

Pregledni članak
UDK 81'33:371.2/4

Primljen 26.8.2012.
Prihvaćen 16.10.2012.

JEZIČNO ISPITIVANJE DANAS

Vesna Bagarić Medve*

Filozofski fakultet u Osijeku

U radu pod naslovom *Testovi u stranim jezicima* objavljenom u ovome časopisu 1971. godine, Mirjana Vilke daje pregled vrsta ispita, svrhe ispita, sadržaja ispita i načina ispitivanja i procjenjivanja u nastavi stranoga jezika. Od 1970-ih do danas došlo je do značajnoga napretka u razumijevanju toga što ispitujemo i kako ispitujemo, tj. u odnosu na određenje konstrukta i metoda ispitivanja. Postojeći pojmovi su preciznije određeni, a predložen je i niz novih koncepata kao rezultat istraživanja koja su dovela do brzoga razvoja jezičnoga ispitivanja¹ kao znanstvene poddiscipline primijenjene lingvistike. U radu se nakon prikaza pristupa jezičnome ispitivanju od sredine 20. stoljeća do danas, vrsta ispitivanja i ispita, iznose trenutne teorijske i empirijske spoznaje u vezi sa svrhom i sadržajem ispita, problematici razvoja zadataka i instrumenata za procjenu, provedbi i analizi ispita.

* Vesna Bagarić, Filozofski fakultet u Osijeku, vbagaric@ffos.hr

¹ U govoru o ovoj disciplini, u literaturi se u puno različitih ili pak identičnih značenja upotrebljavaju izrazi jezično ispitivanje (engl. *language testing*), jezično procjenjivanje (engl. *language assessment*) ili jezično vrjednovanje (engl. *language evaluation*). Iz toga proizlazi da su ključni pojmovi u toj disciplini ispitivanje, procjenjivanje i vrjednovanje. Pojam vrjednovanje se, kako ističu Jelaska i Cvikić (2008) te Škarica Mital (2011), značenjski velikim dijelom preklapa s pojmom procjenjivanja i često upotrebljava istoznačno. U literaturi se na engleskome jeziku, međutim, pojam procjenjivanja upotrebljava još u značenju ispitivanja uz primjenu pismenoga i/ili usmenoga ispita te u značenju procjenjivanja, odnosno donošenja suda o jezičnim znanjima i sposobnostima korisnika jezika. Složenost razumijevanja pojmova postaje veća kada se uzme u obzir da sama disciplina koja se svim tim aspektima bavi nosi naziv jezično ispitivanje (engl. *language testing*). Prema Brown i Abeywickrama (2010), koji su semantički odnos između tih pojmova prikazali i zorno, vrjednovanje se odnosi i na procjenjivanje i na ispitivanje, a procjenjivanje može, a ne mora uključivati ispitivanje, ili bolje rečeno, formalne ispite. U ovom će radu najviše riječi biti o ispitivanju i ispitima. Kada to zahtijeva kontekst izvornoga teksta u kojem se određeni pojam upotrebljavao, posebno će se istaknuti procjenjivanje i/ili vrjednovanje u nastojanju da se, u što je moguće većoj mjeri, smanji moguća dvosmislenost pojmova.

Završni dio rada donosi pregled tematskih područja u okviru kojih se u budućnosti može očekivati intenzivnija istraživačka aktivnost.

Ključne riječi: ispitivanje, procjenjivanje, komunikacijsko jezično ispitivanje, procjenjivanje performanse, standardizirano ispitivanje, ispitivanje u razrednoj sredini

1. POVIJESNI PREGLED

U povijesti jezičnoga ispitivanja razlikuje se nekoliko razdoblja koja se preklapaju s promjenama u razumijevanju prirode jezika, toga kako se jezik usvaja, upotrebljava i poučava (usp. Weir, 1990; Ingram, 1985; Spolsky, 1995; Brown i Abeywickrama, 2010). U prvomu, **tradicionalnomu ili predznanstvenomu** razdoblju, koje je trajalo do sredine 20. stoljeća, ispiti su se temeljili na mjerenju gramatičkoga znanja i sposobnosti prevođenja. S pojavom biheviorističkih teorija jezičnoga učenja i strukturalističkih lingvističkih teorija započelo je **psihometrijsko-strukturalno ili znanstveno** razdoblje u kojemu su se autori ispita usredotočili na individualne jezične elemente poput fonoloških, gramatičkih i leksičkih, ponajviše stoga što se vjerovalo da je znanje pojedinih elemenata jednako ukupnomu znanju jezika. Atomistički način gledanja na jezik odgovarao je psihometrijskim metodama koje su se primjenjivale i dominirale jezičnim testiranjem u tom razdoblju. **Integrativno-sociolingvističko** razdoblje jezičnoga testiranja, koje je uslijedilo u 1970-im, obilježila je spoznaja o međudodnosu i međuzavisnosti komponencijalnih dijelova jezika i svakoga posebnog dijela s lingvističkim i nelingvističkim kontekstom. U skladu s tvrdnjom da je komunikacijska kompetencija unitarna, tj. da se sastoji od jednoga globalnog elementa (usp. Oller 1979), te se kao takva ne može ispitati kao odvojeni niz znanja i sposobnosti, u tom su razdoblju postali popularni integrativni testovi poput diktata i *cloze* testa. Istovremeno se, međutim, naglašavala i važnost autentične komunikacije pa se već krajem 1970-ih naglasak prenio na ispitivanje jezičnoga ponašanja ispitanika u stvarnim ili približno stvarnim situacijama jezične uporabe. Zbog usredotočenja na sve komunikacijske vještine, na ono što ispitanik može napraviti s jezikom u autentičnoj jezičnoj uporabi, na zadatke koje može izvršiti i kako ih može izvršiti, drugim riječima, zbog

usredotočenja na cjelokupno jezično ponašanje ispitanika i njegovu produktivnu sposobnost, Ingram (1985) je to razdoblje u jezičnomu testiranju nazvao bihevioralnim ili autentičnim pristupom. Prema Ingramu, taj je pristup postao popularan s pojavom funkcionalnoga gledišta na jezik, funkcionalno-pojmovnoga ili komunikacijskoga pristupa jezičnome poučavanju. Primjena komunikacijskih ispita – ispita kvalitete ispitanikove jezične performanse, a ne kvantitete, tj. broja točno riješenih zadataka u ispitu ključna je značajka toga razdoblja koje traje i danas pod nazivom **komunikacijsko jezično ispitivanje**. Prema Brown i Abeywickrama (2010) pravi je početak toga razdoblja označila pojava teorijskoga okvira komunikacijske kompetencije koji su predložili Canale i Swain (1980) i Canale (1983), a čija je svrha bila, između ostaloga, potaknuti izradu komunikacijskih jezičnih zadataka u kojima bi se performansa ispitanika podudarala s jezičnom uporabom u stvarnoj komunikacijskoj situaciji. Izrada takvih autentičnih zadataka, odnosno zadataka koji odgovaraju određenoj ciljnoj jezičnoj uporabi, postao je pravi izazov autorima zadataka. Danas takvi zadatci najvećim dijelom odgovaraju poluotvorenim i otvorenim tipovima zadataka (npr. prošireni odgovor, intervju, rasprava, pričanje priče, vođeni sastavak, odgovor na pismo, esej i sl.) kojima se ispituje i procjenjuje usmena i pisana jezična proizvodnja ispitanika, često u kombinaciji sa slušanjem ili čitanjem. Zbog najnovijih tendencija u primjeni takvih zadataka, na današnje se razdoblje u jezičnome ispitivanju referira i u smislu vrjednovanja temeljenoga na performansi ispitanika (engl. *performance-based assessment*, *performance assessment*) (usp. Norris *et al.*, 1998; Bachman, 2002; Leung i Lewkowicz, 2006). Prema Norris *et al.* (1998), tri su glavna obilježja takve vrste vrjednovanja: a) ispitanici moraju izvesti zadatak, b) zadatak treba biti autentičan u najvećoj mogućoj mjeri i c) izvedbu zadatka trebaju procijeniti obučeni ocjenjivači.

2. VRSTE ISPITIVANJA

U knjigama o jezičnome ispitivanju i procjenjivanju koje imaju priručnički karakter (vidi npr. Bachman i Palmer, 1996; Bachman i Palmer, 2010; Fulcher, 2010; Brown i Abeywickrama, 2010) danas se posebno naglašava razlika između standardiziranoga ispitivanja (engl.

standardized testing) i ispitivanja u razrednoj sredini (engl. *classroom testing*)².

Standardizirano ispitivanje, kao što proizlazi iz njegova naziva, pretpostavlja postojanje određenih standarda koji su po svojoj prirodi dvojaki - odnose se na sadržaj ispitivanja i metodologiju ispitivanja. Sadržaj ispitivanja izvodi se iz određenoga teorijskoga modela komunikacijske kompetencije i/ili jezičnoga umijeća koji je operacionaliziran kao niz jasno definiranih kompetencija, odnosno znanja i sposobnosti, ili iz bihevioralnih ljestvica jezičnoga umijeća koje sadrže, primjerice, dokumenti na internacionalnoj razini poput ZEROJ-a (2005), koji je dobro prihvaćen u europskome obrazovnome kontekstu na svim razinama obrazovanja, ili ACTFL-a (1986) u SAD-u. I nastavni planovi i programi na nacionalnoj razini, ako su koncipirani kao standard, mogu postati polazište za određivanje obrazovnih ishoda koji će se ispitivati. Metodologija ispitivanja uključuje niz postupaka za mjerenje opisanih kompetencija, a pri njihovoj se primjeni rukovodi načelima ispitivanja. Navedene činjenice upućuju na to da je standardizirano ispitivanje formalni oblik ispitivanja za koji je karakteristično da se, u vremenskome i organizacijskome pogledu, provodi planski i sustavno, pomoću metrijski provjerenih instrumenata i razrađenih okvira za procjenjivanje i tumačenje rezultata. Kao takav on pruža svim sudionicima u procesu ispitivanja - učenicima, nastavnicima, roditeljima, te obrazovnim i znanstvenim ustanovama - uvid u stečena jezična znanja i sposobnosti na kraju određenih obrazovnih razdoblja. Zbog potonjega obilježja može se reći da se pri standardiziranome ispitivanju provodi sumativno procjenjivanje koje je orijentirano na kraj procesa učenja i na krajnji rezultat toga procesa. Nadalje, u standardiziranome su ispitivanju i danas još uvijek dosta zastupljeni tradicionalni oblici ispitivanja, primjerice pismeni ispiti u kojima prevladavaju zadatci višestrukoga izbora, zadatci dopunjavanja ili alternativnoga izbora, a koji osiguravaju viši stupanj objektivnosti i praktičnosti u provedbi ispita, ali i analizi rezultata. Konačno,

² Uz standardizirano ispitivanje, u literaturi se još češće i većinom istoznačno upotrebljava izraz standardizirano vrjednovanje (engl. *standardized assessment*) te uz ispitivanje u razrednoj sredini izraz vrjednovanje u razrednoj sredini (engl. *classroom assessment*). Potonji se izraz paralelno upotrebljava uz sljedeće izraze: edukacijsko vrjednovanje (engl. *educational assessment*), neformalno vrjednovanje (engl. *informal assessment*), formativno vrjednovanje (engl. *formative assessment*), nastavnikovo vrjednovanje (engl. *teacher assessment*), alternativno vrjednovanje (engl. *alternative assessment*) i drugi.

standardizirano se ispitivanje ponajviše veže uz normativni pristup procjenjivanju u okviru kojega rezultat jednoga ispitanika dobiva značenje tek u usporedbi s rezultatima drugih ispitanika, iako je danas, u razdoblju komunikacijskoga jezičnoga ispitivanja i procjenjivanja usmjerenoga na performansu ispitanika, kriterijski pristup procjenjivanju, tj. procjenjivanje stupnja usvojenosti znanja i sposobnosti s obzirom na unaprijed zadani kriterij, sve prisutniji u standardiziranome ispitivanju.

Uz standardizirane ispite vezuje se sintagma „ispit visokoga rizika“. Upravo se zbog toga danas mogu sve češće čuti kritike standardiziranoga oblika ispitivanja. Primjerice, napredovanje na obrazovnoj ljestvici može ovisiti o uspjehu na jednome standardiziranome ispitu pa mu zbog toga ispitanici, ali i svi drugi sudionici u procesu ispitivanja, pristupaju pod velikim pritiskom i s osjećajem anksioznosti (Lin, 2002; Huhta *et al.*, 2006); u standardiziranome su ispitu svi zadatci isti za svakoga ispitanika, zbog čega se zanemaruju individualne razlike među ispitanicima (Brown i Abeywickrama 2010.); budući da su nastavnicima i ispitanicima unaprijed poznati sadržaji i metode ispitivanja, moguća je pojava isključivoga učenja/poučavanja za ispit što može dovesti do negativnoga povratnoga učinka ispita (Shohamy 1997, Phelps 2008); standardizirani ispit se ponekad iz političkih razloga upotrebljava u svrhu za koju nije razvijen (Shohamy 2001).

Ipak, standardizirani ispiti imaju jednu vrlo važnu pozitivnu stranu, a to je da u najvećoj mogućoj mjeri udovoljavaju načelima ispitivanja: valjani su i pouzdani, objektivni i praktični, a mogu imati i dobar povratni učinak na nastavnu praksu (Brown i Abeywickrama 2010). Da bi pozitivne karakteristike standardiziranoga ispitivanja još više došle do izražaja, predlaže se sustavno provođenje postupka postavljanja standarda (Kantarcioglu i Papageorgiu 2011), pridavanje veće uloge nastavnicima u razvoju standardiziranih ispita (Brindley 2001), i provjere njihove valjanosti u kontekstima u kojima se primjenjuju (Weir 2005, O'Sullivan i Weir 2011).

Budući da nije vremenski određeno ili ograničeno, nego je stalni, sastavni dio procesa poučavanja određene skupine učenika, ispitivanje u razrednoj sredini ima više neformalni karakter i obilježje formativnoga vrjednovanja. Svrha mu je učenika osvijestiti, a nastavniku pružiti

povratnu informaciju o postojećem stanju u vezi s razinom ovladanosti obrađenoga jezičnoga sadržaja, pa kao takav oblikuje, potiče i usmjerava daljnji proces učenja i poučavanja prema krajnjem obrazovnom ishodu³. U vezi s ispitivanjem u razredu danas se osobito naglašava vrijednost alternativnih oblika vrjednovanja kao što su, na primjer, jezični portfolio, dnevници, rasprave, dramatizacije, samoprocjenjivanje i suradničko procjenjivanje⁴. Primjenom tih oblika podržava se pristup poučavanju koji je usmjeren na učenika i njegovu aktivnu ulogu u procesu učenja jezika te uporaba jezika u stvarnim komunikacijskim situacijama što je, kao što smo već naglasili, i glavno obilježje komunikacijskoga jezičnoga ispitivanja. Navedeni oblici vrjednovanja zahtijevaju kriterijski pristup procjenjivanju, kojega nastavnici i inače najviše primjenjuju pri ispitivanju u razrednoj sredini. No, i normativni pristup procjenjivanju učeničkih znanja i sposobnosti, obilježje je ne samo standardiziranoga ispitivanja, nego i ispitivanja u razrednoj sredini.

Za razliku od standardiziranoga ispitivanja, ispitivanje u razrednom okruženju ima izrazito naglašenu pedagošku notu. Vjerojatno je i to razlog što je, s obzirom na broj priručnika posvećenih jezičnom ispitivanju i procjenjivanju u razrednoj sredini (npr. Cohen, 1994; Hughes, 2003; Brown i Abeywickrama, 2010 itd.), istraživanja s tom tematikom relativno malo. Uglavnom su posvećena analizi nastavnikovih postupaka u procesu ispitivanja i procjenjivanja (npr. Leung i Teasdale, 1997; Rea-Dickens, 2001) ili usporedbi nastavnikovih postupaka s drugim vrstama ili oblicima procjenjivanja kao što je, primjerice, samoprocjenjivanje i suradničko procjenjivanje (npr. Matsuno, 2009). U vezi s tim posebno bismo naglasili istraživanje koje je provela Škarica Mital (2011), a u kojemu je usporedila procese i ishode u standardiziranome ispitivanju i ispitivanju u razrednoj sredini s obzirom na engleski kao strani jezik. Zbog malo raznovrsnih istraživanja u području ispitivanja u razrednoj sredini, Hill i McNamara (2012) su nedavno predložili okvir za istraživanja u području ispitivanja i

³ Značajan prinos teorijskim i empirijskim spoznajama o vrijednosti formativnoga vrjednovanja u jezičnom ispitivanju dali su Black *et al.* (2003) pokrenuvši pokret „Vrjednovanje u svrhu učenja“ (engl. *Assessment for learning*). Formativno vrjednovanje nalazi se i u korijenu tzv. Dinamičkoga vrjednovanja (engl. *Dynamic assessment*) koje, međutim, polazi od pojavnosti koju je Vygotsky nazvao područje približnoga razvoja (engl. *zone of proximal development*), a razlikuje se od Vrjednovanja u svrhu učenja po tome što se učenika ne upoznaje nužno s krajnjim ishodom učenja, nego ga se u vidu posredničkoga djelovanja nastavnika upozorava na prazninu u slijedu usvajanja jezičnih oblika, čime se kod učenika potiče opažanje (usp. Lantolf i Poehner 2008).

⁴ Detaljniji pregled primjene oblika alternativnoga vrjednovanja može se naći u Law i Eckes (2007).

procjenjivanja u razrednoj sredini koji će zasigurno oživjeti istraživačku djelatnost u tom vrlo važnom području.

3. VRSTE ISPITA

U standardiziranome ispitivanju najčešće se razvijaju i primjenjuju ispit jezičnoga umijeća (engl. *proficiency tests*). Svrha im je procjena općega jezičnoga umijeća a izrađuju se s obzirom na određeni standard, dakle neovisno o vrsti prethodnoga obrazovanja potencijalnih ispitanika. Izrada kvalitetnih ispita jezičnoga umijeća je dugotrajan i skup proces, i to ne samo zbog složenosti postupaka koji se primjenjuju u njihovoj izradi nego i zbog velikoga broja stručnjaka koji te postupke provode. Zbog svoje prirode i kvalitete izrade vrlo su pogodni za uporabu u istraživačke svrhe i to ne samo u području jezičnoga ispitivanja nego i u području ovladavanja inim jezikom. Danas postoji veliki broj komercijalno razvijenih ispita jezičnoga umijeća za sve svjetske jezike, a osobito za engleski⁵.

U ispitivanju u razrednoj sredini nastavnici obično upotrebljavaju ispite postignuća (engl. *achievement tests*) i dijagnostičke ispite (engl. *diagnostic tests*). Svrha testova postignuća je procjena količine i kakvoće usvojenosti obrađenih sadržaja u određenome predmetu, odnosno ovladanosti ishodima učenja u tom predmetu nakon ograničenoga vremenskoga razdoblja (npr. na kraju nastavne jedinice, cjeline ili na kraju nastavnoga programa). Ispiti postignuća mogu poslužiti nastavniku i kao dijagnostički ispiti, a kojima je svrha otkrivanje aspekata jezičnoga znanja i sposobnosti učenika koji su nedostatno usvojeni. Da bi bili svrsishodni, dijagnostičkim se ispitima ispituju, prije svega, određena, a ne opća jezična znanja i sposobnosti. Na taj način nastavnik može dobiti jasniju sliku o tome kojim znanjima i sposobnostima učenik još treba ovladati u okviru određenoga područja jezičnih znanja i jezičnih vještina. Istraživanja razvoja i kvalitete dijagnostičkih ispita, za razliku od ispita jezičnoga umijeća, još su u povojima (Alderson 2005), a neka od istraživanja (vidi npr. Knoch 2009)

⁵ U Hrvatskoj se od poznatijih ispita jezičnoga umijeća u engleskome jeziku provode, npr. FCE (Cambridge First Certificate in English), CPE (Cambridge Certificate of Proficiency in English), TOEFL (Test of English as a Foreign Language), a u njemačkome jeziku DSD (Deutsches Sprachdiplom), Zertifikat Deutsch, Test DaF i mnogi drugi.

provedenih na široko prihvaćenim dijagnostičkim ispitima upućuju na nedostatke, prije svega u području valjanosti.

Neki ispiti jezičnoga umijeća i ispiti postignuća mogu se primjenjivati i kao razredbeni ispiti (engl. *placement tests*) sa svrhom smještanja ispitanika u određeni obrazovni program ili na određeni stupanj programa, a s obzirom na razinu njihova predznanja. Na taj bi se način ispitanik trebao naći u skupini u kojoj se obrađuju sadržaji koji su za njega optimalne težine, što bi mu potom omogućilo odgovarajući napredak. Da se ta svrha ispita ne postiže uvijek, dokazali su Green i Weir (2004) u istraživanju na primjeru probne verzije *Global Placement Test-a*, kojim se mjeri gramatičko znanje. Tim su istraživanjem navedeni autori upozorili na problematiku izrade razredbenih ispita i činjenicu da se općenito u istraživanjima vrlo malo pažnje posvećuje ovoj vrsti ispita, iako ti ispiti u određenim kontekstima njihove uporabe pripadaju kategoriji ispita visokoga rizika.

Kao što smo naglasili u uvodu ovoga rada, od kraja 1970-ih pa sve do danas traje razdoblje komunikacijskoga jezičnoga ispitivanja, a posljednjih godina i sve izraženije razdoblje procjenjivanja koje se temelji na performansi ispitanika⁶. Bez obzira o kojoj se vrsti ispitivanja i procjenjivanja radi, danas bi svi ispiti trebali imati predznak „komunikacijski“, budući da se komunikacijskim jezičnim ispitima mjeri ispitanikova sposobnost uporabe jezičnoga znanja u stvarnim komunikacijskim situacijama. Kao i u vrijeme kada su predloženi, tako se i danas vjeruje da na komunikacijskim jezičnim ispitima najbolje dolazi do izražaja prava priroda ispitanikove komunikacijske kompetencije.

S obzirom na tu svrhu, u literaturi (npr. Weir, 1990; Bachman, 1990; Skehan 1991) se navodi kako bi komunikacijski jezični ispiti, za razliku od tradicionalnih jezičnih ispita, trebali biti:

1. izravni, a to znači da zadatci odgovaraju u najvećoj mogućoj mjeri zahtjevima jezične uporabe u stvarnim lingvističkim, situacijskim, kulturalnim i afektivnim uvjetima;

⁶ Iako je pitanje na to je li moguće praćenjem performanse ispitanika dobiti uvid u njegova jezična znanja i sposobnosti i danas bez jasnoga odgovora, čini se da smo, a sudeći prema tendencijama u području izrade komunikacijskih jezičnih ispita, više nego ikada blizu potvrdnoga odgovora, odnosno priznanja da performansa odražava kompetenciju, ako ne izravno, a onda, barem, neizravno, te da je moguće izvesti zaključke o kompetenciji na temelju procjenjivanja performanse. Više o tome u Bagarić (2007).

2. integrativni, tj. da se u zadatcima traži uporaba različitih jezičnih znanja i sposobnosti u okviru različitih jezičnih djelatnosti;
3. kriterijski, a to znači da se performansa ispitanika procjenjuje s obzirom na unaprijed određena jezična znanja i sposobnosti koji se zadatkom ispituju i koji su jasno opisani za svaku razinu.

Pored tih osnovnih obilježja, razni teoretičari (npr. Bachman, 1990; Fulcher, 2000) naglašavaju kako bi komunikacijski jezični ispiti trebali imati i ove karakteristike:

- autentičnost, a to znači da bi ispitanici trebali moći prepoznati komunikacijsku svrhu u tekstovima i zadatcima; ona bi za njih trebala biti smislena i u skladu s njihovim potrebama, kako bi na nju mogli moći prikladno reagirati; tekstovi i zadatci ne bi trebali biti pojednostavljeni, barem ne u mjeri da izgube autentična obilježja; jezični i situacijski kontekst trebao bi varirati kako bi se ispitala sposobnost ispitanika da se nosi s različitim kontekstima jezične uporabe;
- uključenost performanse, a koja podrazumijeva da se zadatci u testu trebaju izvesti na način da ispitanici pri izvršenju zadatka trebaju, u pravom smislu riječi, upotrijebiti jezik, npr. u interakciji za koju je karakteristična, između ostalog, nepredvidiva i kreativna uporaba jezika;
- procjena ispitanikovih sposobnosti s obzirom na to je li ispitanik bio u stanju postići komunikacijski učinak, tj. je li komunikacija bila uspješna.
- oslanjanje na kvalitativan, odnosno indirektno kvantitativan način ocjenjivanja, budući da se kvalitativne ocjene mogu ili, nužno, moraju pretvoriti u numeričke ocjene.

Međutim, komunikacijski jezični ispiti - zbog težnje za njihovom lakom provedljivošću - u pravilu nemaju sva gore navedena obilježja u isto vrijeme. Često ne sadrže niti sva osnovna obilježja. Drugim riječima, rijetko su istodobno izravni, integrativni i kriterijski. Obično su, zapravo, više integrativni, tj. mjerila globalne jezične performanse, a manje zasebnih lingvističkih pojava, više su izravni, a manje neizravni, te više kriterijski, a manje normativni (Davis, 1988). Budući da se komunikacijski jezični ispiti mogu nalaziti na različitim mjestima na navedenim kontinuumima, vrlo su raznoliki. Neki su teoretičari (npr. Hulstijn, 1985) čvrstoga uvjerenja da bi se uz zadatke karakteristične za

izravno ispitivanje (npr. intervju), trebalo uključiti i zadatke koji su objektivniji, i to zbog mogućnosti variranja postupaka prikupljanja podataka (odabir između manje ili više strukturiranih postupaka) i postupaka procjenjivanja (odabir između subjektivnih i objektivnih postupaka). Time bi se, ujedno, spriječilo djelovanje jedne metode na rezultate. S druge strane, danas su sve usuglašenija mišljenja o potrebi izrade ispita koji bi sadržavao zadatke čiju izvedbu trebaju procijeniti obučeni ocjenjivači uz pomoć stručno razvijenih ljestvica za procjenjivanje. Norris *et al.* (1998: 9) vrlo jasno opisuju obilježja takvih ispita:

1. Zadatci se trebaju temeljiti na analizi potreba ispitanika, biti autentični, poticati ispitanike na interakciju, biti kontekstualizirani i kompleksni, integrirati vještine i sadržaj ispitivanja, biti prikladni u smislu njihova broja, vremena koje je potrebno za njihovo izvršenje i, općenito, učestalosti jezičnoga ispitivanja, te u skladu sa zadatcima s kojima su se ispitanici upoznali tijekom nastavnoga procesa.
2. Prikladnost ocjenjivača, u smislu da ih je dovoljan broj, da su stručni, upoznati s ljestvicom za procjenjivanje i obučeni za njezinu uporabu.
3. Prikladnost ljestvica za procjenjivanje, a to znači da sadrži kategorije koje proizlaze iz teorijskih postavki o učenju jezika i jezičnoga razvoja, da su te kategorije dobro opisane i da su u skladu sa standardima koji su ispitanicima jasni.
4. Kombinacija s drugim metodama, kao što su, na primjer, portfolio, samoprocjenjivanje, rasprave i druge alternativne metode, a radi povećanja pouzdanosti i valjanosti odluka koje će proizaći iz rezultata ispitivanja.

Iz svega navedenog se, dakle, može zaključiti da je pojam komunikacijskoga jezičnoga ispita relativan. Pretpostavljamo da je Bachman (1990) upravo zbog toga komunikacijski jezični ispit definirao općenito - kao ispit koji uključuje raspon sposobnosti i tipova zadataka koji su u skladu s aktualnim gledištem na komunikacijsku kompetenciju i komunikacijsku jezičnu uporabu. U toj je definiciji komunikacijskoga jezičnoga ispita Bachman upozorio na dvije važne stvari. Prvo, komunikacijskim bi se ispitom trebalo mjeriti više jezičnih sposobnosti, što je u skladu sa spoznajama o višedimenzionalnoj prirodi i strukturi

komunikacijske kompetencije. Drugo, određeni jezični aspekt se različitim načinima mjerenja i u različitim uvjetima mjerenja mjeri na različite načine. Stoga je bitno razmisliti koje aspekte jezičnih znanja i sposobnosti ispitivati i kojim se metodama komunikacijskoga jezičnoga ispitivanja može dobiti točan odraz ciljnih kompetencija.

4. METODOLOGIJA JEZIČNOGA ISPITIVANJA

Proces izrade ispita danas je izuzetno dobro razrađen i opisan, kako za primjenu u okviru standardiziranoga ispitivanja tako i za primjenu u ispitivanju u razrednoj sredini. U prvoj vrsti ispitivanja radi se o složenijem i dugotrajnijem procesu. Tako razvoj standardiziranoga ispita uključuje nekoliko etapa u okviru kojih se provodi sljedeće: određenje svrhe ispita, definiranje konstrukta, odabir sadržaja ispita, oblikovanje ispitne specifikacije ili ispitnoga kataloga, izrada zadataka, probno ispitivanje i analiza zadataka uz primjenu kvalitativnih i kvantitativnih metoda, izrada kriterija procjenjivanja i ljestvica za procjenjivanje, analiza probne provedbe ispita, provedba ispita te praćenje, analiza i interpretacija rezultata ispita, te informiranje svih sudionika u jezičnom ispitivanju o rezultatima. Izrada takvoga ispita je cikličke prirode, što znači da se tijekom razvoja ispita neki postupci ponavljaju nakon što se dobiju određene informacije iz nekih koraka izrade ispita i provedbe ispita. Ta su ponavljanja postupaka nužna radi prilagodbe ispita kontekstu njegove uporabe i podizanja kvalitete ispita na najvišu moguću razinu.⁷

Iako se za potrebe ispitivanja u razrednoj sredini ne provode tako opsežne radnje vezane uz proces izrade ispita, sam je postupak u njegovim općim crtama vrlo sličan, a uključuje, pojednostavljeno rečeno, sljedeće korake: određivanje svrhe ispitivanja, odabir sadržaja, izradu zadataka, provedbu ispita, procjenjivanje i pružanje povratne informacije. Tijekom i nakon izrade ispita za primjenu u razrednoj sredini provodi se i niz kvalitativnih, ali i neki jednostavniji kvantitativni postupci, primjerice utvrđivanje indeksa težine zadatka⁸. Ipak, budući da je primarna svrha ispitivanja u razrednoj sredini prikupljanje informacija

⁷ Detaljan opis procesa izrade standardiziranoga ispita može se naći u Bachman (1990), Bachman i Palmer (2010) i Milanović (2002), a izvrstan uvod u izradu ne samo standardiziranih ispita, nego i ispita koji se razvijaju u razrednoj sredini daju Brown i Abeywickrama (2010.).

⁸ Indeks težine zadatka se odnosi na postotak ispitanika koji (ne)točno rješavaju zadatak. Što je taj postotak veći, zadatak je za zadanu skupinu ispitanika teži.

o učeničkim jezičnim znanjima i sposobnostima i uporaba tih informacija pri donošenju odluka o daljnjim koracima u poučavanju, u prvom planu ne stoje kvantitativne analize rezultata (Lynch, 2001).

4.1. Svrha ispitivanja

Carroll je, kako prenosi Fulcher (2010:1), u jednom od programatskih tekstova o jezičnome ispitivanju iz 1961. godine naglasio kako je primarna svrha jezičnoga ispitivanja dobiti podatke koji će omogućiti „donošenje inteligentnih odluka o mogućem tijeku aktivnosti“. Promatrano iz perspektive jezičnoga ispitivanja kao društveno značajne aktivnosti, moguć je vrlo širok spektar takvih odluka. Bachman (1990) ih razmatra prije svega u okviru obrazovnoga konteksta u kojemu - na temelju informacija koje pružaju jezični ispiti - odluke donose sljedeći sudionici obrazovnoga procesa: učenici, nastavnici, roditelji te obrazovne i znanstvene ustanove. Primjerice, učenicima i nastavnicima je informacija o razini i kvalitativnim aspektima ovladanosti ispitivanim jezičnim sadržajima korisna za razumijevanje postojećega stanja i usmjeravanje i oblikovanje daljnjega procesa učenja, odnosno poučavanja. Roditelje rezultati jezičnih ispita upućuju na vrstu pomoći i suradnje koja je njihovomu djetetu potrebna u određenom trenutku, a obrazovne ustanove na eventualne potrebe preoblikovanja i dopuna nastavnih programa i kurikula te pružanje dodatnoga stručnoga usavršavanja nastavnicima ili dodatnih nastavnih sadržaja učenicima. Znanstvenim ustanovama jezični ispiti služe u svrhu istraživanja uporabe jezika i procesa ovladavanja jezikom, a pružaju uvid, doduše neizravan, i u način poučavanja jezika. Sve navedene informacije koje predstavljaju izvor za donošenje odluka imaju neposredni utjecaj i na daljnje postupke u jezičnome ispitivanju. U tom se smislu na svrhu jezičnoga ispitivanja gleda kao na polazišnu točku u procesu razvoja ispita koja utječe ne samo na određenje sadržaja ispitivanja, nego i vrstu ispitivanja i ispita te tumačenje ishoda ispitivanja (Bachman 1990). Fulcher (2010) upozorava da se i danas u nastavnoj praksi u procesu izrade ispita često izostavlja određivanje svrhe, pa čak i odabir sadržaja ispita, te odmah pristupa izradi zadataka. Rezultat toga je, obično, slaba valjanost i pouzdanost ispita, a zavisno od vrste ispitivanja, takav pristup izradi ispita negativno se odražava i na druga načela ispitivanja.

Iako se o tome malo raspravlja, jezično ispitivanje - osobito u školskome kontekstu - ima i motivacijsku svrhu. S tom su primarnom svrhom ispiti bili prvotno i uvedeni u obrazovni sustav - da pružaju informaciju o napretku i motiviraju za daljnji napredak u učenju (Fulcher, *ibid.*). I u današnje doba bi ta svrha jezičnoga ispitivanja u razrednome okruženju, ali i cijelome društvu, trebala biti izražene naglašena.

4.2. Sadržaj ispitivanja

Nakon određenja svrhe ispitivanja, proces izrade ispita nastavlja se promišljanjem o sadržaju ispita, odnosno određenjem obrazovnih ishoda čija će se usvojenost ispitivati ili, rečeno u terminologiji standardiziranoga ispitivanja, u ovome se koraku pristupa definiranju konstrukta - znanja i sposobnosti ispitanika koji stoje u pozadini njegove performanse, a koja se odvija u određenom kontekstu.

Chapelle (1998) konstrukt definira kao smislenu interpretaciju promatranoga jezičnoga ponašanja. Razlikuje teoretičare s obzirom na tri vrste perspektiva iz kojih pristupaju definiranju konstrukta. Teoretičari koji polaze od kompetencija (engl. *trait theorists*) definiraju konstrukt s obzirom na unutrašnje kognitivne strukture i procese za koje se pretpostavlja da su u osnovi performanse ispitanika. U danas najreferentnije modele kojima se pokušavaju opisati kompetencije, odnosno jezična znanja i sposobnosti koja se aktiviraju u performansu korisnika jezika ubrajaju se model komunikacijske kompetencije Canalea i Swainove (1980; 1983) i Bachmanov i Palmerov model komunikacijske jezične sposobnosti (1990; 1996). Operacionalizacija navedenih modela je, međutim, otežana, jer nisu empirijski validirani⁹. Bihevoristi definiraju konstrukt s obzirom na uvjete pod kojima se promatra performansa ispitanika. Opis kompetencija ispitanika se u takvim pristupima obično vrši s obzirom na ono što ispitanik može u određenome kontekstu jezične uporabe, a vezanom uz jezične djelatnosti slušanja, govorenja, čitanja i pisanja. Primjer takvoga opisa, kao što smo već prije naglasili, može se naći u ZEROJ-u (2005). Interakcionalisti povezuju prethodna dva gledišta, odnosno gledaju na performansu kao na rezultat kompetencija, kontekstualnih značajki i njihove interakcije u kojoj

⁹ Za prikaz i analizu modela komunikacijske kompetencije vidi Bagarić (2007).

odlučujuću ulogu imaju metakognitivne strategije. S novijim zauzimanjem za pristup procjenjivanju temeljenom na performansu ispitanika, do izražaja sve više dolazi bihevioristička, ali i interakcionalistička perspektiva u definiranju konstrukta. Stoga su i današnja istraživanja sve više usmjerena na utvrđivanje komponencijalne strukture kompetencija koje se aktiviraju pri uporabi jezika u određenoj jezičnoj djelatnosti.

Rad na određenju konstrukta, kamena temeljca u procesa razvoja standardiziranoga ispita, može potrajati tjednima, mjesecima, pa čak i više godina. Po njegovu završetku pristupa se određivanju sadržaja ispita, tj. odabire se korpus znanja i sposobnosti koji će se ispitati, te pristupa izradi zadataka.

4. 3. Izrada zadataka

Ključna odluka u procesu izrade zadataka koji će se uključiti u ispit je izbor zadatka kojim će se uistinu i mjeriti željena jezična znanja i sposobnosti. Koliko je ta odluka važna pokazuju istraživanja djelovanja varijable zadatka pri ispitivanju i procjenjivanju komunikacijske kompetencije i/ili jezičnoga umijeća, čiji se broj progresivno povećavao posljednjih dvadesetak godina. Zanimanje istraživača za proučavanjem zadatka potaknula je, prije svega, činjenica da ispit čini zadatak ili niz zadataka, ali još više od toga spoznaja da priroda konstrukta komunikacijske kompetencije i/ili jezičnoga umijeća nije stalna u svim zadacima. Različiti zadatci, primijenjeni u različitim kontekstima, odražavaju različite aspekte ispitanikova jezičnoga znanja i sposobnosti što je, u međuvremenu, empirijski dokazano u brojnim studijama (npr. Chalhoub-Deville, 1995; Gebril, 2009; Roevers, 2011). Tim je studijama ujedno potvrđen učinak metode ispitivanja na konstrukt komunikacijske kompetencije i/ili jezičnoga umijeća (neke dimenzije postaju dominantne zavisno od toga kako je konstrukt operacionaliziran u zadacima), te činjenica da se samo jednim zadatkom, a ponekad ispiti sadrže samo jedan tip zadatka, ne može dobiti sasvim valjan i pouzdan uvid u jezična znanja i sposobnosti ispitanika.

Različiti zadatci ne mjere samo različite aspekte jezičnoga znanja i sposobnosti, nego ih obilježava i različiti stupanj težine. Istraživanja težine zadatka imaju polazište u istraživanjima i raspravama o varijabilnosti u izvršavanju zadataka u području ovladavanja inim

jezikom. Tim se istraživanjima nastoji ispitati utječu li i u kojoj mjeri obilježja zadatka i uvjeta izvođenja zadatka na ispitanikovu jezičnu performansu. U većini dosadašnjih empirijskih predviđanja težine zadatka istraživači su primjenjivali psiholingvističke kriterije koje je opisao Skehan (1998):

1. Kompleksnost jezičnoga koda koja uključuje lingvističku kompleksnost i raznolikost, raspon i raznolikost vokabulara, redundantnost i gustoću.
2. Kognitivna kompleksnost na koju utječe razina kognitivne upoznatosti, tj. stupanj poznavanja teme i mogućnosti predviđanja teme, stupanj poznavanja diskursnoga žanra, stupanj poznavanja vrste zadatka, što može aktivirati već razvijene strategije, i razina potrebne kognitivne obrade, a koja zavisi od načina na koji je informacija strukturirana, stupnja potrebne transformacije ili manipulacije informacijama, o jasnoći i dostatnosti pružene informacije i o vrsti informacije (konkretna-apstraktna, kontekstualizirana-nekontekstualizirana itd.).
3. Komunikacijski stres odražava uvjete pod kojima se izvodi zadatak, a uključuje sljedeće čimbenike: vremenska ograničenja za izvršenje zadatka, učenikovu percepciju o tome pod kakvim se vremenskim pritiskom treba izvršiti zadatak, kratkoću vremena za predstavljanje učinjenoga, broj sudionika, dužinu teksta, vrstu odgovora (jednostavni ili složeni), mogućnost nadgledanja interakcije (npr. usporediti je kako bi se pozornost mogla usmjeriti na oblik i time povećati točnost u izvršenju zadatka i sl.).

Zbog svih navedenih aspekata koje je potrebno uzeti u obzir pri analizi težine zadatka, istraživanja su težine zadataka vrlo složena. Ta činjenica odrazila se i na različite rezultate istraživanja, a koji su uvelike ovisili o kontekstu i ispitanicima s kojima su se zadatci provodili (usp. Fulcher i Reiter 2003). Danas se u istraživanju težine zadataka nastoje uzeti u obzir i sociolingvističke i sociokulturalne, a ne samo psiholingvističke kategorije za određivanje težine zadatka, jer su prve kategorije osjetljivije na to koliko je određen zadatak težak ispitanicima iz određenih kulturalnih sredina.

U vezi sa zadatcima u usmenim ispitima, vrlo su zanimljiva istraživanja provedena u vezi s dva oblika ispita – individualni ispit i skupni ispit, a s ciljem utvrđivanja prihvatljivosti skupnoga ispita, prije

svega, u standardiziranome ispitivanju. Iako su istraživanja tih dvaju oblika ispitivanja u zamahu, dosadašnji rezultati upućuju na neke važne prednosti skupnoga usmenoga ispita. Primjerice, u tom se obliku ispitivanja pospješuju interakcijske vještine učenika (Gan 2010), a u slučaju da su u paru ili skupini jezično slabiji i jezično bolji ispitanici, taj oblik ispitivanja ne utječe negativno niti na jezično bolje ispitanike, niti na jezično slabije ispitanike, štoviše, potonji uporabljaju više riječi (Davis 2009). Pozitivan utjecaj skupnoga rada, osobito na aspekt točnosti, utvrđen je i u pisanju (Wiggelsworth i Storch 2009). Istraživači stoga zaključuju da skupni ispit može biti brzo sredstvo za procjenjivanje govorne sposobnosti velikoga broja ispitanika i da je u interakcijskom smislu djelotvorniji od individualnoga ispita. Ipak, zbog puno neistraženih sociolingvističkih i psiholingvističkih varijabli u tom obliku ispitivanja (npr. razlike među ispitanicima s obzirom na spol, dob, društveni status, osjećaj straha, razinu stidljivosti, spremnost na razgovor sa strancima, itd.) koje bi mogle imati utjecaja na jezičnu produkciju ispitanika, možda je opravdan oprez autora ispita u vezi s tim oblikom ispitivanja. Kombinacija dijaloških i monoloških zadataka pri provedbi skupnoga usmenoga ispita može smanjiti navedeni problem te skupni ispit učiniti prihvatljivim, ponajviše zbog toga jer se tu radi o jednom od najkomunikativnijih oblika ispita u kojem su učenici uključeni u spontanu i, ponekad, uzavrelu raspravu o raznim temama.

Paralelno s izradom zadataka otvorenoga tipa razvijaju se i sredstva za njihovo procjenjivanje na koje se najčešće referira u smislu ljestvica za procjenjivanje. Puno je vrsta i načina klasifikacije ljestvica za procjenjivanje, a mogu se odnositi na izgled ljestvice, elemente koji se procjenjuju ili, pak, na sadržaj opisivača. Bachman i Palmer (2010) ističu dva osnovna pristupa procjenjivanju, a time i dvije vrste ljestvica za procjenjivanje: holističko ili globalno procjenjivanje i analitičko procjenjivanje, odnosno procjenjivanje s obzirom na pojedinačne elemente komunikacijske jezične sposobnosti. Analizirajući načine razvoja ljestvica, Brindley (1998) je uočio dva tipa ljestvica. Prvi je tip ljestvica koja se izvodi iz teorijskoga modela komunikacijske kompetencije i/ili jezičnoga umijeća i pruža analitičke ocjene njezinih raznih sastavnica. Ta je ljestvica određena nezavisno od sadržaja i konteksta stvarne jezične uporabe. Nasuprot tomu, drugi tip ljestvice, tzv. biheavioristička ljestvica, sastoji se od niza opisa jezičnoga ponašanja

u određenim kontekstima jezične uporabe pri čemu se jezična sposobnost opisuje na nekoj vrsti kontinuuma rastuće sposobnosti koja obično ima raspon od „nula“ do „izvorni govornik“. Potrebno je napomenuti da se danas kategorija izvornoga govornika zamjenjuje kategorijom „osobe koja vlada jezikom na razini bliskoj razini izvornoga govornika“. Također, u razvoju opisivača na ljestvici za procjenjivanje komunikacijske kompetencije i/ili jezičnoga umijeća koriste se razni pristupi. Pored teorijskoga pristupa, koji se temelji na prosudbi stručnjaka, posljednjih se godina u standardiziranome ispitivanju preporuča empirijsko izvođenje ljestvice, i to s obzirom na određene zadatke i određenu populaciju ispitanika, odnosno kontekst provedbe ispita.

Bez obzira na vrstu ljestvica i način na koji su ljestvice razvijene, njihova je svrshodnost neupitna, kako u području standardiziranoga ispitivanja tako i u području ispitivanja u razrednoj sredini: ljestvice pružaju korisnicima ispita informaciju o značenju ocjena, a ocjenjivačima su sredstvo za što objektivniju procjenu kvalitete ispitanikove jezične performanse.

4.4. Provedba ispita

Pored tehničkih uvjeta u kojim se provodi ispit, dva su vrlo važna čimbenika u provedbi zadataka otvorenoga tipa koji mogu utjecati na rezultate ispita, a to su, u slučaju dijaloških zadataka na usmenome ispitu, komunikacijski partner koji može biti neki drugi ispitanik ili stručna osoba, te ocjenjivač usmene ili pisane jezične proizvodnje ispitanika.

U teoriji varijabilnosti stoji da komunikacijski partner najviše utječe na proizvodnju korisnika inoga jezika. Stoga se u području jezičnoga ispitivanja pristupilo istraživanju utjecaja osobina komunikacijskoga partnera na ispitanika i njegov uspjeh na komunikacijskome jezičnome ispitu. Utvrđeno je da dob, spol i status komunikacijskoga partnera mogu značajno utjecati na ispitanike, ali da to ovisi o kulturalnome kontekstu (O'Sullivan, 2002). Brown (2003) je k tome u svojem istraživanju upozorila na razlike među ispitivačima s obzirom na strukturu razgovora, postavljanje pitanja i tip odgovora koje su pobuđivali. Učinak sugovornika na jezičnu performansu ispitanika je,

dakle, složen i danas predstavlja, uz ulogu ocjenjivača, jedno od najaktualnijih područja istraživanja u jezičnome ispitivanju.

Uloga ocjenjivača u procesu ispitivanja je od izuzetnoga značaja, tim više što u ispitivanju u razrednoj sredini, a ponekad i u standardiziranome ispitivanju, jedna osoba ima istovremeno dvije uloge: ocjenjivača i ispitivača. Mnoge su studije posvećene toj tematici (npr. Weigle, 1998; Kondo-Brown, 2002; Bonk i Ockey, 2003; Zhang i Elder, 2011) i sve su upozorile na razlike u procjenama i ponašanju ocjenjivača prilikom procjenjivanja. Te su razlike obično bile uzrokovane nekim od sljedećih čimbenika: samim procesom ocjenjivanja, podrijetlom ocjenjivača, količinom prethodne obuke ili iskustva u procjenjivanju jezičnih sposobnosti u govorenju ili pisanju, učinkom obuke i pružanja povratne informacije, stupnjem pristranosti (strogosti ili blagosti) ocjenjivača, pogriješcima u interpretaciji ljestvice, spolom ocjenjivača itd. Pored razlika u razini strogosti ocjenjivača, koja se u standardiziranome ispitivanju može minimalizirati obukom i izborom ocjenjivača, potencijalno najveći problem, prema našem mišljenju, krije se u nalazima kako ocjenjivači, čak i oni koji su obučeni za analitički pristup procjenjivanju, jezičnu proizvodnju ispitanika u pravilu ocjenjuju holistički te kako je pri tome određeni aspekt jezičnih znanja i sposobnosti presudan pri procjeni. U svakom slučaju, toj bi problematici u određenim obrazovnim kontekstima trebalo posvetiti više istraživanja.

4.5. Analiza ispita

Nakon provedbe ispita (u slučaju standardiziranoga ispitivanja, već nakon probne provedbe ispita) provode se analize kojima se želi utvrditi ispituje li se zadatcima u ispitu ono što se želi ispitati, u kojoj su mjeri zadatci teški, u kojoj mjeri razlikuju slabije od boljih ispitanika, jesu li primjereni kontekstu ispitivanja, drugim riječima primjenom kvalitativnih i kvantitativnih, često vrlo složenih psihometrijskih postupaka (npr. Rasch analiza, strukturalno modeliranje, teorija odgovora na zadatak itd.) traži se odgovor na pitanje udovoljava li ispit načelima ispitivanja¹⁰. Najvažnije načelo je svakako valjanost - mjeri li ispit ono što je predmet mjerenja. U kojoj je mjeri to načelo važno, govori činjenica da u teoriji jezičnoga ispitivanja posebno mjesto zauzimaju

¹⁰ Vrlo detaljan prikaz načela u jezičnome ispitivanju daje Škarica Mital (2011).

teorije valjanosti (usp. Davies i Elder, 2005). Drugo načelo po važnosti, a koje je usko povezano s načelom valjanosti jest pouzdanost – pruža li ispit u nekoliko uzastopnih mjerenja istoga sadržaja iste ili što sličnije rezultate mjerenja. Načelo objektivnosti je u rangu važnosti s prvim dvama načelima, a prema njemu rezultat pojedinoga ispitanika treba ovisiti samo o njegovim jezičnim znanjima i sposobnostima. U načela ispitivanja koja su manje ili više povezana s navedenim trima načelima, ubrajaju se još: osjetljivost - je li zadatak osjetljiv na razlike između boljih i slabijih ispitanika; autentičnost – u kojoj su mjeri zadatci u skladu sa svrhom i potrebama ispitanika, tj. u kojoj mjeri traže od ispitanika da primjene i pokažu one jezične sposobnosti koje se primjenjuju u stvarnim komunikacijskim situacijama; povratni učinak – u kojoj mjeri ispit utječe na nastavni proces, što predstavlja jednu od izuzetno aktualnih tema istraživanja u područja standardiziranoga jezičnoga ispitivanja; praktičnost – u kojoj je mjeri ispit jednostavan za provedbu s obzirom na potrebne materijalne i ljudske resurse; transparentnost – u kojoj je mjeri postupak provedbe za sve sudionike u procesu ispitivanja transparentan; i konačno, sigurnost – u kojoj je mjeri provedba ispita sigurna, što značajno može utjecati na valjanost ispita i pravednost prema svim ispitanicima. Pravednost je, inače, u jezičnome ispitivanju danas vrlo aktualna tema rasprava.

Kada su u pitanju znanstvena istraživanja u području jezičnoga ispitivanja, svakako je potrebno napomenuti kako se danas, uz već gore spomenute kvantitativne analize, sve više naglašava primjena kvalitativnih metoda istraživanja i analize, primjerice, diskursne analize i introspektivnih postupaka.

5. Umjesto zaključka

Nakon što smo pokušali pružiti uvid u osnovne pojmove i sadržaje u okviru jezičnoga ispitivanja danas, rad bismo zaključili navođenjem još nekih nedovoljno istraženih aspekata u tome području. U tom bismo smislu na prvo mjesto stavili potrebu istraživanja procesa ispitivanja s ispitanicima rane školske dobi, kako u području standardiziranoga ispitivanja tako i u ispitivanju u razrednoj sredini. Druga tema je danas sve aktualnije jezično ispitivanje uz primjenu novih tehnologija. Treća tema je alternativno vrjednovanje i njegova primjena u standardiziranome ispitivanju. Kao četvrtu temu naglasili bismo potrebu

sustavnoga istraživanja ispitivanja i procjenjivanja u razrednoj sredini, a s tim u vezi i istraživanja jezičnoga ispitivanja u određenome obrazovnome i kulturalnome kontekstu.

Vrlo dobar uvid u istraživanja u području jezičnoga ispitivanja pružaju časopisi, od *Language Testing*, vodećeg časopisa u tom području do *Language Assessment Quarterly*, *International Journal of Testing* i drugih. Sve veći broj priloga u tim časopisima govori o brzom razvoju ove relativno mlade znanstvene discipline. Ono što je, međutim, izuzetno važno jest činjenica da se rezultati istraživanja u tome području gotovo istodobno prenose u praksu. To potvrđuje i postojanje relativno velikoga broja priručnika za razvoj i provedbu standardiziranih ispita, ali i ispita u razrednoj sredini. Iako ih je najviše na engleskome jeziku, s isticanjem važnosti kontekstualnih značajki u ispitivanju, pojavljuju se i priručnici na jeziku populacije ispitanika kojima su namijenjeni. Navedeni časopisi i priručnici zasigurno će doprinijeti boljem razumijevanju ispitivanja u (stranome) jeziku koje, prema mišljenju mnogih, predstavlja jedno od najzahtjevnijih područja procesa učenja i poučavanja jezika.

LITERATURA:

- Alderson, J. C. (2005): *Diagnosing Foreign Language Proficiency: The Interface between Learning and Assessment*. London: Continuum.
- American Council on the Teaching of Foreign Languages (1986): *ACTFL proficiency guidelines*. Hastings-on Hudson, NY: ACTFL.
- Bachman, L. F. (1990): *Fundamental Considerations in Language Testing*. Oxford etc.: OUP.
- Bachman, L. F. (2002): Some reflections on task-based language performance assessment. *Language Testing* 19, 4, 453-476.
- Bachman, L. F., Palmer, A. S. (1996): *Language Testing in Practice: Designing and Developing Useful Language Tests*. Oxford etc.: OUP.
- Bachman, L. F., Palmer, A. S. (2010): *Language Assessment in Practice*. Oxford etc.: OUP.
- Bagarić, V. (2007.): *Struktura komunikacijske kompetencije u stranom jeziku*. Neobjavljena doktorska disertacija. Zagreb: Filozofski fakultet.
- Black, P., Harrison, C., Lee, C., Marshall, B., William D. (2003): *Assessment for Learning: Putting it into Practice*. Buckingham, UK: Open University Press.
- Bonk, W. J., Ockey, G. J. (2003): A many-facet Rasch analysis of the second language group oral discussion task. *Language Testing* 20, 89-110.
- Brindley, G. (1998): Outcomes-based assessment and reporting in language learning programmes: a review of the issues. *Language Testing* 15, 1, 45-85.

- Brindley, G. (2001): Outcomes-based assessment in practice: some examples and emerging insights. *Language Testing* 18, 393-407.
- Brown, D. H., Abeywickrama, P. (2010): *Language Assessment. Principles and Classroom Practices* (2. izd.). White Planis, NY: Pearson Longman.
- Brown, A. (2003): Interviewer variation and the co-construction of speaking proficiency. *Language Testing* 20, 1, 1-25.
- Canale, M. (1983): From communicative competence to communicative language pedagogy. U Richards, J. C., Schmidt, R. W. (ur.), *Language and Communication*, 2-27. London: Longman.
- Canale, M., Swain, M. (1980): Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics* 1, 1, 1-47.
- Chalhoub-Deville, M. (1995): A contextualized approach to describing oral language proficiency. *Language Learning* 45, 2, 251-281.
- Chapelle, C. A. (1998): Construct definition and validity inquiry in SLA research. U Bachman, Lyle F. i Cohen, Andrew D. (ur.), *Interfaces between second language acquisition and language testing research*, 32-70. Cambridge: Cambridge University Press.
- Cohen, A. D. (1994): *Assessing Language Ability in the Classroom*. Boston, Massachusetts: Heinle and Heinle Publishers.
- Davies, A. (1988): Communicative Language Testing. U Hughes, A. (ur.), *Testing English for University Study*, 5-15. Hong Kong: Modern English Publications and The British Council (ELT Documents 127)
- Davies, A., Elder, C. (2005): Validity and Validation in Language Testing. U Hinkel, E. (ur.), *Handbook of Research in Second Language Teaching and Learning*, 795-814. Mahwah, New Jersey and London: Lawrence Earlbaum.
- Davis, L. (2009): Interacting in pairs in a test of oral proficiency: Co-constructing a better performance. *Language Testing* 26, 341-364.
- Fulcher, G. (2000): The "Communicative" Legacy in Language Testing. *System*, 28, 483-497.
- Fulcher, G. (2010): *Practical Language Testing*. London: Hodder Education.
- Fulcher, G., Márquez Reiter, R. (2003): Task difficulty in speaking tests. *Language Testing* 20, 3, 321-344.
- Gan, Z. (2010): Interaction in group oral assessment: A case study of higher and lower scoring students. *Language Testing* 27, 585-602.
- Gebril, A. (2009). Score generalizability of academic writing tasks: Does one test method fit it all? *Language Testing* 26, 507-531.
- Green, B. A., Weir, C. J. (2004): Can placement tests inform instructional decisions? *Language Testing* 21, 467-494.
- Hill, K., McNamara, T. (2012): Developing a comprehensive, empirically based research framework for classroom-based assessment. *Language Testing* 29, 395-420.
- Hughes, A. (2003): *Testing for language teacher* (2. izd.). Cambridge: CUP.

- Huhta, A., Kalaja, P., Pitkänen-Huhta, A. (2006): Discursive construction of a high-stakes test: the many faces of a test taker. *Language Testing* 23, 326-350.
- Hulstijn, J. H. (1985.): Testing second language proficiency with direct procedures. A comment on Ingram. U Hyltenstam, K. i Pienemann, M. (ur.), *Modelling and assessing second language acquisition*, 277-281. Avon: Multilingual Matters Ltd.
- Ingram, D. E. (1985): Assessing proficiency: An overview on some aspects of testing. U Hyltenstam, K., Pienemann, M. (ur.), *Modelling and assessing second language acquisition*, 215-276. Avon: Multilingual Matters Ltd.
- Jelaska Z., Cvikić L. (2008). Procjenjivanje i vrjednovanje u hrvatskome jeziku. *Lahor* 5, 115-125.
- Kondo-Brown, K. (2002): A FACETS analysis of rater bias in measuring Japanese second language writing performance. *Language Testing* 19, 3-31.
- Lantolf, J. P., Poehner, M. E. (2007): *Dynamic Assessment in the Foreign Language Classroom: A Teacher's Guide*. Pennsylvania: Calper Publications.
- Law, B., Eckes, M. (2007): *Assessment and ESL: An alternative approach*. Winnipeg, Manitoba, Canada: Portage and Main Press.
- Leung, C., Lewakowicz, J. A. (2006): Expanding horizons and unresolved conundrums: Language testing and assessment. *TESOL Quarterly* 40, 211-234.
- Leung, C., Teasdale, A. (1997): What do teachers mean by speaking and listening: a contextualised study of assessment in the English National Curriculum. U Huhta A., Kohonen V., Kurki-Suonio, L., Luoma S. (ur.), *New contexts, goals and alternatives in language assessment*, 291-324. Jyväskylä: University of Jyväskylä.
- Lin, Q. (2002): Beyond Standardization: Testing and Assessment in Standards-Based Reform. *Action in Teacher Education* 23, 43-49.
- Lynch, B. K. (2001): Rethinking assessment from a critical perspective. *Language Testing* 18, 333-349.
- Matsuno, S. (2009): Self-, peer- and teacher assessment in Japanese university EFL writing classrooms. *Language Testing* 26, 75-100.
- Milanovic, M. (2002): *Language examining and test development*. Strasbourg. Language Policy Division.
- Norris, J. M., Brown, J. D., Hudson, T. D., Yoshioka, J. (1998): *Designing Second Language Performance Assessments*. Honolulu: Second Language Teaching and Curriculum Center, University of Hawai'i at Mānoa.
- Kantarcioğlu, E., Papageorgiou, S. (2011): Benchmarking and standards in language tests. U O'Sullivan, B. (ur.), *Language Testing: Theories and Practices*, 94-110. Basingstoke: Palgrave Macmillan.
- Knoch, U. (2009): Diagnostic assessment of writing: A comparison of two rating scales. *Language Testing* 26, 275-304.
- Oller, J. W. (1979.): *Language Tests at School: A Pragmatic Approach*. London: Longman Group Ltd.
- O'Sullivan, B. (2002.): Learner acquaintanceship and oral proficiency test pair-task performance. *Language Testing* 19, 3, 277-295.

- O'Sullivan, B., Weir, C. J. (2011): Test development and validation. U O'Sullivan, B. (ur.) *Language Testing: Theories and Practices*, 13-32. Basingstoke: Palgrave Macmillan.
- Phelps, R. P. (2008): *Kill the Messenger: The War on Standardized Testing*. New Brunswick, New Jersey: Transaction Publishers.
- Rea-Dickens, P. (2001): Mirror, mirror on the wall: identifying processes of classroom assessment. *Language Testing* 18, 429-462.
- Roeever, C. (2011): Testing of second language pragmatics: Past and future. *Language Testing* 28, 463-481.
- Shohamy, E. (1997): Testing methods, testing consequences: are they ethical? Are they fair? *Language Testing* 14, 340-349.
- Shohamy, E. (2001): Democratic assessment as an alternative. *Language Testing* 18, 373-391.
- Skehan, P. (1991): Progress in Language Testing. The 1990s. U Alderson, C. J., North, B. (ur.), *Language Testing in the 1990s: The Communicative Legacy*, 3-21. London: Modern English Publication in association with The British Council.
- Skehan, P. (1998): *A Cognitive Approach to Language Learning*. Oxford: OUP.
- Spolsky, B. (1995): *Measured Words*. Oxford: Oxford University Press.
- Škarica Mital, I. (2011): *Vrjednovanje komunikacijske kompetencije u stranome jeziku: Usporedba nastavnikovih procjena i rezultata standardiziranih komunikacijskih testova*. Neobjavljeni magistarski rad. Zagreb: Filozofski fakultet.
- Vijeće za kulturnu suradnju, Odbor za obrazovanje, Odjel za suvremene jezike u Strasbourgu (2005): *Zajednički europski referentni okvir za jezike: učenje, poučavanje, vrednovanje*. Zagreb: Školska knjiga. (Prijevod: Valnea Bressan i Martina Horvat).
- Weir, C. J. (1990): *Communicative Language Testing*. New York etc.: Prentice Hall International.
- Weir, C. J. (2005.): *Language Testing and Validation: An Evidence-Based Approach*. Oxford: Palgrave Macmillan.
- Weigle, S. C. (1998): Using FACETS to model rater training effects. *Language Testing* 15, 263-287.
- Wigglesworth, G., Storch, N. (2009): Pair versus individual writing: Effects on fluency, complexity and accuracy. *Language Testing* 26, 445-466.
- Zhang, Y., Elder, C. (2011): Judgements of oral proficiency by non-native and native English speaking teacher raters: Competing or complementary constructs? *Language Testing* 28, 31-50.

LANGUAGE TESTING TODAY

In the paper *Tests in foreign languages* published in this journal in 1971, Mirjana Vilke gives an overview of test types, purposes and contents as well as assessment procedures used in foreign language teaching. Since 1970s, a great progress has been made in understanding the *what* and the *how* in the testing process, i.e. in defining the construct and testing methods. The existing concepts have been redefined and new ones proposed as a result of research within a fast developing scientific subfield of applied linguistics, named *language testing*. The first part of the paper reviews the development of approaches to language testing since the middle of the 20th century as well as types of testing fields and tests. The main part is dedicated to the theoretical and empirical aspects of test development with special attention to test purpose and content, issues in task and assessment procedure development, test administration and analysis. In the end, research areas in which more intensive activities could be expected in the future are identified.

Key words: testing, assessment, communicative language testing, performance assessment, standardized testing, classroom testing