinin l'assoliment d'un objectiu d'aprenentatge: *Com s'interpreta l'actuació?*
3. Per descriure els nivells de domini de la llengua en les proves i exàmens existents, de manera que permetin comparar diferents sistemes de qualificacions: *Com es poden fer comparacions?*

Aquestes qüestions fan referència a diferents tipus d'avaluació. Hi ha molts tipus i tradicions diferents d'avaluació. És un error pensar que un enfocament (per exemple, un examen públic) és necessàriament superior en els seus efectes educatius a un altre enfocament (per exemple, l'avaluació que fa el docent). De fet, és un avantatge molt important disposar d'una sèrie de nivells comuns –com els nivells comuns de referència del *Marc*–, que fan possible relacionar les diferents formes d'avaluació.

La tercera secció del capítol presenta possibles opcions entre diferents tipus d'avaluació. Aquestes opcions es presenten en forma de parells oposats. En cada cas, es defineixen els termes que s'utilitzen i se n'examinen els avantatges i inconvenients pel que fa al propòsit de l'avaluació en el seu context educatiu. També s'estableixen les implicacions de portar a la pràctica cadascuna de les opcions alternatives. A continuació, s'indica l'adequació del *Marc* per a cada tipus d'avaluació concreta.

Un procediment d'avaluació també ha de ser pràctic, *viable*. La viabilitat té a veure especialment amb l'avaluació de les actuacions. Els examinadors actuen sota la pressió del temps. Només veuen una mostra limitada de l'actuació dels aprenents i hi ha uns límits clars pel que fa als tipus i nombre de categories que es poden utilitzar com a criteris. El *Marc* pretén proporcionar un punt de referència, no pas una eina pràctica d'avaluació. El *Marc* ha de ser exhaustiu, però tots els seus usuaris han de ser selectius, fet que pot suposar l'ús d'un esquema operatiu més senzill, que integri les categories que el *Marc* presenta de forma separada. Per exemple, les categories utilitzades en els capítols 4 i 5 sovint són més senzilles quan apareixen en les escales de descriptors que no pas quan s'inclouen en el text d'aquests mateixos capítols. La secció final d'aquest capítol examina aquesta qüestió amb exemples.

## 9.2. El Marc com un recurs per a l'avaluació

### 9.2.1. L'especificació del contingut d'exàmens i proves

Quan s'hagi de preparar l'especificació d'una tasca per a una avaluació comunicativa, es pot consultar la descripció de “L'ús de la llengua i l'usuari o aprenent” feta en el capítol 4 i, especialment, a la secció 4.4, que s'ocupa de les activitats comunicatives. Cada cop és més evident que una avaluació vàlida exigeix el mostratge de diversos tipus de discurs. Pel que fa a l'avaluació de l'expressió oral, per exemple, una prova desenvolupada recentment il·lustra aquest punt. En aquesta prova, en primer lloc, hi ha una *conversa* simulada que serveix d'escalfament; a continuació, hi ha una *discussió informal* sobre temes familiars en què l'examinand defensa un punt de vista. Després, es produeix una fase de *transacció* que pren la forma d'una conversa cara a cara o bé la simulació d'una intervenció per telèfon. A continuació, es produeix una fase de *producció*, basada en un *informe* escrit en què l'examinand fa una *descripció* del seu àmbit acadèmic i dels seus projectes. Finalment, hi ha una tasca de *cooperació centrada en un objectiu*, és a dir, una tasca de consens entre els examinands.

En resum, les categories que el Marc utilitza per a les activitats comunicatives són:

	<b>Interacció</b> (espontània, torns de paraula breus)	<b>Producció</b> (preparada, torns de paraula llargs)
<b>Oral</b>	<i>Conversa</i> Discussió informal <i>Cooperació centrada en un objectiu</i>	<i>Descripció</i> de l'àmbit acadèmic de l'aprenent
<b>Escrit</b>		<i>Informe / Descripció</i> de l'àmbit acadèmic de l'aprenent

A l'hora de detallar les especificacions de la tasca, l'usuari pot consultar la secció 4.1, “El context de l'ús de la llengua” (àmbits, condicions i limitacions, context mental); la secció 4.6, que s'ocupa dels “textos”, i el capítol 7, que analitza “el paper de les tasques en l'ensenyament de llengua”, sobretot la secció 7.3, que estudia “la dificultat de la tasca”.

La secció 5.2, que s'ocupa de la “competència lingüística comunicativa”, pot ajudar a redactar els ítems d'una prova, o a elaborar les fases d'una prova oral per fer evidents les competències lingüístiques, sociolingüístiques i pragmàtiques pertinents. El conjunt d'especificacions de continguts del *Threshold Level* (Nivell Llindar), desenvolupades pel Consell d'Europa per a més de 20 llengües europees (vegeu la bibliografia per al capítol 5, a l'apartat de *Bibliografia general*), així com dels nivells *Waystage* (Bàsic) i *Vantage* (Avançat) per a l'anglès, més els seus equivalents en el cas que s'hagin desenvolupat per a altres llengües i nivells, es poden considerar com a auxiliars al document principal del Marc. En aquestes especificacions s'inclouen exemples més detallats que ofereixen orientacions per elaborar proves dels nivells A1, A2, B1 i B2.

## 9.2.2. Els criteris per a l'assoliment d'un objectiu d'aprenentatge

Les escales serveixen de base per al desenvolupament d'escales de valoració que avaluen l'assoliment d'un objectiu d'aprenentatge concret, i els descriptors poden ajudar a formular els criteris. L'objectiu pot ser, d'una banda, un nivell ampli de domini general de la llengua, expressat com un nivell comú de referència (per exemple, B1). Per altra banda, pot ser un conjunt específic d'activitats, habilitats i competències, tal com s'ha esmentat a la secció 6.1.2, que s'ocupa de "les competències parcials i la variació d'objectius en relació amb el *Marc de referència*". Un objectiu modular semblant es podria perfilar en una graella de categories per nivells, com la que s'ha presentat en el Quadre 2.

A l'hora d'examinar l'ús dels descriptors és essencial distingir entre:

1. Els descriptors d'activitats comunicatives, que es troben en el capítol 4.
2. Els descriptors d'aspectes del domini de la llengua relacionats amb competències concretes, que es troben en el capítol 5.

Els primers descriptors són molt útils per a l'avaluació que fa el professor o per a l'autoavaluació respecte de les tasques de la vida real. Aquests dos tipus d'avaluació es fan sobre la base d'una representació detallada de la capacitat de l'aprenent de llengua, desenvolupada durant el curs en qüestió. Són interessants perquè poden ajudar els aprenents i els professors a adoptar un enfocament centrat en l'acció.

No obstant això, no és gaire aconsellable incloure descriptors d'activitats comunicatives en els criteris que utilitza un examinador per valorar l'actuació en una prova oral o escrita determinada si el que interessa és explicar els resultats pel que fa al nivell de domini de la llengua assolit. Això és així perquè, per informar sobre el domini de la llengua, l'avaluació no ha de basar-se únicament en una actuació determinada, sinó que hauria d'intentar valorar les competències generals que posa en evidència aquesta actuació. Evidentment, hi pot haver raons sòlides de caràcter educatiu per concentrar-se a assolir l'èxit en la realització d'una activitat determinada, especialment entre els usuaris bàsics més joves (nivells A1 i A2). Aquests resultats seran menys generalitzables, però la generalització dels resultats no sol ser el centre d'atenció en les primeres etapes de l'aprenentatge de la llengua.

Això reforça el fet que les avaluacions poden tenir moltes funcions diferents. El que és adequat per a una finalitat d'avaluació pot ser inadequat per a una altra.

### 9.2.2.1. Descriptors d'activitats comunicatives

Els descriptors de les activitats comunicatives (capítol 4) es poden utilitzar de tres maneres diferents pel que fa a l'assoliment d'objectius.

1. **Elaboració de tasques.** Tal com s'ha comentat en la secció 9.2.1, les escales de les activitats comunicatives ajuden a definir una especificació per al disseny de les tasques d'avaluació.

2. **Informació sobre resultats.** Les escales de les activitats comunicatives també poden ser molt útils per informar dels resultats. A les persones interessades en els resultats del sistema educatiu, com ara els empresaris, sovint els interessen més els resultats globals que no pas un perfil detallat de competència.
3. **Avaluació que fa el professor i autoavaluació.** Finalment, els descriptors de les activitats comunicatives es poden utilitzar de diverses maneres tant per a l'avaluació que fa el professor com per a l'autoavaluació:
  - **Llista de control:** Es pot utilitzar per a l'avaluació contínua o per a l'avaluació sumatòria al final d'un curs. Els descriptors d'un nivell concret es poden recollir en forma de llista. Com a alternativa, el contingut dels descriptors es pot "expandir". Per exemple, el descriptor: *Puc demanar i proporcionar informació personal* es podria expandir en les parts implícites constituents *puc presentar-me; puc dir on visc; puc donar la meua adreça en català; puc dir l'edat que tinc*, etc. i *puc preguntar a algú com es diu, i on viu, quants anys té*, etc.
  - **Graella:** Es pot utilitzar per a l'avaluació contínua o sumatòria. Es construeix un perfil seleccionant en una graella unes categories (per exemple, *Conversa; Discussió; Intercanvi d'informació*) definides en diferents nivells (B1+, B2, B2+).

Aquest ús dels descriptors ha esdevingut molt habitual en els últims 10 anys. L'experiència ha demostrat que la consistència amb què professors i aprenents poden interpretar els descriptors augmenta si els descriptors descriuen no només ALLÒ que l'aprenent pot fer, sinó també COM ho fa.

### 9.2.2.2. Descriptors dels aspectes del domini de la llengua referits a competències concretes

Els descriptors dels aspectes del domini de la llengua es poden utilitzar de dues maneres pel que fa a l'assoliment d'objectius.

1. **L'autoavaluació i l'avaluació que fa el professor.** Sempre que els descriptors estiguin formulats "en positiu" i "de forma independent", es poden incloure en llistes de control per a l'autoavaluació o per a l'avaluació que fa el professor. No obstant això, la majoria de les escales existents tenen l'inconvenient que els descriptors sovint s'expressen negativament en els nivells inferiors i amb referència a la norma, situada cap a la meitat de l'escala. Sovint també fan distincions purament verbals entre nivells només canviant una o dues paraules en les descripcions adjacents, i això fa que tinguin poc sentit fora del context de l'escala. A l'annex A es descriuen algunes maneres de desenvolupar descriptors que eviten aquests problemes.
2. **L'avaluació de l'actuació.** L'ús més evident de les escales de descriptors del capítol 5 que tracten aspectes relacionats amb la competència és oferir punts de partida per al desenvolupament de criteris d'avaluació. Si les impressions personals i no sistemàtiques s'articulen com a judicis, aquests descriptors poden contribuir a desenvolupar un marc de referència compartit entre el grup d'examinadors implicats.

Bàsicament hi ha tres maneres de presentar els descriptors perquè es puguin utilitzar com a criteris d'avaluació:

- En primer lloc, els descriptors es poden presentar en una *escala* –sovint combinant descriptors de categories diferents en un paràgraf global per cada nivell. Aquest enfocament és molt habitual.
- En segon lloc, es poden presentar com una *llista de control*, normalment una llista per nivell, sovint amb descriptors agrupats sota títols, és a dir, sota categories. Les llistes de control són menys habituals en l'avaluació directa.
- En tercer lloc, es poden presentar en una *graella* de categories seleccionades com un conjunt d'escala paral·leles per a categories diferents. Aquest enfocament fa possible donar un perfil de diagnòstic. No obstant això, hi ha límits en el nombre de categories amb què els examinadors poden treballar.

Hi ha dues maneres diferents de proporcionar una graella de subescales:

*Escala de domini:* Es proporciona una graella de perfil que defineix els nivells pertinents per a certes categories; per exemple, des del nivell A2 fins al B2. Aleshores l'avaluació en aquests nivells es fa directament amb més precisió mitjançant un segon dígit o signes + per aconseguir una diferenciació més gran si es vol. D'aquesta manera, malgrat que la prova d'actuació se situés al nivell B1, i malgrat que cap dels aprenents arribés al nivell B2, encara seria possible que els aprenents més preparats assolissin el nivell B1+, B1++ o B1.8.

*Escala de puntuació d'una prova.* Per a cada categoria pertinent se selecciona o defineix un descriptor que descriu el nivell d'exigència requerit per aprovar un mòdul o un examen concret d'aquella categoria. A aquest descriptor se l'anomena aleshores "Aprovat" o "3", i l'escala s'estableix respecte d'aquell nivell d'exigència (una actuació molt fluixa = "1", una actuació excel·lent = "5"). La formulació d' "1" i "5" podrien constituir-la altres descriptors extrets o adaptats dels nivells adjacents de l'escala que apareix a la secció apropiada del capítol 5; el descriptor també es podria formular en relació amb la redacció del descriptor definit com a "3".

### 9.2.3. Descripció dels nivells de domini de la llengua per facilitar la comparació de proves i exàmens

Les escales dels nivells comuns de referència intenten facilitar la descripció del nivell de domini de la llengua representat per mitjà de diplomes i certificats existents, i d'aquesta manera permetre la comparació entre sistemes. En la bibliografia sobre *mesura* es reconeixen cinc maneres clàssiques de relacionar avaluacions diferents: (1) l'equiparació, (2) el calibratge, (3) la moderació estadística, (4) els referents externs comunament acceptats i (5) la moderació social.

Els tres primers mètodes són tradicionals: (1) produir versions alternatives de la mateixa prova (equiparació), (2) relacionar els resultats de diferents proves en una escala comuna (calibratge) i (3) corregir la dificultat de les proves o l'exigència dels examinadors (moderació estadística).

Els dos darrers mètodes suposen aconseguir un consens comú a través de la discussió (moderació social) i de la comparació de mostres de treball en relació amb definicions i exemples estandarditzats (referents externs comunament acceptats). Una de les finalitats del *Marc* és donar suport a aquest procés per aconseguir un consens. Per aquesta raó, les escales de descriptors que s'utilitzen per a aquest propòsit s'han estandarditzat amb una metodologia de desenvolupament rigorosa. En l'ensenyament, aquest enfocament es descriu cada vegada més com una avaluació centrada en estàndards. Generalment es reconeix que el desenvolupament d'un enfocament centrat en estàndards demana temps, ja que els usuaris adquireixen la consciència del significat dels estàndards a través d'un procés d'exemplificació i d'intercanvi d'opinions.

Es pot sostenir que aquest enfocament és potencialment el mètode més sòlid de relació perquè suposa el desenvolupament i la validació d'una visió comuna del constructe. La raó fonamental per què és difícil relacionar les avaluacions de llengua, malgrat els malabarismes estadístics de les tècniques tradicionals, és que les avaluacions generalment avaluen elements radicalment diferents fins i tot quan pretenen cobrir els mateixos àmbits. Això és degut en part a (*a*) una conceptualització i plasmació operativa del constructe insuficients, i en part a (*b*) una interferència derivada del mètode d'avaluació.

El *Marc* ofereix un intent raonat i fonamentat de proporcionar la solució al primer problema en relació amb l'aprenentatge de llengües modernes en un context europeu. Del capítol 4 al 7 s'argumenta un esquema descriptiu que intenta conceptualitzar l'ús de la llengua, les competències, i els processos d'ensenyament i d'aprenentatge d'una manera pràctica que ajudi els aprenents a fer funcionar la capacitat lingüística comunicativa que volem promoure.

Les escales de descriptors constitueixen una graella conceptual que es pot utilitzar per:

- a*) Relacionar sistemes nacionals i institucionals entre si, a través del *Marc*.
- b*) Estructurar els objectius d'exàmens concrets i de cursos utilitzant les categories i els nivells de les escales.

L'annex A proporciona als lectors una visió global dels mètodes utilitzats per desenvolupar escales de descriptors, i relacionar-les amb l'escala del *Marc*.

La Guia de l'usuari per a examinadors, elaborada sota la direcció de M. Milanovic (ALTE), proporciona consells detallats sobre la plasmació operativa dels constructes de les proves i sobre com evitar distorsions innecessàries derivades dels efectes del mètode d'avaluació aplicat.

### 9.3. Tipus d'avaluació

Es poden fer diverses distincions importants pel que fa a l'avaluació. La llista següent no és exhaustiva ni és important el fet que un terme de la distinció aparegui a l'esquerra o a la dreta.

<b>1</b>	Avaluació de l'assoliment d'objectius	Avaluació del domini de la llengua
<b>2</b>	Amb referència a una norma (RN)	Amb referència a un criteri (RC)
<b>3</b>	Domini RC	Línia contínua RC
<b>4</b>	Avaluació contínua	Avaluació puntual
<b>5</b>	Avaluació formativa	Avaluació sumatòria
<b>6</b>	Avaluació directa	Avaluació indirecta
<b>7</b>	Avaluació de l'actuació	Avaluació dels coneixements
<b>8</b>	Avaluació subjectiva	Avaluació objectiva
<b>9</b>	Valoració mitjançant una llista de control	Valoració mitjançant una escala
<b>10</b>	Decisions basades en impressions	Decisions basades en criteris
<b>11</b>	Avaluació holística o global	Avaluació analítica
<b>12</b>	Avaluació per sèrie	Avaluació per categories
<b>13</b>	Avaluació feta per altres persones	Autoavaluació

**Quadre 7**

### 9.3.1. Avaluació de l'assoliment d'objectius / Avaluació del domini de la llengua

En l'*avaluació de l'assoliment d'objectius* el que s'avalua són objectius específics, és a dir, allò que s'ha ensenyat. Per tant, es relaciona amb el treball setmanal o mensual, amb el llibre de text, amb el programa del curs. L'avaluació de l'assoliment d'objectius se centra en el curs. Representa una perspectiva interna.

L'*avaluació del domini de la llengua*, d'altra banda, és l'avaluació d'allò que algú sap fer o pot fer realment, és a dir, del grau en què pot aplicar en el món real allò que ha après. Representa una perspectiva externa.

Els professors tenen una tendència natural a interessar-se més per l'avaluació de l'assoliment d'objectius, per tal d'obtenir realimentació per a la docència. Als responsables de recursos humans de les empreses, als administradors educatius i als aprenents adults, en canvi, els sol interessar més l'avaluació del domini de la llengua, és a dir, l'avaluació dels resultats, d'allò que una persona pot fer realment. L'avantatge de l'avaluació de l'assoliment d'objectius rau en el fet que és propera a l'experiència de l'aprenent. L'avantatge de l'avaluació del domini de la llengua és que permet a tots els implicats saber on es troba l'aprenent: els resultats són transparents.

En les proves comunicatives, en un context d'ensenyament i d'aprenentatge centrat en les necessitats, es pot argumentar que la diferència entre l'assoliment d'ob-



jectius (centrat en el contingut del curs) i el domini de la llengua (centrat en la capacitat de desenvolupar-se en situacions reals) hauria de ser, teòricament, ínfima. En la mesura en què l'avaluació de l'assoliment d'objectius avalua la utilització pràctica de la llengua en situacions pertinents i pretén oferir una imatge equilibrada de la competència emergent, té un component de domini. En la mesura en què l'avaluació del domini consisteix en tasques lingüístiques i comunicatives basades en un programa conegut i adequat, i dóna a l'aprenent l'oportunitat de demostrar què ha aconseguit, té un component d'assoliment d'objectius.

Les escales de descriptors es relacionen amb l'avaluació del domini de la llengua: la capacitat de desenvolupar-se en situacions de la vida real. Al capítol 6 s'examina la importància de les proves d'assoliment d'objectius com un reforç per a l'aprenentatge.

### 9.3.2. Amb referència a una norma / Amb referència a un criteri

*L'avaluació amb referència a una norma* és la classificació dels aprenents d'acord amb un ordre establert a partir dels resultats de l'avaluació. La posició de cada aprenent s'estableix respecte dels altres aprenents.

*L'avaluació amb referència a un criteri* és una reacció contra la referència a la norma: s'avalua l'aprenent únicament segons les seves pròpies capacitats en la matèria, independentment de les capacitats dels altres aprenents.

La referència a la norma es pot entendre en relació amb la classe (ets el divuitè) o en el marc d'una població determinada (ets el 21.567; estàs situat a la banda superior del 14%) o en relació amb el grup d'aprenents que fan la prova. En aquest darrer cas, es poden ajustar les puntuacions de les proves per arribar a un resultat "just" comparant el traçat de la corba de distribució dels resultats de la prova amb la corba dels anys precedents a fi de mantenir un cert nivell i assegurar que el mateix percentatge d'aprenents aconsegueix tots els anys qualificacions d'"excel·lent", independentment de la dificultat de la prova o de la capacitat dels alumnes. L'avaluació amb referència a una norma s'utilitza normalment en les proves de nivell que serveixen per formar els grups d'aprenents.

L'avaluació amb referència a un criteri es pot representar amb una gràfica en l'eix vertical de la qual se situaria el grau de domini de la llengua (com en una línia contínua) i en l'eix horitzontal de la qual es recolliria un repertori d'àmbits pertinents de manera que els resultats individuals de la prova se situarien en relació amb el mapa total de criteris. Això comporta (a) la definició dels àmbits pertinents que cobreix la prova concreta i (b) la identificació de les "puntuacions de tall", és a dir, la puntuació o les puntuacions de la prova que es consideren necessàries per assolir el nivell de domini de la llengua que estableix la prova.

En l'esquema descriptiu les escales de descriptors es componen d'enunciats de criteri per a categories. Els nivells comuns de referència presenten un conjunt d'estàndards comuns.

### 9.3.3. Domini RC / Línia contínua RC

L'enfocament *amb referència al criteri de domini* estableix un únic “nivell mínim de competència” o “puntuació de tall” per dividir els aprenents entre els que han assolit el domini i els que no, sense atribuir graus de qualitat en l'assoliment de l'objectiu establert.

L'enfocament *amb referència al criteri de la línia contínua* és un enfocament en què una capacitat individual es classifica amb referència a una línia contínua que conté tots els graus de capacitat pertinents per avaluar una àrea determinada.

De fet, existeixen nombrosos enfocaments de l'avaluació amb referència a un criteri, però la majoria es poden definir principalment des d'una interpretació de *domini* o de *línia contínua*. La confusió prové en gran mesura del fet que s'identifica erròniament l'avaluació amb referència a un criteri exclusivament amb l'enfocament de domini. L'enfocament de domini és un enfocament que té en compte l'assoliment d'un curs o d'un mòdul. Dóna menys importància al fet de situar aquest mòdul (i, per tant, l'assoliment d'aquest mòdul) en la línia contínua del domini de la llengua.

L'alternativa a l'enfocament de domini consisteix a relacionar els resultats de cada prova amb el grau corresponent de domini de la llengua i fer-ho, habitualment, mitjançant una sèrie de qualificacions. En aquest enfocament la línia contínua és el “criteri”, és a dir, la realitat exterior que garanteix que els resultats de la prova tenen sentit. La referència a aquest criteri exterior es pot fer amb una anàlisi estadística (per exemple, el model de Rasch) que permeti relacionar els resultats de totes les proves entre ells i traslladar-los directament a una escala comuna.

El *Marc* es pot utilitzar amb un enfocament de domini o amb un enfocament de línia contínua. L'escala de nivells utilitzada en l'enfocament de línia contínua es pot equiparar als nivells comuns de referència; l'objectiu que s'ha d'assolir en l'enfocament de domini es pot representar en una graella conceptual que contingui les categories i els nivells del *Marc*.

### 9.3.4. Avaluació contínua / Avaluació puntual

L'*avaluació contínua* és l'avaluació que fa el professor i, possiblement, l'aprenent, de les actuacions a classe, els treballs i els projectes realitzats durant el curs. La nota final reflecteix, així, tot el curs, l'any o el semestre.

L'*avaluació puntual* consisteix en el fet d'atorgar notes o prendre decisions sobre la base d'un examen o d'un altre procediment d'avaluació que té lloc en una data determinada, generalment al final o a l'inici d'un curs. El que ha passat abans és irrelevant; només compta el que l'aprenent és capaç de fer aquí i ara.

Sovint es considera l'avaluació com un fet extern al curs, que té lloc en moments precisos i que té la finalitat de prendre decisions. L'avaluació contínua és una avaluació que s'integra en el curs i que contribueix, de manera acumulativa, a l'avaluació de final de curs. A banda de la correcció dels deures i de les proves d'assoliment per

reforçar l'aprenentatge, que poden ser ocasionals o regulars, l'avaluació contínua es pot fer de diverses maneres: mitjançant llistes de control o gralles que emplen els docents i/o els aprenents, mitjançant l'avaluació d'una sèrie de tasques específiques, mitjançant l'avaluació formal del treball del curs i mitjançant el recurs a un recull de mostres de treballs en fases diferents de la seva realització i/o en diferents moments del curs.

Tots dos enfocaments tenen avantatges i inconvenients. L'avaluació puntual assegura que les persones encara siguin capaces de fer coses que podrien haver après dos anys abans, però provoca "ansietat" en la realització dels exàmens i afavoreix certs tipus d'aprenents. L'avaluació contínua permet tenir més en compte la creativitat i les aptituds diferents, però depèn molt de la capacitat que tingui el professor de ser objectiu. En el pitjor dels casos, pot transformar la vida de l'aprenent en una prova permanent i esdevenir un malson burocràtic per al docent.

Les llistes d'especificacions de criteris que descriuen la capacitat respecte a les activitats comunicatives (capítol 4) poden ser útils per a l'avaluació contínua. D'altra banda, les escales de puntuació desenvolupades a partir dels descriptors d'aspectes de la competència (capítol 5) es poden utilitzar per atorgar notes en l'avaluació puntual.

### 9.3.5. Avaluació formativa / Avaluació sumatòria

L'*avaluació formativa* és un procés continu que permet recollir informació sobre els punts forts i els punts febles de l'aprenentatge; una informació que el professor pot utilitzar en la planificació del curs i transmetre als aprenents. L'avaluació formativa sovint s'utilitza en un sentit molt ampli a fi d'incloure-hi la informació no quantificable procedent de qüestionaris i consultes.

L'*avaluació sumatòria* resumeix l'assoliment al final del curs i li atribueix una nota. No es tracta forçosament d'una avaluació de domini de la llengua. De fet, l'avaluació sumatòria és sovint una avaluació puntual amb referència a una norma i avalua l'assoliment.

La qualitat que té l'avaluació formativa és que pretén millorar l'aprenentatge. La seva feblesa és inherent a la metàfora de la realimentació. La realimentació només funciona si el destinatari està en situació (a) *de poder reconèixer-la*, és a dir, està atent, motivat i acostumat a la forma en què arriba la informació; (b) *de poder rebre-la*, és a dir, no l'ofega la informació i té possibilitat de registrar-la, organitzar-la i personalitzar-la; (c) *de poder interpretar-la*, és a dir, té coneixements previs suficients i és prou conscient per comprendre quin és el punt en qüestió i no dur a terme accions contraproduents, i (d) *de poder integrar* la informació, és a dir, té el temps, l'orientació i els recursos adequats per reflexionar, integrar i, d'aquesta manera, recordar la informació nova. Això suposa una certa capacitat d'aprendre a ser autònom, de controlar el propi aprenentatge i de saber actuar segons la realimentació rebuda.

Aquesta formació de l'aprenent o increment de la conscienciació s'ha anomenat *avaluació formativa*. Es poden utilitzar diverses tècniques per formar aquesta cons-

cienciació. Un principi bàsic consisteix a comparar la impressió (és a dir, el que l'aprenent diu que pot fer en una llista de control) amb la realitat (per exemple, escoltar realment el material del tipus esmentat que apareix a la llista de control i comprovar si el comprèn). El projecte DIALANG relaciona d'aquesta manera l'autoavaluació amb l'actuació en una prova. Una altra tècnica important consisteix a examinar mostres de treball –tant exemples neutres com mostres procedents dels aprenents– i fomentar en els aprenents el desenvolupament d'un metallenguatge propi sobre aspectes relacionats amb la qualitat. Aleshores podran utilitzar aquest metallenguatge per controlar el seu treball, cercar-hi els punts forts i els punts febles i formular un pla d'aprenentatge de manera autònoma.

La majoria de les avaluacions formatives o de diagnòstic funcionen en un nivell molt detallat d'aspectes lingüístics o habilitats que s'han ensenyat recentment o que s'ensenyaran aviat. Per a l'avaluació de diagnòstic, les llistes d'exponents presentades a la secció 5.2 encara són massa generals perquè siguin pràctiques; caldria referir-se a una especificació concreta que fos pertinent (Nivell Bàsic, Nivell Llindar, etc.). No obstant això, les escales de descriptors que defineixen aspectes diferents de la competència en nivells diferents (capítol 4) poden ser útils per oferir realimentació formativa procedent de l'avaluació de l'expressió oral.

Sembla que els nivells comuns de referència siguin del tot pertinents per a l'avaluació sumatòria. Malgrat això, tal com demostra el projecte DIALANG, fins i tot la realimentació procedent d'una avaluació sumatòria pot ser de diagnòstic i, per tant, formativa.

### 9.3.6. Avaluació directa / Avaluació indirecta

L'*avaluació directa* avalua allò que el candidat està fent realment. Per exemple, un grup reduït discuteix sobre algun tema i l'examinador observa, compara amb una graella de criteris, confronta les actuacions dels aprenents amb les categories més adequades de la graella i valora.

L'*avaluació indirecta*, d'altra banda, utilitza una prova, generalment escrita, que sovint avalua les capacitats de l'aprenent.

L'avaluació directa es limita, de fet, a l'expressió oral i escrita, i a la comprensió de la interacció oral, ja que és impossible observar directament les activitats de recepció. La comprensió lectora, per exemple, només es pot avaluar indirectament demanant als aprenents que demostrin la comprensió omplint caselles, completant frases, responent preguntes, etc. El repertori i el control lingüístics es poden avaluar o bé directament, valorant si encaixen amb els criteris que es defineixen, o bé indirectament, interpretant i generalitzant les respostes a les preguntes d'un examen. L'entrevista és una prova directa clàssica; mentre que un text amb buits és una prova indirecta clàssica.

Els descriptors del capítol 5, que defineixen aspectes de la competència en nivells diferents, es poden utilitzar per desenvolupar criteris d'avaluació en proves directes.

Els paràmetres del capítol 4 poden servir per a la selecció de temes, textos i tasques a realitzar per a proves directes de les habilitats productives i per a proves indirectes de comprensió oral i lectora. Els paràmetres del capítol 5 poden permetre, a més, identificar les competències lingüístiques clau que es poden incloure en una prova indirecta de coneixements lingüístics i les competències clau de caràcter pragmàtic, sociolingüístic i lingüístic per formular les preguntes que poden aparèixer en proves de les quatre habilitats basades en ítems.

### 9.3.7. Avaluació de l'actuació / Avaluació dels coneixements

*L'avaluació de l'actuació* exigeix que l'aprenent produeixi una mostra de discurs oral o escrit a través d'una prova directa.

*L'avaluació dels coneixements* exigeix que l'aprenent respongui preguntes que poden ser d'una sèrie de diferents tipus d'ítem per fer evident l'abast dels seus coneixements i el seu grau de domini de la llengua.

Malauradament, les competències no es poden avaluar mai directament. Hem de basar-nos en l'avaluació d'una sèrie d'actuacions a partir de les quals s'intenta generalitzar sobre el domini de la llengua. El domini de la llengua es pot considerar com la posada en acció de la competència. En aquest sentit, per tant, totes les proves avaluen només l'actuació, encara que, a partir de l'evidència aportada, es pot intentar deduir-ne les competències subjacents.

De tota manera, una entrevista exigeix més "actuació" que omplir els buits d'unes frases, i omplir buits, per la seva banda, exigeix més "actuació" que triar entre diverses opcions. Aquí la paraula "actuació" s'utilitza en el sentit de producció de llengua. Però la paraula "actuació" s'utilitza en un sentit més restringit en l'expressió "proves d'actuació". En aquest cas la paraula es refereix a una actuació pertinent en una situació (relativament) real i sovint relacionada amb la feina o els estudis. Utilitzant aquest terme "avaluació de l'actuació" d'una manera més imprecisa es podria dir que els procediments d'avaluació oral són proves d'actuació en el sentit que fan generalitzacions sobre el domini de la llengua a partir d'actuacions en una sèrie d'estils discursius que es consideren adequats al context d'aprenentatge i a les necessitats dels aprenents. Algunes proves equilibren l'avaluació de l'actuació amb una avaluació del coneixement de la llengua com a sistema; d'altres no ho fan.

Aquesta diferència és molt semblant a la que hi ha entre proves directes i indirectes. El *Marc* es pot utilitzar d'una manera similar. Les especificacions del Consell d'Europa per a nivells diferents (Nivell Bàsic, Nivell Llindar, Nivell Avançat) ofereixen a més, en les llengües en què estan disponibles, una descripció detallada i adequada del coneixement de la llengua objecte d'estudi.

### 9.3.8. Avaluació subjectiva / Avaluació objectiva

*L'avaluació subjectiva* és una valoració que fa un examinador i això significa normalment valorar la qualitat d'una actuació.

En l'avaluació *objectiva* s'elimina la subjectivitat. Aquest tipus d'avaluació se sol materialitzar en una prova indirecta en què els ítems només tenen una resposta correcta, com ocorre, per exemple, en les proves amb exercicis d'elecció múltiple.

Tanmateix, la qüestió de la subjectivitat i l'objectivitat és força més complexa.

La prova indirecta se sol descriure com a "prova objectiva" quan el corrector consulta una clau definitiva per decidir si ha d'acceptar o rebutjar les respostes, i a continuació fa el recompte de les correctes per donar el resultat de la prova. Alguns tipus de proves van una mica més enllà en no permetre més d'una resposta correcta per a cada pregunta (per exemple, les proves amb exercicis d'elecció múltiple i les proves de tipus *c*, que es van desenvolupar a partir dels exercicis de tipus *cloze* per aquesta raó), i sovint s'adopta la correcció automàtica a fi d'eliminar qualsevol error del corrector. De fet, l'objectivitat de les proves descrites com a "objectives" és sobrevalorada, ja que ha estat una persona qui ha decidit limitar l'avaluació a tècniques que ofereixen més control sobre la situació de la prova (una decisió subjectiva amb la qual no tothom ha d'estar necessàriament d'acord). Després algú ha fet l'especificació de la prova i algú altre pot haver redactat un ítem amb l'intent de fer operatiu un punt concret de l'especificació. Finalment, algú ha seleccionat un ítem entre tots els ítems possibles per a una prova en concret. Ja que totes aquestes decisions inclouen un element de subjectivitat, val més descriure aquestes proves com a proves puntuades objectivament.

En l'avaluació directa de l'actuació les notes s'atorguen generalment segons un judici. Això significa que la decisió sobre la qualitat de l'actuació de l'aprenent es pren subjectivament tenint en compte els factors que es consideren importants i amb referència a unes directrius o criteris, o a l'experiència del corrector. Com que la llengua i la comunicació són molt complexes, difícils d'atomitzar i no són només el resultat de la suma de les seves parts, utilitzar un enfocament subjectiu té avantatges. Sovint és molt difícil establir què avalua realment un ítem d'una prova. Per tant, fer correspondre els ítems d'una prova amb aspectes específics de competència o actuació és molt menys senzill del que sembla.

Malgrat tot, i per ser justos, cal dir que qualsevol avaluació hauria de ser al més objectiva possible. Els efectes de les valoracions personals que intervenen en les decisions subjectives sobre la selecció del contingut i la qualitat de l'actuació s'haurien de reduir tant com fos possible, especialment quan es tracta d'una avaluació sumatòria, perquè els resultats de la prova els utilitzen molt sovint terceres persones per prendre decisions sobre el futur dels aprenents que han estat avaluats.

La subjectivitat de l'avaluació es pot reduir, i d'aquesta manera augmentar-ne la validesa i la fiabilitat, seguint els passos següents:

- Desenvolupar una ***especificació del contingut*** de l'avaluació basada, per exemple, en un marc de referència comú en el context en qüestió.
- Utilitzar *valoracions compartides* per seleccionar el contingut i/o per puntuar les actuacions.

- Adoptar **procediments estandarditzats** per portar a terme l'avaluació.
- Proporcionar **claus de respostes definitives** per a proves indirectes i basar les valoracions de les proves directes en **criteris específics definits**.
- Exigir **valoracions múltiples** i/o la **ponderació de factors diferents**.
- Proporcionar una **formació** adequada en relació amb les **directrius d'avaluació**.
- Comprovar la qualitat de l'avaluació (validesa, fiabilitat) amb l'**anàlisi de les dades d'avaluació**.

Tal com s'ha dit a l'inici d'aquest capítol, el primer pas per reduir la subjectivitat de les valoracions fetes en totes les etapes del procés d'avaluació és desenvolupar una comprensió comuna del constructe implicat, un marc comú de referència. El *Marc* pretén oferir la base per a l'**especificació del contingut** i ser un recurs per al desenvolupament de **criteris específics definits** per a les proves directes.

### 9.3.9. Puntuació d'acord amb una escala / Puntuació d'acord amb una llista de control

En la *puntuació d'acord amb una escala* es valora si una persona es troba en una franja o nivell concret d'una escala constituïda per diversos nivells o franges.

En la *puntuació d'acord amb una llista de control* es valora una persona segons una llista d'elements que s'han considerat pertinents per a un nivell o un mòdul concret.

En la “puntuació d'acord amb una escala” es posa l'accent en el fet de situar la puntuació d'una persona en una sèrie de franges o nivells. S'accentua la dimensió vertical: a quina alçada de l'escala ha arribat? Els descriptors de nivells han d'aclarir el significat dels diferents nivells o franges. Hi pot haver diverses escales per a diferents categories i poden aparèixer en forma de graella en una mateixa pàgina o en pàgines diferents. Hi pot haver una definició de cada nivell o franja, o de nivells i franges alterns, o dels nivells superior, inferior i mitjà.

L'altra possibilitat és una llista de control que mostra el camí recorregut, és a dir, té una dimensió horitzontal: quina proporció del contingut del mòdul s'ha completat amb èxit? La llista de control es pot presentar com una llista d'elements, com un qüestionari. També es pot presentar com si fos una roda o d'alguna altra manera. La resposta pot ser *sí* o *no*. La resposta es pot diferenciar més amb una sèrie d'esglaons (per exemple, de 0 a 4) preferiblement identificats amb títols i amb definicions que expliquin com s'haurien d'interpretar aquests títols.

Atès que els descriptors constitueixen enunciats de criteri independents que han estat calibrats als nivells pertinents, es poden utilitzar com una font per produir tant una llista per a un nivell concret –tal com ocorre en algunes versions del Carnet de Llengües– com escales de puntuació o gralles que cobreixin tots els nivells rellevants, tal com es presenten en el capítol 3, per a l'autoavaluació en el Quadre 2, i per a l'avaluació per part de l'examinador en el Quadre 3.

### 9.3.10. Valoració basada en impressions / Valoració basada en criteris

*Valoració basada en impressions:* valoració completament subjectiva basada en l'observació de l'actuació de l'aprenent a classe sense cap referència a criteris específics relatius a una avaluació específica.

*Valoració basada en criteris:* valoració en què es redueix la subjectivitat pròpia de l'examinador complementant la simple impressió amb una avaluació conscient en relació amb uns criteris específics.

El terme "impressió" s'utilitza aquí per referir-se al que passa quan un docent o un aprenent puntuen simplement sobre la base de la seva experiència de l'actuació a classe, dels deures fets a casa, etc. Moltes formes de puntuació subjectiva, especialment les utilitzades en l'avaluació contínua, suposen valorar una impressió sobre la base de la reflexió o la memòria centrades possiblement en l'observació conscient de la persona implicada durant un cert període de temps. Molts sistemes escolars funcionen sobre aquesta base.

L'expressió "valoració basada en criteris" descriu la situació en què s'utilitza un enfocament d'avaluació que transforma una impressió en un judici raonat. Un enfocament d'aquest tipus suposa *a)* una activitat d'avaluació amb algun tipus de procediment i/o *b)* un conjunt de criteris definits que distingeixen entre les diferents notes o resultats, i *c)* algun tipus de formació dels examinadors en aplicació estàndard de criteris. L'avantatge d'un enfocament com el de la "valoració basada en criteris" és que si s'estableix un marc comú de referència per al grup d'examinadors implicats, la consistència de les decisions pot augmentar considerablement. Això ocorre especialment si es proporcionen com a referents mostres d'actuació i s'estableixen vincles amb altres sistemes. Aquestes orientacions són importants perquè, tal com ha demostrat repetidament la recerca en diverses disciplines, si no es disposa d'uns criteris que serveixin de guia, les diferències de rigor dels examinadors poden explicar gairebé tantes diferències en l'avaluació dels aprenents com les que explica la seva capacitat real, deixant els resultats pràcticament a l'atzar.

Les escales de descriptors dels nivells comuns de referència es poden expandir per obtenir un grup de criteris definits, tal com s'ha dit anteriorment en l'apartat *b)*, o per descriure els nivells d'exigència representats per criteris existents en funció dels nivells comuns. En el futur, pot ser que es proporcionin mostres d'actuació en els diferents nivells que es puguin utilitzar com a referents en la formació per a l'aplicació de criteris estàndards.

### 9.3.11. Avaluació holística / Avaluació analítica

L'*avaluació holística* consisteix en una valoració sintètica global. L'examinador pondera intuïtivament diferents aspectes.

L'*avaluació analítica* considera separatament els diferents aspectes.



Hi ha dues maneres d'establir la distinció entre avaluació holística i avaluació analítica: *a)* segons el que es pretén avaluar; *b)* segons com s'arriba al resultat final. De vegades els sistemes combinen els dos enfocaments.

- a)* Què s'avalua: Certs enfocaments avaluen una categoria global com, per exemple, "l'expressió oral" o "la interacció", i li atribueixen una puntuació o una nota. D'altres, més analítics, exigeixen que l'examinador qualifiqui per separat diversos aspectes de l'actuació. En altres enfocaments l'examinador ha de partir d'una impressió general, analitzar les diferents categories i després fer una valoració global. L'avantatge de les categories independents en un enfocament analític és que animen l'examinador a observar amb atenció. També proporcionen un metallenguatge per a la negociació entre examinadors i per a la realimentació que es dona als aprenents. Tanmateix, té l'inconvenient que l'experiència tendeix a provar que els examinadors són incapaços de mantenir les categories separades d'una valoració global i que pateixen una sobrecàrrega mental quan han de treballar amb més de 4 o 5 categories.
- b)* Càlcul del resultat: Alguns enfocaments relacionen de manera global les actuacions observades i els descriptors d'una escala de valoració, ja sigui l'escala holística (de caràcter global) o l'analítica (de 3 a 6 categories en una graella). Aquests enfocaments no comporten cap càlcul. Els resultats són traslladats sota la forma d'una sola xifra o com un "número de telèfon", per categoria. Altres enfocaments més analítics exigeixen atorgar qualificacions a diversos elements diferents i després fer-ne la suma per obtenir un resultat, que es pot convertir en una nota. Una de les característiques d'aquest enfocament és que es ponderen les categories, és a dir, no tenen totes el mateix valor.

Els quadres 2 i 3 del capítol 3 proporcionen exemples de l'autoavaluació i de l'avaluació feta per l'examinador, respectivament, de les escales analítiques de criteris (graelles) utilitzades amb una estratègia de puntuació de caràcter global (és a dir, l'examinador relaciona el que es pot deduir de l'actuació amb les definicions i fa una valoració).

### 9.3.12. Avaluació per sèrie / Avaluació per categoria

L'*avaluació per categoria* consisteix en una sola tasca d'avaluació (que pot tenir diferents fases per generar discursos diferents, tal com s'ha vist a la secció 9.2.1) en què es valora l'actuació en funció de les categories d'una graella d'avaluació: l'enfocament analític subratllat a la secció 9.3.11.

L'*avaluació per sèrie* consisteix en una sèrie de tasques d'avaluació aïllades (sovint jocs de rol amb altres aprenents o amb el docent) que es qualifiquen amb una simple nota global segons una escala de punts, per exemple de 0 a 3 o d'1 a 4.

L'avaluació per sèrie intenta corregir la tendència existent en les avaluacions per categories que fa que els resultats d'una categoria afectin els d'una altra. En els ni-

vells inferiors es posa l'accent en l'assoliment de la tasca: la finalitat és emplenar una llista de control amb allò que l'aprenent pot fer sobre la base de l'avaluació feta pel docent o per l'aprenent de les actuacions reals, més que no pas sobre una simple impressió. En els nivells superiors es poden programar tasques en les quals es manifestin aspectes concrets del domini de la llengua en l'actuació. S'informa dels resultats mitjançant un perfil.

Les escales per a diferents categories de la competència lingüística incloses en el text del capítol 5 ofereixen un recurs per al desenvolupament dels criteris per a una avaluació per categories. Atès que els examinadors només poden treballar amb un nombre reduït de categories, s'han de fer algunes concessions en aquest procés. L'elaboració de tipus pertinents d'activitats comunicatives a la secció 4.4 i la llista dels diferents tipus de competència funcional recollida a la secció 5.2.3.2 poden contribuir a determinar tasques convenients per a una avaluació per sèrie.

### 9.3.13. Avaluació feta per altres persones / Autoavaluació

*Avaluació feta per altres persones:* valoració que fa el docent o l'examinador.

*Autoavaluació:* valoració que fa una persona del seu propi domini de la llengua.

Els aprenents poden participar en la majoria de les tècniques d'avaluació descrites anteriorment. La recerca suggereix que, si no hi ha un risc important (per exemple, poder ser admès o no en un curs), l'autoavaluació pot ser un complement eficaç de les proves i de l'avaluació que fa el docent. La precisió en l'autoavaluació augmenta *a)* quan l'avaluació es fa en relació amb descriptors que defineixen clarament patrons estàndards de domini de la llengua i/o *b)* quan l'avaluació es relaciona amb una experiència concreta. Aquesta experiència pot ser fins i tot una activitat d'examen. Probablement també es fa amb més precisió quan els aprenents reben alguna formació específica. Una autoavaluació estructurada com aquesta pot tenir correlacions amb les avaluacions que fan els docents i amb les proves, de la mateixa manera que hi sol haver correlació (nivell de validesa concurrent) entre els mateixos docents, entre les proves, i entre l'avaluació que fa el docent i les proves.

No obstant això, el potencial més important de l'autoavaluació es troba en el seu ús com a eina per a la motivació i la presa de consciència, ja que pot ajudar els aprenents a apreciar els seus punts forts i reconèixer els seus punts febles, i a orientar el seu aprenentatge d'una manera més efectiva.

Els usuaris del *Marc de referència* han de tenir present i, si escau, determinar:

- quins tipus d'avaluació, dels esmentats en aquest capítol, són:
  - més adequats a les necessitats dels aprenents en el seu sistema;
  - més adequats i viables en la cultura pedagògica del seu sistema;
  - més productius, per l'efecte de rebot, per a la formació dels docents;
- com s'equilibren i es completen en el seu sistema l'avaluació de l'assoliment

(centrada en l'escola i en l'aprenentatge) i l'avaluació del domini de la llengua (orientada al món real; orientada als resultats), i fins a quin punt s'avaluen tant les actuacions comunicatives com els coneixements lingüístics;

- fins a quin punt s'avaluen els resultats de l'aprenentatge segons les normes i els criteris definits (amb referència a un criteri) i fins a quin punt s'atribueixen les notes i les avaluacions en funció de la classe en què es troba l'aprenent (amb referència a la norma);
- fins a quin punt:
  - s'informa els docents dels nivells d'exigència (per exemple, facilitant descriptors comuns, mostres d'actuacions...);
  - s'anima els docents a ser conscients de l'existència d'una sèrie de tècniques d'avaluació;
  - es forma els docents en tècniques d'avaluació i en interpretació de resultats;
- fins a quin punt és desitjable i viable desenvolupar un enfocament integrat de l'avaluació contínua del treball de classe i de les avaluacions puntuals segons uns nivells relacionats entre si i unes definicions de criteris;
- fins a quin punt és desitjable i viable implicar els aprenents en l'autoavaluació referida a descriptors definits de tasques i a aspectes de domini de la llengua en diferents nivells, així com en l'aplicació d'aquests descriptors en, per exemple, l'avaluació per sèries;
- la importància de les especificacions i de les escales que proporciona el *Marc* per als seus contextos, i la manera com es podrien complementar o ampliar.

Les versions de les graelles de puntuació de l'autoavaluació i de l'examinador es presenten en els quadres 2 i 3 del capítol 3. La distinció més notable entre les dues –deixant de banda formulacions com *Puc fer* o *Pot fer*– és que mentre el Quadre 2 se centra en les activitats comunicatives, el Quadre 3 se centra en els aspectes generals de la competència manifestats en qualsevol actuació oral. Això no obstant, es pot imaginar fàcilment una versió simplificada del Quadre 3 per a l'autoavaluació. L'experiència indica que almenys els aprenents adults són capaços de fer aquest tipus de valoracions qualitatives sobre la seva pròpia competència.

#### 9.4. L'avaluació viable i un metasistema

Les escales dels capítols 4 i 5 presenten un exemple d'un conjunt de categories simplifiades relacionades amb l'esquema descriptiu més exhaustiu que es desenvolupa en els capítols 4 i 5. La intenció dels autors no és que tothom utilitzi, en un enfocament pràctic d'avaluació, totes les escales en tots els nivells. Per als examinadors és difícil treballar amb un gran nombre de categories, i, a més, la sèrie completa de nivells exposats pot ser que no sigui adequada al context concret. El conjunt d'escales més aviat té la intenció de ser una eina de referència.

Independentment de l'enfocament que s'adopti, qualsevol sistema d'avaluació pràctic ha de reduir el nombre de categories possibles a un nombre viable. L'experiència demostra que més de 4 o 5 categories provoquen una sobrecàrrega mental, i que 7 categories és, psicològicament, el límit màxim. Això vol dir que cal escollir. Pel que fa a l'avaluació oral, si es considera que les estratègies d'interacció són un aspecte qualitatiu de la comunicació pertinent en l'avaluació oral, aleshores les escales il·lustratives contenen 14 categories pertinents per a l'avaluació oral:

- Estratègies d'intercanvi de torns de paraula
- Estratègies de cooperació
- Demandar aclariments
- Fluïdesa
- Flexibilitat
- Coherència
- Desenvolupament temàtic
- Precisió
- Competència sociolingüística
- Repertori lingüístic general
- Repertori de vocabulari
- Correcció gramatical
- Domini de vocabulari
- Domini fonològic

És evident que, encara que els descriptors de moltes d'aquestes categories es podrien incloure en una llista general, 14 categories són massa per a l'avaluació de qualsevol actuació. En qualsevol enfocament pràctic, per tant, la llista de categories s'hauria de plantejar de manera selectiva. Les categories s'han de combinar, reformular i reduir a un conjunt més petit de criteris d'avaluació adequat a les necessitats dels aprenents implicats, a les exigències de la tasca d'avaluació concreta i a l'estil de la cultura pedagògica específica. Els criteris resultants es podrien ponderar equitativament o, com a alternativa, es podria donar més pes específic a certs factors que són determinants per a la tasca concreta.

Els quatre exemples següents presenten les maneres de portar-ho a terme. Els tres primers mostren breument de quina manera es poden utilitzar les categories com a criteris d'examen en els enfocaments d'avaluació existents. El quart exemple mostra de quina manera els descriptors de les escales del *Marc* es van combinar i reformular per proporcionar una graella d'avaluació per a un propòsit concret en una situació determinada.

**Exemple 1**

*Cambridge Certificate in Advanced English (CAE), prova 5: Criteris d'avaluació (1991)*

<b>Criteris d'avaluació</b>	<b>Escales</b>	<b>Altres categories<sup>1</sup></b>
Fluïdesa	Fluïdesa	
Correcció i repertori	Repertori lingüístic general Repertori de vocabulari Correcció gramatical Domini de vocabulari	
Pronunciació	Domini fonològic	
Assoliment de la tasca	Coherència Adequació sociolingüística	Èxit en la tasca Necessitat de l'ajuda de l'interlocutor
Comunicació interactiva	Estratègies d'intercanvis de torns de paraula Estratègies de cooperació Desenvolupament temàtic	Facilitat de mantenir la participació i grau en què es manté.

**Exemple 2**

*International Certificate Conference (ICC): Certificat d'anglès per a afers professionals, prova 2: Conversa professional (1987)*

<b>Criteris d'avaluació</b>	<b>Escales</b>	<b>Altres categories</b>
Escala 1 (sense títol)	Adequació sociolingüística Correcció gramatical Domini de vocabulari	Èxit en la tasca
Escala 2 (utilització dels elements discursius per iniciar i mantenir una conversa)	Estratègies d'intercanvis de torns de paraula Estratègies de cooperació Adequació sociolingüística	

**Exemple 3**

*Eurocentres: avaluació de la interacció en grups petits (RADIO) (1987)*

<b>Criteris d'avaluació</b>	<b>Escales</b>	<b>Altres categories</b>
Repertori	Repertori lingüístic general Repertori de vocabulari	
Correcció	Correcció gramatical Domini de vocabulari Adequació sociolingüística	
Expressió	Fluïdesa Domini fonològic	
Interacció	Estratègies d'intercanvis de torns de paraula Estratègies de cooperació	

<sup>1</sup> En les escales, els enunciat sobre l'èxit de la tasca es relacionen amb les *Activitats comunicatives*. La *facilitat de mantenir la participació i el grau en què es manté* s'inclou dins la categoria de *Fluïdesa* en aquestes escales. L'intent de redacció i de calibratge dels descriptors corresponents a *Necessitat de l'ajuda de l'interlocutor* per incloure'ls en el grup d'escales no va tenir èxit.

### Exemple 4

Consell Nacional Suís per a la Recerca: Avaluació d'actuacions en vídeo.

Context: A l'annex A s'explica com es van classificar per escales els descriptors en un projecte de recerca que es va dur a terme a Suïssa. Al final del projecte de recerca, els docents que hi van participar van ser convidats a un congrés per presentar-hi els resultats i començar l'experimentació del Carnet Europeu de Llengües a Suïssa. En el congrés dos dels temes de debat van ser: *a)* la necessitat de relacionar l'avaluació contínua i les llistes de control per a l'autoavaluació amb un marc general, i *b)* les maneres d'utilitzar en l'avaluació els descriptors del projecte classificats per escales. En el curs d'aquesta discussió es van valorar vídeos d'alguns dels aprenents de l'estudi en una graella d'avaluació presentada com a Quadre 3 en el capítol 3. Presenta una selecció de descriptors en una forma combinada i corregida.

Críteris d'avaluació	Escales	Altres categories
Repertori	Repertori lingüístic general Repertori de vocabulari	
Correcció	Correcció gramatical Domini de vocabulari	
Fluïdesa	Fluïdesa	
Interacció	Interacció global Torns de paraula Cooperació	
Coherència	Coherència	

Diferents sistemes, amb aprenents diferents, en contextos diferents, simplifiquen, seleccionen i combinen els elements de maneres diferents per a tipus d'avaluació diferents. De fet, en lloc de ser molt llarga, la llista de 14 categories probablement no pot tenir en compte totes les variants que les persones trien i s'hauria d'augmentar per tal de ser completament exhaustiva.

Els usuaris del *Marc de referència* han de tenir present i, si escau, determinar:

- de quina manera se simplifiquen les categories teòriques per esdevenir enfocaments funcionals en el seu sistema;
- fins a quin punt els principals factors utilitzats com a criteris d'avaluació en el seu sistema es poden situar en la sèrie de categories introduïdes en el capítol 5 (per a les quals es proporcionen escales de mostra a l'annex), sempre que es desenvolupin amb més detall per tal de tenir en compte els àmbits específics d'ús.