

9 Hodnocení

9.1 Úvod

Pojem hodnocení je v této kapitole užíván ve smyslu hodnocení úrovně ovládnutí jazyka, které dosáhl uživatel jazyka. Všechny jazykové testy jsou formou hodnocení, ale existují i způsoby hodnocení, jež nemohou být považovány za testování (například tzv. kontrolní seznamy pro průběžné hodnocení nebo neformální pozorování ze strany učitele). Termín evaluace vyjadřuje hodnocení v širším smyslu. Každé hodnocení je formou evaluace, avšak v kontextu jazykové výuky bývá vedle ovládnutí jazyka studentem evaluováno mnoho dalších aspektů, například efektivita učebních metod či materiálů, druh a kvalita promluv, které vznikají v průběhu výuky, kvalita vztahů mezi učitelem a studentem, efektivita výuky atd. Tato kapitola se zabývá hodnocením v užším smyslu a pomíjí evaluaci výukového programu jako celku.

Žádná diskuse o hodnocení se neobejde bez tří pojmů, které jsou již tradičně považovány za zásadní. Jsou to validita, reliabilita a praktičnost. Vzhledem k obsahu této kapitoly bude užitečné, seznámíme-li se nyní s tím, co tyto pojmy označují, s jejich vzájemným vztahem a s jejich významem ve Společném evropském rámci.

Validita je z pohledu Rámce skutečně nejdůležitější. Jakýkoliv test nebo hodnotící procedura mohou být považovány za validní, lze-li doložit, že to, co je reálně hodnoceno (konstrukt), je to, co hodnoceno *být má*, a že výsledek hodnocení přesně odráží úroveň ovládnutí jazyka hodnocených kandidátů.

Reliabilita je čistě technický pojem. Vyjadřuje míru, v níž je zachováno výkonnostní pořadí hodnocených studentů při dvou nezávislých aplikacích (skutečných nebo simulovaných) téže hodnotící procedury.

Mnohem důležitější než reliabilita je při hodnocení *přesnost rozhodování* vzhledem k daným standardům požadavků. Je-li hodnocení vyjádřeno výsledky jako „vyhověl/nevyhověl“ nebo přiřazením ke stupňům A2+/B1/B1+, nakolik jsou učiněná rozhodnutí přesná? Přesnost rozhodování je závislá na validitě standardních požadavků (např. pro stupeň B1) v daném kontextu. Kromě toho také závisí na validitě kritérií, jimiž se rozhodování řídí, a na validitě postupu, na jehož základě byla tato kritéria vytvořena.

Užívají-li dvě různé organizace či regiony pro rozhodování při hodnocení téže jazykové dovednosti kritéria, která se vztahují k týmž standardním požadavkům, jsou-li tyto požadavky validní a přiměřené v obou daných kontextech a jsou-li patřičně zakomponovány do tvorby zkušebních úloh a do procesu hodnocení, budou si výsledky hodnocení v těchto dvou systémech vzájemně odpovídat. Korelace mezi dvěma testy určenými k hodnocení téhož konstruktů se tradičně označuje pojmem „*souběžná validita*“.

Tento pojem je samozřejmě úzce svázán s pojmem *reliabilita* (spolehlivost), protože nespolehlivé testy si vzájemně odpovídat nemohou. Nejdůležitější však je, nakolik si takové dva testy odpovídají v tom, *co se hodnotí a jak je interpretován výkon studenta*.

Právě na tyto dva body se Společný evropský referenční rámec soustředí nejvíce. Následující oddíl textu naznačuje tři hlavní oblasti, v nichž lze v tomto směru z Rámce čerpat:

- | | |
|---|-------------------------------------|
| 1. Specifikace obsahu testů a zkoušek: | <i>co se hodnotí</i> |
| 2. Stanovování kritérií splnění učebního cíle: | <i>jak je výkon interpretován</i> |
| 3. Určení úrovně ovládnutí jazyka u již existujících testů a zkoušek, což umožňuje srovnávání mezi různými hodnoticími systémy: | <i>jakým způsobem lze srovnávat</i> |

Různé formy hodnocení se k těmto bodům staví různě. Existuje mnoho druhů a tradičních způsobů hodnocení a není správné považovat jeden přístup (např. veřejnou zkoušku) v jeho vzdělávacím účinku za nadřazený jinému (např. hodnocení učitelem). Hlavní výhodou souboru společných standardních požadavků – jakým jsou například Společné referenční úrovně popsané v Rámci – je právě to, že umožňuje srovnání mezi nejrůznějšími formami hodnocení.

Třetí oddíl této kapitoly vypočítává různé typy hodnocení, mezi nimiž lze volit. Tyto možnosti jsou prezentovány formou kontrastních dvojic. U každé dvojice termínů je vysvětlen jejich význam a relativní výhody a nevýhody obou možností vzhledem k účelu hodnocení a vzdělávacímu kontextu. Dále se hovoří o důsledcích volby mezi těmito dvěma alternativními postupy a posléze je u obou možností hodnocení uvedeno, jak při nich může pomoci Rámec.

Každá hodnoticí procedura musí být také *praktická*, tedy uskutečnitelná. Praktičnost je důležitá především při výkonovém testování. Zkoušející disponují omezeným časem, hodnotí pouze vzorek výkonnosti zkoušeného a omezen je také počet a typ kategorií, které jim mohou sloužit jako kritéria. Rámec má sloužit jako konzultační materiál, nikoliv jako praktická příručka pro hodnocení. K širokému okruhu informací, které Rámec pokrývá, musejí jeho uživatelé přistupovat selektivně. Selektivní přístup může někdy spočívat i ve zjednodušení pracovního schématu, čímž dochází ke splynutí některých kategorií, které Rámec rozlišuje jako samostatné. Například kategorie uvedené v modelových stupnicích deskriptorů, které doprovázejí text v kapitolách 4 a 5, jsou často ve skutečnosti mnohem jednodušší než jejich podoba, o níž se hovoří v textu. Této problematice je věnován poslední oddíl této kapitoly, kde jsou uvedeny i konkrétní příklady.

9.2 Rámec jako zdroj informací o hodnocení

9.2.1 Specifikace obsahu testů a zkoušek

Při navrhování specifického obsahu úloh pro komunikativní hodnocení může pomoci pasáž „Užívání jazyka a uživatelé jazyka“ v kapitole 4 a především oddíl 4.4 s názvem „Komunikativní jazykové činnosti a strategie“. Stále více platí, že má-li být hodnocení validní, je nutné posuzovat více vzorků z okruhu relevantních typů promluv. V oblasti testování mluveného projevu ilustruje zmíněné tvrzení například tento nedávno navržený postup zkoušky: začíná se simulovanou *konverzací*, která slouží jako „zahřívací kolo“, pak přichází *neformální*

diskuse o aktuálních otázkách, o něž student projeví zájem, následuje *vyjednávací fáze*, která má formu simulovaného nebo skutečného telefonátu, jehož cílem je získat určité informace, poté přichází na řadu *produkční fáze* založená na písemné *zprávě*, v níž student *popisuje* zkoušejícím své studijní zájmy a záměry, a celá zkouška končí *fází cílené spolupráce*, úkolem, v němž se mají studenti dohodnout na společném řešení.

Shrneme-li to, Společný evropský referenční rámec rozlišuje u zmíněných komunikativních činností následující kategorie:

	Interakce (krátké spontánní repliky)	Produkce (delší připravené repliky)
Ústní projev	<i>konverzace</i> <i>neformální diskuse</i> <i>účelová spolupráce</i>	<i>popis</i> oblasti studijních zájmů
Písemný projev		<i>zpráva/popis</i> oblasti studijních zájmů

Při detailní specifikaci obsahu úloh může uživatel Rámce nahlédnout také do oddílu 4.1 „Kontext užívání jazyka“ (oblasti užívání jazyka, podmínky a omezení, vnitřní kontext), do oddílu 4.6 „Texty“, a do kapitoly 7 „Učební úlohy a jejich role v jazykové výuce“, především do oddílu 7.3 „Obtížnost úloh“.

Oddíl 5.2 „Komunikativní jazykové kompetence“ může pomoci sestavit jednotlivé části testu či zkoušky tak, aby bylo možné zhodnotit jejich prostřednictvím žádoucí jazykové, sociolingvistické a pragmatické kompetence. Za doplňky k hlavnímu dokumentu Rámce lze považovat soubor obsahových specifikací pro úroveň ovládnutí B1 (*Threshold*), sestavený Evropskou radou pro více než 20 evropských jazyků (viz bibliografie ke kapitole 5), a pro úrovně A2 (*Waystage*) a B2 (*Vantage*) u anglického jazyka, a také ekvivalenty těchto dokumentů pro jiné jazyky a jejich úrovně ovládnutí. Zmíněné materiály nabízejí příklady a podrobnější rady k tvorbě testů pro stupně A1, A2, B1 a B2.

9.2.2 Kritéria splnění učebního cíle

Stupnice uvedené v Rámci mohou napomoci při tvorbě hodnotících stupnic, jejichž prostřednictvím bude možné zhodnotit splnění určitého učebního cíle, a z uvedených deskriptorů lze čerpat při vytváření hodnotících kritérií. Učební cíl může být pojímán buď ze širšího hlediska Společných referenčních úrovní (např. B1), nebo může mít podobu specifické sestavy jazykových činností, dovedností a kompetencí, jak o tom hovoří oddíl 6.1.4 „Dílčí kompetence a obměny v učebních cílech ve vztahu k Rámci“. Takto formulovaný cíl může být znázorněn tabulkou jazykových kategorií v rámci jednotlivých úrovní ovládnutí. Příklad uvádí Tabulka 2.

V diskusi o používání deskriptorů je velmi důležité rozlišovat:

1. Deskriptory komunikativních činností, které jsou uvedeny v kapitole 4.
2. Deskriptory různých aspektů ovládnutí jazyka ve vztahu k určitým kompetencím, které jsou uvedeny v kapitole 5.

První skupina je velmi užitečná pro hodnocení studenta učitelem a pro sebehodnocení u úloh z reálného života. Obě tyto formy hodnocení se zakládají na detailním obrazu jazykových schopností, které si student vytváří v průběhu celého jazykového kursu. Tento typ deskriptorů je oblíbený, protože pomáhá učitelům i studentům soustředit se na akčně zaměřený přístup k výuce.

Není však obvykle vhodné, jsou-li deskriptory komunikativních činností součástí kritérií hodnotitele, který má posuzovat výkon studenta v ústním nebo písemném testu, pokud chceme zjistit celkovou úroveň ovládnutí jazyka. Chceme-li totiž hodnotit ovládnutí jazyka, neměli bychom se primárně soustředit na hodnocení onoho jednotlivého výkonu, ale spíše na to, o jakých zobecnitelných kompetencích tento výkon svědčí. Samozřejmě, že někdy nás oprávněné důvody vedou i k hodnocení úspěšnosti v jedné určité činnosti, především u mladších uživatelů na základní úrovni (stupně A1 a A2). Výsledky pak mohou být obtížněji zobecnitelné, ale v počátečních stádiích jazykové výuky není zobecnitelnost výsledků tolik důležitá.

To jen znovu dokazuje, že hodnocení může mít různou funkci. Co je vhodné pro jeden účel hodnocení, nemusí být vhodné pro jiný.

9.2.2.1 Deskriptory komunikativních činností

Deskriptory komunikativních činností (kapitola 4) lze při hodnocení dosažení výukových cílů využívat ve třech samostatných oblastech.

1. **Tvorba testů:** Jak již bylo řečeno v oddíle 9.2.1, stupnice komunikativních činností pomáhají při specifikaci obsahu hodnotících úloh a při jejich vytváření.
2. **Prezentace výsledků:** Stupnice pro hodnocení komunikativních činností mohou být užitečné také pro prezentaci výsledků studenta. Uživatelé produktů vzdělávacího systému, například zaměstnavatelé, se často zajímají více o celkové výsledky než o detailní profil kompetencí.
3. **Hodnocení učitelem nebo sebehodnocení:** Deskriptory komunikativních činností mohou být různým způsobem využívány při hodnocení učitelem a při sebehodnocení studenta, například v podobě:
 - **Kontrolního seznamu:** užívá se při průběžném nebo souhrnném hodnocení na konci jazykového kursu. U každé úrovně ovládnutí může být uveden seznam deskriptorů. Obsah deskriptorů může být popřípadě konkrétně rozveden. Například deskriptor *Dokážu zjistit i poskytnout osobní údaje* může být rozveden do podoby konkrétních složek *Dokážu se představit. Dokážu říci, kde bydlím. Dokážu říci svou adresu francouzsky. Dokážu říct, kolik je mi let.* atd. Dále *Dokážu se někoho zeptat, jak se jmenuje, kde bydlí a kolik je mu let.*
 - **Tabulky:** užívají se při průběžném nebo souhrnném hodnocení. Profil schopností je zaznamenáván do tabulky podle vybraných kategorií (např. *konverzace, diskuse, výměna informací*) a popsanych pro jednotlivé úrovně (např. B1+, B2, B2+).

Takové využití deskriptorů je v posledních deseti letech stále oblíbenější. Zkušenost ukázala, že interpretace deskriptorů ze strany učitelů i studentů se zpřesňuje a zjednodušuje, popisují-li deskriptory nejen CO student DOKÁŽE, ale také JAK to DOKÁŽE.

9.2.2.2 Deskriptory různých aspektů ovládání jazyka ve vztahu k určitým kompetencím

Pokud jde o dosažení výukových cílů, mohou deskriptory pomoci s hodnocením především v následujících dvou oblastech:

1. **Sebehodnocení nebo hodnocení učitelem:** Mají-li deskriptory podobu *pozitivně formulovaných nezávislých kritérií*, mohou být zahrnuty do kontrolních seznamů pro hodnocení učitelem a pro sebehodnocení.

Nedostatkem většiny užívaných hodnotících stupnic je právě to, že deskriptory bývají u nižších úrovní ovládání formulovány negativně a kolem středu hodnotící stupnice mají normativní povahu. Mezi úrovněmi ovládání bývá často jen nepatrný rozdíl jednoho nebo dvou slov, která jsou u dvou po sobě následujících deskriptorů pouze obměněna, takže tyto ztrácejí mimo kontext dané stupnice svůj význam. Jak se při tvorbě deskriptorů vyhnout těmto úskalím, je stručně uvedeno v Příloze A.

2. **Výkonové hodnocení:** Stupnice deskriptorů různých aspektů jazykové kompetence uvedené v kapitole 5 lze však především využít jako východisko při vytváření hodnotících kritérií. Tím, že osobní a nesystémové dojmy převádějí do podoby uvážlivého úsudku, pomáhají tyto deskriptory vytvořit pro skupinu hodnotitelů společný referenční rámec.

Existují v zásadě tři formy prezentace deskriptorů jako hodnotících kritérií:

- Lze je prezentovat formou *stupnice*. Deskriptory jednotlivých kategorií tvoří u každé úrovně ovládání samostatný odstavec. Tento přístup je velmi oblíbený.
- Lze je prezentovat formou *kontrolního seznamu*. Pro každou úroveň ovládání je obvykle vytvořen samostatný seznam a deskriptory bývají seskupeny podle kategorií. Kontrolní seznamy jsou méně obvyklé u hodnocení aktuálně probíhajícího výkonu.
- Lze je prezentovat formou *tabulky* vybraných kategorií, tedy jako soubor paralelních stupnic pro jednotlivé kategorie. Tento přístup umožňuje zaznamenat diagnostický profil studenta. Počet kategorií, na něž se zkoušející mohou soustředit, je však omezený.

Stupnice v tabulce mohou být ještě dále rozděleny na dílčí stupně dvěma zcela odlišnými způsoby:

Stupnice úrovní ovládání jazyka vznikne tak, že profilová tabulka udává pro určité kategorie relevantní úrovně ovládání, například v rozmezí stupňů A2 a B2. Hodnocení pak probíhá přímo na těchto úrovních s možným dalším zpřesněním, které v případě potřeby umožňuje přesnější rozlišení v podobě desetinného místa nebo znaménka +. Výsledkem je, že ačkoliv byl test zaměřen např. na stupeň B1 a žádný ze studentů nedosáhl stupně B2, schopnější studenti mohli získat hodnocení B1+, B1++ nebo B 1,8.

Známkovací stupnice vznikne tak, že vybereme pro každou relevantní kategorii určitý deskriptor, který vyjadřuje standardní požadavek nebo normu pro složení zkoušky v dané kategorii. Tento deskriptor pak můžeme označit známkou „vyhověl“ nebo „3“ a zbytek stupnice se bude vztahovat k této normě (velmi slabý výkon = „1“, vynikající výkon = „5“). Deskriptory v rozmezí „1“ až „5“ mohou být formulovány samostatně nebo adaptovány podle posloupnosti úrovní ovládání uvedené v příslušném oddíle kapitoly 5, popř. mohou být formulovány ve vztahu k deskriptoru označenému známkou „3“.

9.2.3 Popis úrovní ovládání jazyka u testů a zkoušek za účelem snadnějšího srovnávání

Stupnice Společných referenčních úrovní mají za úkol usnadnit popis úrovně ovládání jazyka u různých typů zkoušek, a tak umožnit jejich vzájemné porovnávání. Literatura o poměřování rozeznává pět klasických způsobů propojení dvou samostatných hodnocení: 1. vyvážení (různých stupnic), 2. kalibrace, 3. statistické upřesnění, 4. referenční porovnávání (benchmarking) a 5. sociální upřesnění.

První tři metody jsou tradiční: 1. jsou vzájemně porovnávány stupnice různých testů (vyvážení), 2. výsledky různých testů se zhodnotí podle téže stupnice (kalibrace) a 3. provede se upřesnění vzhledem k obtížnosti testů nebo náročnosti zkoušejících (statistické upřesnění).

Zbývající dvě metody spočívají v nastolení vzájemného porozumění prostřednictvím diskuse (sociální upřesnění) a v porovnání pracovních vzorků se standardizovanými definicemi a referenčními ukazateli (benchmarking). Podpořit tento proces budování vzájemného porozumění je jedním z cílů Rámce. Proto byly stupnice deskriptorů navržené k tomuto účelu vypracovány přesným a standardizovaným metodickým postupem. V pedagogice je tento přístup k hodnocení stále častěji označován jako hodnocení zaměřené na standardy. Zavést tento přístup k hodnocení je jistě časově náročné, protože jednotliví hodnotitelé se musejí prostřednictvím vzájemné diskuse a srovnáváním různých příkladů postupně naučit standardy přesně rozlišovat.

Lze říci, že takový přístup je potenciálně nejlepší metodou vzájemného propojování, protože spočívá ve vytvoření a validaci společného pohledu na jazykový konstrukt. Základní příčinou problémů při srovnávání různých jazykových hodnocení je navzdory „statistickým kouzlům“ tradičních postupů fakt, že hodnocení se obvykle týká radikálně odlišných jevů, ačkoliv má za cíl pokrývat tutéž oblast. Příčinou je částečně a) nedostatečný pojmový aparát a nedostatečná funkčnost konstruktů a b) s tím související rušivý vliv testovacích metod.

Rámec se pokouší nabídnout pro tento závažný problém, jež zatěžuje výuku moderních jazyků v evropském kontextu, možné systémové řešení. V kapitolách 4 a 7 je rozpracováno popisné schéma, které se snaží zavést do oblasti užívání jazyka, jazykových dovedností a procesu vyučování a učení praktické pojmy, což má partnerům v hodnotícím procesu pomoci lépe pracovat s komunikativními jazykovými schopnostmi, jejichž rozvoj se snažíme podporovat.

Stupnice deskriptorů tvoří pojmovou tabulku, kterou lze využít:

- a) ke vzájemnému srovnávání národních a institucionálních referenčních rámců prostřednictvím Společného evropského rámce;
- b) ke zmapování cíle určité zkoušky nebo modulů kursu za pomoci kategorií a úrovní ovládání na stupnicích.

Příloha A nabízí přehled metod, jak vytvářet stupnice deskriptorů a jak je srovnávat se stupnicí uvedenou v Rámci.

Průvodce pro examinátory u testu ALTE (dokument CC-Lang (96) 10 rev) nabízí podrobné rady, jak pracovat s jazykovým konstruktem v testech a jak se vyhnout zbytečnému zkreslení vlivem metody testování.

9.3 Typy hodnocení

Hodnocení lze dělit na různé typy podle mnoha důležitých hledisek. Následující výčet není v žádném případě vyčerpávající. Umístění jednotlivých typů do pravého a levého sloupce je náhodné.

Tabulka 7. *Typy hodnocení*

1	Didaktické hodnocení	Hodnocení způsobilosti (úrovně ovládní)
2	Normativní hodnocení	Kriteriální hodnocení
3	Kriteriální hodnocení zvládacího učení	Kontinuální kriteriální hodnocení
4	Průběžné hodnocení	Hodnocení ve stanovených termínech
5	Formativní hodnocení	Sumativní hodnocení
6	Přímé hodnocení	Nepřímé hodnocení
7	Výkonové hodnocení	Hodnocení znalostí
8	Subjektivní hodnocení	Objektivní hodnocení
9	Hodnocení pomocí kontrolního seznamu	Celkové hodnocení výkonu
10	Hodnocení na základě dojmu	Hodnocení na základě řízeného úsudku
11	Holistické hodnocení	Analytické hodnocení
12	Hodnocení souborné	Hodnocení kategoriální
13	Hodnocení jinou osobou	Sebehodnocení

9.3.1 Didaktické hodnocení / hodnocení způsobilosti (úrovně ovládní jazyka)

Didaktické hodnocení (hodnocení školního výkonu) udává míru dosažení určitého cíle – hodnotí to, co bylo předmětem výuky. Vztahuje se tedy k jednotlivým týdnům či semestrům nebo k obsahu učebnice či sylabu. Didaktické hodnocení je zaměřeno na určitý kurs. Představuje vnitřní perspektivu.

Hodnocení způsobilosti (úrovně ovládní jazyka) na druhé straně hodnotí to, jak umí student své znalosti aplikovat ve skutečném životě. Takové hodnocení představuje vnější perspektivu.

Učitelé mívají přirozený sklon více se soustředit na hodnocení školního výkonu, které jim poskytuje zpětnou vazbu pro další výuku. Zaměstnavatelé, pracovníci školské administrativy a dospělí studenti se oproti tomu více zajímají o jazykovou způsobilost (stupeň ovládní jazyka), o celkové výsledky, o to, co člověk v dané chvíli dokáže.

Výhoda didaktického hodnocení spočívá v tom, že má bližší vztah k aktuální zkušenosti studenta. Výhodou hodnocení úrovně ovládní jazyka je to, že pomáhá studentovi celkově a jasně zhodnotit jeho současné schopnosti.

Pokud jde o komunikativní testování v kontextu výuky zaměřené na potřeby studenta, mohli bychom tvrdit, že rozdíl mezi úspěšností (v jazykovém kursu) a úrovní ovládní jazyka (v reálném životě) musí být malý. Platí, že je-li předmětem hodnocení úspěšnosti praktické užívání jazyka v relevantní situaci a snaží-li se toto hodnocení podat vyrovnaný obraz

rozvíjející se jazykové kompetence, odráží také úroveň ovládnutí jazyka. Platí také, že hodnotí-li se úroveň ovládnutí jazyka prostřednictvím jazykových a komunikativních úloh vycházejících z relevantního a jasněho sylabu, může student ukázat, čemu se naučil, a hodnocení tak odráží i jeho úspěšnost.

Stupnice vzorových deskriptorů jsou zaměřeny na hodnocení úrovně ovládnutí jazyka – na soubor schopností použitelných v reálném životě. O významu hodnocení úspěšnosti pro podporu procesu učení se hovoří v kapitole 6.

9.3.2 Hodnocení normativní / hodnocení kritériální

Normativní hodnocení spočívá v sestavení pořadí úspěšnosti studentů, kteří jsou pak hodnoceni vzhledem ke svým spolužákům.

Kritériální hodnocení vzniklo jako reakce na hodnocení normativní. Student je hodnocen čistě na základě svých schopností v daném předmětu, nezávisle na schopnostech svých spolužáků.

Normativní hodnocení lze užít v rámci třídy (jsi na 18. místě) nebo v určitém demografickém měřítku (jsi na 21 567. místě, náležíš k horním 14 %) nebo ve skupině testovaných studentů. V druhém případě mohou základní výsledky testu získat spravedlivější podobu tak, že se bodovací křivka testu promítne na bodovací křivky z minulých let, což zajistí, že např. známka „A“ bude každoročně udělena stejnému procentu studentů, a to bez ohledu na obtížnost testu a schopnosti žáků. Normativní hodnocení se často užívá u rozmísťovacích testů při zařazování studentů do tříd.

Kritériální hodnocení spočívá v posuzování průběhu ovládnutí jazyka (vertikální směr) a okruhu relevantních jazykových oblastí (horizontální směr), takže výsledky testu lze vyznačit v prostoru mezi těmito souřadnicemi. Je nutné a) definovat relevantní oblast či oblasti, které má daný test pokrýt a b) stanovit „rozhraní“, tedy počet či počty bodů, které testovaný musí získat, aby vyhověl danému souboru standardních požadavků.

Stupnice vzorových deskriptorů představují schéma, v němž jsou charakterizována kritéria pro jednotlivé kategorie. Společné referenční úrovně představují soubor společných standardních požadavků.

9.3.3 Kritériální hodnocení zvládacího učení / kontinuální kritériální hodnocení

Kritériální hodnocení zvládacího učení (mastery learning) probíhá tak, že se stanoví určitá minimální úroveň kompetence nebo „rozhraní“ a studenti jsou rozděleni na ty, kteří „zvládli“ a na ty, kteří „nezvládli“, a to bez podrobnějšího hodnocení kvality jejich výkonu.

Kontinuální kritériální hodnocení probíhá tak, že schopnosti studenta jsou srovnávány s definovanou posloupností všech relevantních stupňů dovednosti v testované oblasti.

Ke kritériálnímu hodnocení lze přistupovat různě, ale většinou je takové hodnocení zaměřeno právě na „zvládnutí určitého učiva“ nebo na „posloupnost kritérií“. Kritériální hodnocení bývá často mylně ztotožňováno výlučně s hodnocením podle kritéria zvládnutí, což způsobuje mnohé nejasnosti. Hodnocení podle kritéria zvládnutí je zaměřeno na úspěšnost a vztahuje se k obsahu určitého jazykového kursu. Klade menší důraz na zařazení tohoto obsahu do posloupnosti úrovní ovládnutí jazyka.

Alternativní možností u kriteriálního hodnocení je ohodnotit výsledky každého testu podle posloupnosti úrovní ovládnání jazyka, většinou v podobě známek. V této variantě je „kritériem“ právě ona posloupnost, je vnější realitou, která zajišťuje, že výsledky testu mají nějaký význam. K hodnocení podle takového vnějšího kritéria lze využít stupnicovou analýzu (např. Raschův model), která umožňuje porovnat výsledky všech testů mezi sebou a zhodnotit je podle jedné společné stupnice.

Rámec nabízí pomoc u obou forem kriteriálního hodnocení. Stupnici úrovní ovládnání jazyka, která se užívá při průběžném kriteriálním hodnocení, lze srovnat se Společnými referenčními úrovněmi. Cílový požadavek u „kritéria zvládnutí“ lze vyznačit v pojmové tabulce kategorií a úrovní ovládnání, která je také součástí Rámce.

9.3.4 Průběžné hodnocení / hodnocení ve stanovených termínech

Průběžné hodnocení je takové hodnocení, kdy učitel, popřípadě samotný student, hodnotí v průběhu celého kursu výkony v hodině, písemné práce a práci na projektech. Závěrečná známka tedy odráží celý kurs, rok, pololetí.

Hodnocení ve stanovených termínech je takové hodnocení, kdy se v určitý den, většinou na konci kursu, popřípadě na jeho začátku, přidělují známky na základě zkoušky nebo jiné hodnotící procedury. Předchozí skutečnosti nejsou brány v úvahu, rozhodující je pouze aktuální výkon studenta.

Hodnocení se často považuje za něco, co se děje v určitých termínech mimo samotný kurs. Termín průběžné hodnocení naznačuje, že v tomto případě se jedná o nedílnou součást kursu, která kumulativním způsobem přispívá k závěrečnému hodnocení. Kromě známkování domácích úkolů a příležitostných či pravidelných výkonnostních testů, které podporují proces učení, může průběžné hodnocení probíhat také formou kontrolních seznamů či tabulek, které vyplňují učitelé nebo samotní studenti, nebo formou hodnocení série tematických úloh. Lze také formálně hodnotit průběžnou práci studenta v hodinách nebo vést složku jeho písemných prací, nejlépe z různých časových údobí a v různém stádiu rozpracování.

Oba přístupy mají své výhody i nevýhody. Hodnocení ve stanovených termínech zajišťuje, že studenti budou stále schopni předvést něco, co bylo součástí sylabu před dvěma lety. Vede však ke zkouškovým traumatům a zvýhodňuje určitý typ studenta. Průběžné hodnocení umožňuje lépe zohlednit individuální přednosti a tvořivost, spoléhá však na schopnost učitele být objektivní. Je-li přivedeno do extrému, může proměnit život studenta v jeden nekonečný test a život učitele v byrokratickou noční můru.

Pro průběžné hodnocení mohou být užitečné kontrolní seznamy kritérií popisující jazykové schopnosti ve vztahu ke komunikativním činnostem (kapitola 4). Hodnotící stupnice vypracované ve vztahu k deskriptorům různých aspektů kompetence (kapitola 5) mohou pomoci při známkování u hodnocení ve stanovených termínech.

9.3.5 Formativní hodnocení / sumativní hodnocení

Formativní hodnocení spočívá v průběžném shromažďování informací o stavu vědomostí a o silných a slabých stránkách studenta. Učitel těchto poznatků využívá při plánování dalších hodin a při poskytování zpětné vazby studentům. Termín formativní hodnocení se

často užívá v širokém významu, kdy jsou brány v úvahu i informace z dotazníků a konzultací, které nelze kvantifikovat.

Sumativní hodnocení znamená, že znalosti studenta jsou na konci kursu ohodnoceny jedinou známkou. Nemusí jít pokaždé o hodnocení úrovně ovládnutí jazyka. Sumativní hodnocení je často hodnocení normativní, výkonové a odehrává se ve stanoveném termínu.

Předností formativního hodnocení je to, že jeho cílem je zkvalitnit proces učení. Nedostatky tohoto hodnocení vyplývají z povahy samotné zpětné vazby. Ta funguje jen tehdy, je-li ji příjemce schopen a) *rozpoznat*, tj. je-li pozorný, motivovaný a obeznámený s formou, jíž je informace podávána; b) *přijmout*, tj. není-li přehlacen informacemi, má-li možnost si je uložit, zařadit a vstřebat; c) *interpretovat*, tj. má-li dostatečné předcházející znalosti, aby dané informaci rozuměl a neuchýlil se ke kontraproduktivnímu jednání a d) *integrovat* tyto informace, tj. disponuje-li časem, zaměřením a dostatečnými zdroji, aby mohl nové informace promyslet, zařadit a tak si je zapamatovat. To vše vyžaduje autoregulaci, k níž je nutné studenty vést, aby byli schopni sami sledovat a ovlivňovat proces svého učení a naučili se zpětnou vazbu patřičně využívat.

Toto vedení studentů k zvyšování zodpovědnosti, či podporování uvědomělého přístupu se nazývá *formativní hodnocení*. Uvědomělé učení studenta lze podporovat nejrůznějšími metodami. Základním principem je porovnání subjektivního názoru (např. kterou z dovedností v kontrolním seznamu si označíme jako zvládnutou) se skutečností (např. poslechneme si nahrávku odpovídající typu, který jsme v seznamu vyznačili, a přesvědčíme se, zda ji skutečně rozumíme). Právě tímto způsobem je porovnáváno sebehodnocení se skutečnou úspěšností v testu DIALANG. Další důležitou metodou je rozbor jak neutrálních vzorků, tak vzorků skutečných prací studentů, při němž si studenti osvojují metajazyk nutný k hodnocení různých kvalitativních aspektů. Tento metajazyk pak mohou využít při zjišťování silných a slabých stránek vlastní práce a mohou si jeho prostřednictvím formulovat své vlastní učební cíle.

Většina formativního neboli diagnostického hodnocení probíhá na velmi podrobné úrovni určitých jazykových znalostí či dovedností, které právě byly nebo mají brzy být předmětem výuky. Seznamy jazykových jevů v oddíle 5.2 jsou pro praktické užití v diagnostickém hodnocení příliš všeobecné a je nutné je porovnat se specifikací příslušných úrovní ovládnutí (A2, B1 atd.) Tabulky deskriptorů různých aspektů jazykové kompetence u různých úrovní ovládnutí (kapitola 4) však mohou pomoci formulovat formativní zpětnou vazbu při hodnocení mluveného projevu.

Společné referenční úrovně mají význam především pro sumativní hodnocení. Projekt DIALANG však ukazuje, že i u tohoto typu hodnocení může mít zpětná vazba diagnostickou funkci, a tedy formativní účel.

9.3.6 Přímé hodnocení / nepřímé hodnocení

Přímé hodnocení je hodnocení právě probíhající činnosti studenta. Skupinka studentů například o něčem diskutuje, hodnotitel je pozoruje, porovná jejich výkony s tabulkou kritérií, zařadí je do příslušné kategorie a zhodnotí.

Nepřímé hodnocení probíhá naproti tomu formou testu, většinou písemného, u něhož se často hodnotí pomocné dovednosti.

Přímé hodnocení je z hlediska účinnosti omezeno na mluvení, psaní a interaktivní poslech, protože receptivní dovednosti nelze přímo pozorovat. Čtení lze například hodnotit

pouze nepřímo, a to tak, že studenti musejí porozumění textu prokázat plněním různých úkolů, například zaškrťováním možností, dokončováním vět a odpovídáním na otázky. Rozsah a ovládnutí jazykových prostředků lze testovat buď přímo – porovnáním s určitými kritérii, nebo nepřímo – posuzováním a interpretací odpovědí studenta na otázky testu. Klasickým přímým testem je pohovor/interview, klasickým nepřímým testem je cloze test (souvislý text, v němž je pravidelně vynecháno n-té slovo).

Z deskriptorů různých aspektů jazykové kompetence u jednotlivých úrovní ovládnutí v kapitole 5 lze čerpat při vytváření hodnotících kritérií pro přímé testy. Parametry uváděné v kapitole 4 mohou pomoci výběru témat, textů a testovacích úloh pro přímé testy produktivních dovedností a pro nepřímé testování poslechu a čtení. Parametry uváděné v kapitole 5 mohou navíc pomoci při výběru klíčových jazykových kompetencí, které mají být předmětem nepřímých testů jazykových znalostí, a při výběru klíčových pragmatických, sociolingvistických a jazykových kompetencí, na něž je nutné se zaměřit při formulování testových otázek u položkových testů čtyř základních jazykových dovedností.

9.3.7 Výkonové hodnocení / hodnocení znalostí

Výkonové hodnocení vyžaduje, aby student předvedl při přímém ústním nebo písemném testu vzorek jazyka.

Hodnocení znalostí spočívá v tom, že student odpovídá na otázky nejrůznějšího zaměření, aby doložil rozsah svých jazykových znalostí a úroveň ovládnutí jazyka.

Jazykovou kompetenci však bohužel nelze testovat přímo. Vždy lze hodnotit jen určitý počet výkonů, z nichž je nutné učinit si představu o celkovém ovládnutí jazyka. Úroveň ovládnutí jazyka lze považovat za kompetenci uvedenou do praxe. V tomto smyslu tedy všechny testy hodnotí pouze výkon (performanci), je však možné učinit si na jeho základě kvalitě příslušných kompetencí.

Je však pravda, že pohovor/interview vyžaduje náročnější „výkon“ než vyplňování mezer ve větách, což zase vyžaduje náročnější „výkon“ než pouhé zaškrťování správné výběrové odpovědi. Slovo „výkon“ zde tedy znamená produkci jazyka. Ve spojení „výkonnostní testy“ má užší význam; označuje jeden určitý výkon v (relativně) autentické situaci, často v pracovním nebo studijním prostředí. Přisoudíme-li termínu „hodnocení výkonu“ poněkud volnější význam, můžeme za výkonnostní testy považovat také procedury hodnocení ústního projevu, protože i zde se úroveň ovládnutí jazyka posuzuje na základě výkonu v různých stylistických typech promluv, které jsou v daném kontextu výuky považovány za důležité a odpovídající potřebám studentů. Některé testy hodnotí stejnou měrou jazykový výkon i znalost jazyka jako systému, jiné nikoliv.

Rozdílnost těchto typů hodnocení se velmi blíží rozdílu mezi přímým a nepřímým testováním, Rámec může být proto i zde využíván podobným způsobem. Specifikace jednotlivých úrovní ovládnutí (A1, B1, B2) vydané Radou Evropy navíc nabízejí podrobné informace o cílových jazykových znalostech v jazycích, pro něž jsou k dispozici.

9.3.8 Subjektivní hodnocení / objektivní hodnocení

Subjektivní hodnocení je hodnocení osobou zkoušejícího. Obvykle se tímto hodnocením míní posuzování kvality výkonu.

Objektivní hodnocení je hodnocení oproštěné od subjektivního pohledu. Obvykle má podobu nepřímého testu, kde pro každou otázku existuje jediná správná odpověď, jako je tomu například v testu s mnohonásobným výběrem odpovědí.

Otázka subjektivity a objektivity je však mnohem složitější.

Nepřímý test je často považován za „objektivní“, má-li hodnotící osoba k dispozici jednoznačný klíč, podle něhož se rozhoduje, zda určitou odpověď přijme nebo zamítne, a konečný výsledek pak vznikne součtem správných odpovědí. Některé typy testů tento proces dále rozvíjejí, takže pro každou otázku existuje jediná možná odpověď (např. testy s mnohonásobným výběrem odpovědí a c-testy¹, které vznikly z cloze testů právě za tímto účelem). Odpovědi bývají zpracovávány počítačově, aby se předešlo případným omylům při opravování. Takové testy jsou sice považovány za „objektivní“, jejich objektivita však není dokonalá. Někdo musel rozhodnout, že budou užity jen takové metody hodnocení, které umožňují lepší kontrolu nad průběhem a zpracováním testu (což bylo subjektivní rozhodnutí, s nímž ostatní nemuseli souhlasit). Někdo navrhl zaměření testu, někdo jiný se toto zaměření pokusil zpracovat do podoby položek testu a konečně někdo musel z množiny všech možných položek vybrat ty, které se v testu skutečně objeví. Protože všechna tato rozhodnutí zahrnují subjektivní prvek, měly by takové testy být označovány spíše jako objektivně hodnocené testy.

Přímé hodnocení výkonu obecně spočívá v udělování známek na základě úsudku. To znamená, že rozhodování o kvalitě studentova výkonu se děje subjektivně, po zvážení všech relevantních faktorů a na základě určitých pokynů, kritérií či zkušeností. Výhodou subjektivního přístupu je skutečnost, že jazyk a komunikace jsou složité jevy, které nelze atomizovat, protože nejsou pouhým součtem svých částí. Bývá obtížné přesně stanovit, co vlastně určitá položka testu testuje. Zaměřit testové položky na specifické aspekty kompetence nebo jazykového výkonu je proto mnohem složitější, než se zdá.

Přesto je v zájmu spravedlnosti nutné, aby bylo veškeré hodnocení co nejobjektivnější. Vliv osobního subjektivního úsudku na výběr obsahu testu a na hodnocení výkonu by měl být co nejvíce omezen, především jedná-li se o sumativní hodnocení, protože na základě výsledků takových testů často rozhoduje třetí strana o budoucnosti testovaných osob.

Subjektivní přístup k hodnocení pomohou omezit následující opatření, která zároveň posílí validitu a spolehlivost testu:

- vypracovat **specifikace obsahu hodnocení**, například na základě společného referenčního rámce pro daný kontext;
- zavést **týmové rozhodování** při výběru obsahu testu a při hodnocení výkonu;
- přijmout **standardní proceduru** stanovující, jak má být hodnocení studentů prováděno;
- vypracovat **jednoznačný klíč** pro hodnocení nepřímých testů a specificky definovaná kritéria pro přímé hodnocení;
- vyžadovat **posouzení více osobami** a/nebo použít **vážení různých faktorů**;
- **proškolit** zkoušející **v oblasti pokynů pro hodnocení**;
- kontrolovat kvalitu hodnocení (validitu, reliabilitu) formou **analýzy výsledků hodnocení**.

¹ tj. cloze testy s uvedením počátečního písmene slova, které má být doplněno do mezery.

Jak jsme se zmínili již na začátku této kapitoly, prvním krokem k omezení subjektivního postoje při hodnocení je ve všech stádiích hodnotícího procesu snaha o jednotný výklad daného konstruktů, jinými slovy budování společného referenčního rámce. Společný evropský referenční rámec se snaží nabídnout společnou základnu pro **specifikaci obsahu testů** a pro vytváření **specificky definovaných kritérií** pro přímé testování.

9.3.9 Hodnocení pomocí stupnice / hodnocení pomocí kontrolního seznamu

Hodnocení pomocí stupnice spočívá v tom, že zařazujeme studenta na určitý stupeň nebo do určitého pásma v rámci stupnice skládající se z několika stupňů či pásem.

Hodnocení pomocí kontrolního seznamu spočívá v tom, že student je hodnocen podle seznamu položek relevantních pro určitou úroveň ovládnutí nebo modelovou situaci.

Cílem „hodnocení pomocí stupnice“ je zařadit studenta do jednoho z řady pásem. Důraz je tedy vertikální: jak vysoko se student na stupnici dostane? Význam jednotlivých pásem či stupňů by měl být definován deskriptory dané stupnice. Stupnic může být několik, pro každou kategorii jedna, a lze je prezentovat společně formou tabulky nebo na samostatných listech. Je možné definovat samostatně každé pásmo či stupeň, popřípadě každé druhé, nebo pouze výkonnostní maximum, minimum a střed.

Alternativou stupnice je kontrolní seznam, jehož cílem je zjistit, zda hodnocený student obsáhl relevantní okruh znalostí. Důraz je tedy horizontální: v kolika položkách seznamu student vyhověl? Kontrolní seznam může mít dotazníkovou formu sloupce položek, lze ho však prezentovat také v jiné podobě, například ve tvaru kruhu. Hodnocení pak zní buď Ano/Ne nebo je odstupňované v rozmezí několika kroků (např. 0–4). Je žádoucí, aby pro každý krok existoval určitý požadavek a také definice, jak tento požadavek interpretovat.

Protože modelové deskriptory představují nezávislá kritéria, která byla kalibrována vzhledem k příslušným úrovním ovládnutí, mohou být pomůckou *nejen* při sestavování kontrolních seznamů pro určitou úroveň ovládnutí, jak je tomu v určitých verzích Jazykového portfolia, *ale* také pro vytváření hodnotících stupnic či tabulek, které pokrývají všechny relevantní úrovně ovládnutí, jak ukazují tabulka 2 pro sebehodnocení a tabulka 3 pro hodnocení zkoušejícím v kapitole 3.

9.3.10 Hodnocení na základě dojmu / hodnocení na základě řízeného úsudku

Hodnocení na základě dojmu je ryze subjektivním rozhodováním na základě sledování výkonu studenta v hodině, a to bez specifických kritérií pro specifické hodnocení.

Hodnocení na základě řízeného úsudku spočívá v tom, že subjektivní dojem zkoušejícího usměrňuje a doplňuje vědomé hodnocení podle specifických kritérií.

Slovo dojem zde znamená, že učitel či student hodnotí čistě na základě svého názoru na práci v hodině, plnění domácích úkolů atd. Mnoho forem subjektivního hodnocení, především ty využívané při průběžném hodnocení, spočívá v hodnocení dojmu, který si zkoušející, jenž daného studenta v ideálním případě po určitou dobu cíleně pozoroval, vybaví v paměti. Na tomto základě funguje velmi mnoho školských systémů.

Termín „řízený úsudek“ popisuje situaci, kdy je za pomoci hodnotícího přístupu dojem zformován do podoby uvážlivého úsudku. Takový přístup vyžaduje, a) aby měla hodnotící

činnost určitou proceduru, anebo b) soubor jasných kritérií, s jehož pomocí lze rozlišovat mezi různým bodovým nebo známkovým hodnocením, a c) zaškolení ve standardizovaném hodnocení. Výhodou řízeného přístupu k rozhodování je, že vytvoří-li si skupina hodnotitelů společný referenční rámec, radikálně se zvýší shoda v hodnocení ze strany jednotlivých členů. Platí to především tehdy, jsou-li k dispozici společně vypracované standardy v podobě vzorků výkonu a jsou-li přesně stanoveny souvislosti s jinými systémy hodnocení. Velký význam řízeného hodnocení podtrhují i výzkumy na poli mnohých disciplín, které opakovaně prokázaly, že u nezaškolených hodnotitelů mohou být rozdíly v přísnosti hodnocení ve stejném rozpětí, jako jsou schopnosti studentů, takže výsledky jsou téměř zcela nahodilé.

Stupnice deskriptorů pro jednotlivé Společné referenční úrovně mohou pomoci při přípravě souboru jasných kritérií zmiňovaných v bodě (b), nebo k vyjádření standardních požadavků u již existujících kritérií prostřednictvím společných úrovní. V budoucnu mohou být k dispozici standardní vzorky výkonu pro různé úrovně ovládání, což přispěje k výcviku ve standardizaci hodnocení.

9.3.11 Holistické hodnocení / analytické hodnocení

Holistické hodnocení spočívá v globálním, celkovém posouzení výkonu. Hodnotitel zvažuje různé aspekty výkonu intuitivně.

Analytické hodnocení zvažuje různé aspekty výkonu odděleně.

Toto dělení může mít dvojí podobu: a) vzhledem k tomu, co je sledováno, b) vzhledem ke způsobu zařazování do pásem nebo přidělování známek či bodů. Některé systémy kombinují analytický přístup na jedné úrovni s holistickým přístupem na úrovni druhé.

- a) Co se hodnotí: některé přístupy hodnotí globální kategorii, například „mluvení“ nebo „interakci“ a výsledek má podobu určitého počtu bodů nebo známky. Jiné, analytičtější přístupy vyžadují, aby hodnotitel posoudil jednotlivé nezávislé aspekty výkonu odděleně. Další přístupy požadují, aby si hodnotitel poznamenal celkový dojem, analyzoval ho podle různých kategorií a posléze dospěl k uvážlivému holistickému úsudku. Výhodou analytického přístupu s oddělenými kategoriemi je to, že vede hodnotitele k pečlivému sledování výkonu. Poskytuje potřebný metajazyk pro diskusi mezi jednotlivými zkoušejícími a pro poskytnutí výsledků studentům. Nevýhodou je, že pro hodnotitele je prokazatelně obtížné soustředit se na jednotlivé kategorie a nenechat se ovlivnit celkovým dojmem. Kromě toho, hodnotí-li se více než 4–5 kategorií, kognitivní kapacita hodnotitelů je přetížena.
- b) Výpočet výsledku: některé přístupy porovnávají celkový výkon s deskriptory na hodnotící stupnici, která je buď holistická (jediná globální stupnice), nebo analytická (samostatné stupnice pro 3–6 kategorií v tabulce). Tyto přístupy tedy nevyžadují matematické zpracování výsledků.

Výsledky mají podobu jediného čísla nebo „telefonního čísla“, které pokrývá všechny kategorie. Jiné, analytičtější přístupy vyžadují, aby byla určitá skupina položek obodována samostatně, bodová hodnocení se sečtou a výsledné skóre je případně převedeno na známku. Pro tento přístup je charakteristické, že různé kategorie jsou různě významné a nejsou tedy hodnoceny stejným počtem bodů.

Tabulky 2 a 3 v kapitole 3 jsou příkladem *analytických* stupnic kritérií pro sebehodnocení a pro hodnocení zkoušejícím používaných při holistickém přístupu k hodnocení (výkon studenta je porovnán s příslušnými definicemi a na tomto základě se rozhodne o konečném výsledku).

9.3.12 Hodnocení kategoriální / hodnocení souborné

Kategoriální hodnocení spočívá v hodnocení jediné úlohy (která může mít samozřejmě několik částí, aby pokrývala různé druhy promluv, jak se o tom hovoří v oddíle 9.2.1). Výkon se posuzuje podle kategorií v hodnotící tabulce, což je analytický přístup popsáný v oddíle 9.3.11.

Souborné hodnocení spočívá v hodnocení řady samostatných úloh (často hraní rolí s jinými studenty nebo s učitelem), které jsou hodnoceny jednoduchou celkovou známkou na definované stupnici – např. 0–3 nebo 1–4.

Souborné hodnocení je jedním ze způsobů, jak předejít u kategoriálního hodnocení tomu, že výsledky v jedné kategorii ovlivňují výsledky v jiných kategoriích. Na nižších úrovních výuky je kladen důraz především na úspěšné splnění jednotlivých úloh a cílem je vyplnit kontrolní seznam toho, co student ovládá, a to na základě učitelova či studentova hodnocení aktuálního výkonu, nikoliv pouze na základě dojmu. Na vyšších úrovních mohou být úlohy zaměřeny na různé aspekty ovládání jazyka v rámci výkonu studenta. Výsledky pak mají podobu výkonnostního profilu.

Ze stupnic pro různé kategorie jazykové kompetence doprovázejících text v kapitole 5 lze čerpat při přípravě kritérií pro kategoriální hodnocení. Protože se hodnotitelé mohou soustředit pouze na malý počet kategorií, je nutné hledat určitý kompromis. Podrobné rozpracování relevantních typů komunikativních jazykových činností v oddíle 4.4 a výčet různých typů funkční kompetence v oddíle 5.2.3.2 mohou pomoci při výběru vhodných úloh pro souborné hodnocení.

9.3.13 Hodnocení jinou osobou / sebehodnocení

Hodnocení jinou osobou znamená, že studenta hodnotí učitel nebo zkoušející.

Sebehodnocení znamená, že své ovládání jazyka posuzuje sám student.

Studenti se mohou aktivně podílet na mnoha metodách hodnocení, o nichž již byla řeč. Výzkumy naznačují, že není-li v sázce příliš mnoho (např. přijetí do kursu), může být sebehodnocení užitečným doplňkem jazykových testů a hodnocení učitelem. Přesnost sebehodnocení se zvyšuje, a) je-li hodnocení podloženo jasnými deskriptory, které vyjadřují standardní požadavky na ovládání jazyka, a b) vychází-li hodnocení z určité činnosti. Tato činnost sama může být testovaným úkolem. Hodnocení se také pravděpodobně zpřesní, budou-li studenti alespoň částečně zaškoleni. Takovéto strukturované sebehodnocení pak může dosahovat korelace mezi hodnocením učitelů a výsledky testů, která se rovná běžně udávané korelaci (míře souběžné validity) mezi jednotlivými učiteli, mezi testy a mezi hodnocením učitelů a testy.

Hlavní význam sebehodnocení však spočívá v tom, že slouží jako nástroj ke zvyšování motivace a zodpovědnosti studentů. Pomáhá jim poznat vlastní přednosti a nedostatky a uvědomit si, na co mají více zaměřit svou učební činnost.

Uživatelé Rámce mohou zvážit a, kde je to vhodné, uvést:

- *které z výše popsanych typů hodnocení:*
 - *lépe vyhovují potřebám studentů v jejich systému;*
 - *jsou vhodnější a praktičtější vzhledem ke zvyklostem jejich vzdělávacího systému;*
 - *svým zpětným účinkem na učitele lépe stimulují jejich další odborný růst;*
- *nakolik je v jejich systému zastoupeno didaktické hodnocení (zaměřené na školu a na studium) a hodnocení celkového ovládnutí jazyka (zaměřené na reálný život a celkový výsledek studia), jak se tyto dva typy hodnocení vzájemně doplňují a nakolik se vedle jazykových znalostí hodnotí také komunikativní výkon studenta;*
- *nakolik jsou studijní výsledky hodnoceny podle jasných standardních požadavků a kritérií (kriteriální hodnocení) a nakolik se známky a závěrečná hodnocení přidělují na základě výkonnostního pořadí studentů ve třídě (normativní hodnocení);*
- *nakolik jsou učitelé:*
 - *obeznámeni se standardním požadavky (společné deskriptory, vzorky výkonu studentů),*
 - *vedeni k tomu, aby se seznamovali s nejrůznějšími metodami hodnocení,*
 - *školeni v metodách hodnocení a v interpretaci výkonu;*
- *nakolik žádoucí a praktické by bylo vypracovat na základě vhodných standardních požadavků a definovaných kritérií integrovaný přístup k průběžnému hodnocení práce v hodinách a k hodnocení v určitých termínech;*
- *nakolik žádoucí a praktické by bylo zapojit studenty do sebehodnocení podle jasných deskriptorů testových úloh a deskriptorů různých aspektů jednotlivých úrovní ovládnutí jazyka a využít tyto deskriptory například při souborném hodnocení;*
- *nakolik se v kontextu jejich systému uplatní specifikace a stupnice, které nabízí Společný evropský rámec, a jak mohou být doplněny či rozpracovány.*

Tabulky 2 a 3 v kapitole 3 představují dvě verze hodnotících tabulek, první je určena pro sebehodnocení, druhá pro hodnocení zkoušejícím. Liší se – kromě čistě verbálního rozdílu mezi „*dokážu ...*“ a „*dokáže ...*“ – především v tom, že zatímco tabulka 2 je zaměřená na komunikativní činnosti, tabulka 3 se zaměřuje na generické aspekty kompetence, které se projevují v každém mluveném projevu. Tabulku 3, verzi pro sebehodnocení, si lze snadno představit i v mírně zjednodušené podobě. Zkušenost naznačuje, že přinejmenším dospělí studenti jsou schopni tímto způsobem kvalitativně posoudit stupeň své kompetence.

9.4 Praktický přístup k hodnocení a metasystém

Stupnice doprovázející text kapitol 4 a 5 jsou příkladem zjednodušeného souboru kategorií, jejichž podrobnější schéma je popsáno v samotném textu. V praxi není při hodnocení cílem, aby každý hodnotitel použil pro každou úroveň ovládnutí vždy všechny stupnice. Pro hodnotitele je obtížné soustředit se na větší počet kategorií a kromě toho nemusí být uvedené rozpětí stupňů vhodné pro každý kontext. Proto má soubor stupnic sloužit spíše jako referenční zdroj.

Každý praktický systém hodnocení, ať je založen na jakémkoliv přístupu, musí zredukovat počet kritérií, která se nabízejí, na zvládnutelné číslo. Je prověřeno, že více než 4 nebo 5 kategorií již může přetěžovat vnímání a že 7 kategorií je z psychologického hlediska maximum. Proto je nutné vybírat. Modelové stupnice pro hodnocení ústního projevu, považujeme-li za kvalitativní aspekt komunikace i schopnost interakce, zahrnují 12 kvalitativních kategorií:

- *strategie ujímání se slova,*
- *strategie spolupráce,*
- *žádost o objasnění,*
- *plynulost promluvy,*
- *přízřusobivost,*
- *logická uspořádanost promluvy,*
- *rozvoj tematických složek,*
- *přesnost,*
- *sociolingvistická kompetence,*
- *obecný rozsah jazykových prostředků,*
- *rozsah slovní zásoby,*
- *gramatická správnost,*
- *ovládání slovní zásoby,*
- *fonologická správnost.*

Je zřejmé, že deskriptory mnohých z těchto kategorií by mohly být součástí všeobecného kontrolního seznamu, 12 kategorií je však příliš mnoho pro jakékoliv hodnocení výkonu. Každý praktický přístup k hodnocení bude tedy tento seznam zohledňovat selektivně. Jednotlivé položky je nutné zkombinovat a přejmenovat a celý seznam zredukovat do podoby méně početného souboru hodnotících kritérií přiměřených potřebám studentů, povaze hodnocených úloh a pedagogickému stylu daného systému. Nově vytvořená kritéria mohou mít všechna stejnou závažnost, nebo budou některá z nich v kontextu dané úlohy považována za důležitější než jiná.

Následující čtyři příklady ukazují, jak lze postupovat. První tři stručně nastiňují, jakým způsobem jsou kategorie využívány jako testovací kritéria v již fungujících systémech hodnocení. Čtvrtý příklad ukazuje, jak byly deskriptory stupnic uvedených v Rámci sloučeny a přeformulovány, aby mohla vzniknout hodnotící tabulka pro určitou příležitost a účel.

Příklad 1

Cambridge Certificate in Advanced English (CAE), Test 5: Kritéria pro hodnocení (1991)

Testovací kritéria	Modelové stupnice	Jiné kategorie
plynulost promluvy	plynulost promluvy	
správnost a rozsah	obecný rozsah rozsah slovní zásoby gramatická přesnost ovládání slovní zásoby	
výslovnost	fonologická správnost	
zvládnutí úlohy	uspořádanost promluvy sociolingvistická přiměřenost	splnění úlohy potřeba pomoci partnera v komunikaci
interaktivní komunikace	strategie ujímání se slova strategie spolupráce rozvoj tematických složek	rozsah a přirozenost účasti v komunikaci

Poznámka k jiným kategoriím: Modelové stupnice a kritéria splnění úlohy lze vyhledat podle druhu činnosti v oddíle *Komunikativní činnosti*. Kategorie *rozsah a přirozenost účasti v komunikaci* je v těchto stupnicích zařazena pod *plynulost projevu*. Pokus vypracovat a kalibrovat deskriptory pro kategorii *potřeba pomoci partnera v komunikaci* skončil nezdarem.

Příklad 2

International Certificate Conference (ICC): Zkouška z anglického jazyka pro obchodní účely, Test 2: Obchodní konverzace (1987)

Kritéria testu	Modelové stupnice	Jiné kategorie
stupnice 1 (bez názvu)	sociolingvistická přiměřenost gramatická správnost ovládání slovní zásoby	splnění úlohy
stupnice 2 (využití prostředků promluvy k zahájení a udržování konverzace)	strategie ujímání se slova strategie spolupráce sociolingvistická přiměřenost	

Příklad 3

Eurocentres – hodnocení interakce v malé skupině (RADIO) (1987)

Kritéria testu	Modelové stupnice	Jiné kategorie
rozsah	obecný rozsah rozsah slovní zásoby	
správnost	gramatická přesnost ovládání slovní zásoby sociolingvistická přiměřenost	
podání (promluvy)	plynulost promluvy fonologická správnost	
interakce	strategie ujímání se slova strategie spolupráce	

Příklad 4

Švýcarská Národní rada pro výzkum: Hodnocení studentů na základě videonahrávek

Kontext: Jak vysvětluje Příloha A, jeden ze švýcarských výzkumných projektů se zabýval sestavováním vzorových deskriptorů do stupnic. Na závěr projektu byli zúčastnění učitelé pozváni na konferenci, kde prezentovali výsledky svého výzkumu a zahájili projekt experimentálního využití Evropského jazykového portfolio ve Švýcarsku. Konference se zabývala mimo jiné dvěma následujícími tématy: a) potřebou uvést kontrolní seznamy pro průběžné hodnocení a pro sebehodnocení do souladu s celkovým rámcem; b) otázkou, jak lze stupnice deskriptorů, které jsou výsledkem projektu, různým způsobem využívat při hodnocení. Součástí diskuse bylo promítnutí videonahrávek výkonu několika studentů, které byly posléze hodnoceny podle tabulky 3 v kapitole 3. Tato tabulka obsahuje vybrané vzorové deskriptory ve sloučené a upravené podobě.

Kritéria testu	Modelové stupnice	Jiné kategorie
rozsah	obecný rozsah rozsah slovní zásoby	
správnost	gramatická správnost ovládání slovní zásoby	
plynulost projevu	plynulost promluvy	
interakce	celková úroveň interakce strategie ujímání se slova spolupráce	
uspořádanost promluvy	uspořádanost promluvy	

Různé testovací systémy s různými studenty a v různých kontextech hodnotící kategorie zjednodušují, vybírají a kombinují dle potřeb různých druhů hodnocení. Uvedený seznam 12 kategorií, ačkoliv se může zdát dlouhý, pravděpodobně nepokrývá všechny varianty, které si lidé volí, a kdyby měl být vyčerpávající, bylo by ho nutné rozšířit.

Uživatelé Rámce mohou zvážit a, kde je to vhodné, uvést:

- *jak jsou v rámci pracovního přístupu jejich systému teoretické kategorie zjednodušeny;*
- *nakolik se hlavní faktory, které jsou v jejich systému používány jako hodnotící kritéria, mohou stát součástí souboru kategorií uvedených v kapitole 5, jejichž modelové stupnice jsou uvedeny v příloze. Předpokládá to ovšem, že podrobnější rozpracování specifických oblastí užívání jazyka bude respektováno.*