

Testen euskaratzea. Balizko alborapenaren iturriak¹

Paula Elosua Oliden

Psikologia Fakultatea
Gizarte Psikologia eta Portaera Zientzien Metodologia Saila

GAKO-HITZAK: Ebaluazio. Test. Baliokidetasuna. Funtzionamendu diferentziala. Alborapen.

1. SARRERA

Hizkuntzen arteko proben egokitzapena, testen sorrera bertan abiapuntua duen jarduera da. Horren lekukorik nabariena Binet-Simon-en *Intelligence Scale for children* litzake, zein 1911. urtean frantsesetik ingelesera itzuli ondoren, 1916.enean jada zazpi hizkuntza desberdinetan erabiltzen zen. Ebaluazio psiko-pedagogiakoan egokitzapena aktibitate ohikoa izan arren, horren kalitatearen azterketa berria dugu ikerketa psikometrikoan. XXgarren mendearen azkeneko hamarkadan kokatuko genuke jatorrizko eta egokituriko bertsioen arteko baliokidetasuna bermatzearen beharrak zabaldu duen ikerketa eremua. Gaiaren gailentasuna besteak beste, azken urteotan indartzen ari diren nazioarteko hezkuntza sistemen ebaluaketa proiektuekin bat dator. Testigutza modura TIMSS (*Third International Mathematics and Science Study*) edo PISA (*Programme for International Student Assessment*) ikerketak aipatuko genituzke, zeinen barne, proba berdinak 30 hizkuntza baino gehiagotara itzuliak izan diren.

Egokitzapenaren arrazoiak desberdinak izan badaitezke ere, kultura-arteko ikerketak, gutxiengo linguistikoen ebaluazioa, berariazko proben ga-bezia..., xedea puntuazioen arteko erkaketa denean, ezinbestekoa da horien ezaugarri metrikoen arteko parekotasuna bermatzea. Nola berdindu esanahi desberdinak barneratzen dituzten puntuazioak? Zein ondorio atera daitezke barne egitura diferenteak islatzen dituzten testen alderatzetik? Egokitzapenaren helburua konparazioa denean, erkaketan parte hartzen duten testen

¹ Lan hau EHU/UPV-ren dirulaguntzaz gauzatu da:1/UPV 00109.231-H-14855/2002.

ezaugarri psikometrikoen arteko baliokidetasuna da aztergaia, eta hortaz balizko alborapena ere. Alderakizun guztien oinarrian bi kontzeptu hauek daude; baliokidetasuna eta alborapena. Horiek ongi aztertutik ez dauden bitartean erkaketei uko egin beharko litzaieke.

2. BALIOKIDETASUNA

Egokitzapenaren azterketaren muina baliokidetasuna da. Bi test balioki-deak dira baldin eta horien bidez eskuratzen diren puntuazioetatik eratortzen diren inferentziak konparagarriak badira; hau da, horien eredu formal eta substantiboen arteko parekotasuna lortzen bada (Elosua, 2003). Ez du honek esanahi inolaz ere, egokituriko bertsioak eta jatorrizkoak, baremo edo banaketa berdinak izan behar dituztenik bi taldetan. Ez da hori kontua. Tipifikazio taulak desberdinak izango baitira, populazioen gaitasun banaketan arabera.

Baliokidetasunaren azterketan, honek izan ditzaken mailak ezagutu behar dira. Izan ere, bakoitzak berariazko ezaugarriak ditu, eta horiek dira puntuazioen interpretazio eta balizko tratamendu diferenteak gidatuko dituztenak. Hiru dira nagusiki bereizten ditugunak; konstruktoen arteko baliokidetasuna, baliokidetasun metrikoa eta metodoarekin loturikoa.

a) *Baliokidetasuna konstruktoan*: Berdintasunaren lehen baldintza, neurketaren xede diren konstruktoen arteko ekibalentzia da; *egiturazko baliokidetasuna* edo *baliokidetasun funtzionalaz* ere ezagutzen dena. Honek esan nahi du, aldagai berdina neurtzen dela talde, kultura nahiz hizkuntza desberdinetan. Hau da, konstruktoaren unibertsalitatea onartzen da; kultura arteko ikerketetan *etic* izenez ezagutzen dira aldagai orokor hauek.

Aztergai diren populazioetan, neurketaren xede modu desberdinetan ulertu, edo adierazten bada, ez litzake baliokidetasunaren gutxienezko kondizioa beteko. Esate baterako, adimenaren kontzepzioan diferentziak daude mendebalde eta ekialdetako kulturen artean. Gure inguruan adimen probek interesa arrazonamenduan, oroimenean eta abarreko alderdietan finkatzen duten bitartean, ekialdeko kulturek alderdi sozialei ematen diete gailentasuna (Super, 1983). Neurketa objektuak kulturen arabera definizio desberdinak baldin baditu, hau da *emic* motakoa bada, testaz kanpoko inferentziak baliogabetuak geratuko lirateke.

Baina konstruktoen arteko baliokidetasun ezaz hitz egiteko ez da derri-gorrez urrun dauden bi kultura alden behar. Testaren erabilzaileak, jatorrian proposaturiko erabilpenetan aldaketak sartzean, nahiz xedetan nahiz norentzakoan, konstruktoaren balizko parekotasuna zalantzan jar dezake. Adibide gisa, B.A.D.Y.G.-E (Yuste, 1988) probaren «Comprensión Verbal» atala ekarri dugu. Probaren helburua gaztelarazko hizkuntzaren ezagupenaren neurria ematea da, hizkuntza hori ama- eta hezkuntza-hizkuntza duten haurrengan. Testa, euskarazko lagin batean aplikatuko bagenu, ez genuke jada ama- eta hezkuntza-hizkuntzaren menerapen neurririk jasoko. Honela jardutean, proba-

ren erabilera egokiak eskatzen duen kanpo testuingurua hausten da. Haurrak euskaldunak izanik proba horrek bigarren hizkuntza baten ezagupena antzemango bailuke. Egoera hori enpirikoki aztertuz, proba osatzen duten itemen %50ek bi laginetan dituzten ezaugarrien arteko baliokidetasun eza egiaztatu dugu (Elosua, López eta Egaña, 2000; Elosua eta lag., 2000).

Konstruktoen arteko baliokidetasunaz aritzerakoan beraz, baliokidetasun *teorikoaz*, eta *erabilpenarekin* loturikoak bereiztuko genituzke. Lehen kasuan, neurtu nahi den konstruktoarekiko definizio baliokiderik ez bada go neurketa tresnen arteko parekotasuna ezinezkoa da. Horregatik, erabat desegokia da kasu hauetan puntuazioen arteko erkaketa, horien esangura psikologikoa desberdina baita. Bigarren kasuan, testaren *erabilpen* desegokian dago baliokidetasun ezaren jatorria. Honek puntuazio alboratuak sortuko lituzke, baliagarritasuna bera kolokan jarri.

b) *Itemen arteko baliokidetasuna, edo baliokidetasun metrikoa*. Nahiz konstruktoen arteko baliokidetasun teorikoa ziurtaturik egon, hori neurtzeko erabiltzen diren itemak populazioen arabera modu diferentetan erlazionatu daitezke xedearekin. Kasu hauetan, itemen barne ezaugarri psikometrikoak desberdinak izaten dira taldetan. Aurreko atalean aitaturiko proba berdinarekin jarraituz, oraingoan euskaratu eta D ereduan dihardutenei eginarazten zaie. Test baten bi bertsio ditugu beraz, «Comprensión Verbal» eta «Hitzezko Ulermena». Biek konstrukto berdina dute xedetzat, «Hizkuntza nagusiaren hitzezko ulermen», alegia. Itzulpena irizpide linguistikoei jarraituz gauzatu denez, euskaraturiko proba osatzen duten itemen kontzeptu eta terminoek ez dituzte jatorrizkoarekiko eduki semantiko berberak izango. Hizkuntzak ez baitira paraleloak. Horregatik horien bidez lorturiko puntuazioek ez lukete baliokidetasun metrikoaren baldintza beteko. Hala erakusten du itemen barne egituraren azterketak. Erabilitako itemetatik %76k ez dituzte ezaugarri psikometriko berdinak bi taldetan (Elosua eta López, 1999a).

c) Bi maila hauezz gain zenbait autorek *metodoarekin erlazionaturiko baliokidetasuna* definitzen dute (van der Vijver eta Poortinga, 1997; van der Vijver eta Tanzer, 1997). Honen barne, ikerketan parte hartzen duten laginen arteko berdintasunarekin, erabilitako tresna motarekiko familiartasunarekin eta administrazio egoera berarekiko hurbiltasunarekin loturiko aspektuak aitaturiko genituzke (Hui eta Triandis, 1989).

Ikusten denez, hirugarren maila hau, aztergai diren populazioen arteko distantzia handitzearekin batera azaldu eta areagotzen doa. Ez da hori gure kasua ordea, non aitaturiko faktoreetan populazioen arteko berdintasunaren hipotesia onartuko genukeen.

3. PROBEN EGOKITZAPENERAKO GOMENDIOAK

Baliokidetasunaren helburua bermatzeko *International Test Comission*-ek (ITC) 1992garren urtean test psikologiko eta pedagogikoen itzulpenerako

gomendioak idaztearen proiektuari ekin zion. Hauek 1999.ean argia ikusi zuten Georgetown-eko Unibertsitatean (Washington) gauzaturiko egokitzapenari buruzko Nazioarteko Konferentzian. Gomendioen inguruan lan asko sortu da, horiek zabaldu eta jakitera emateko asmoekin; Hambleton (2001), Hambleton eta Patsuala (1999), edo van de Vijver eta Leung (1997). Hedatzeaz gain, hor biltzen diren irizpideak azken urteotan hezkuntza sistemak ebaluatzeko indarra hartu duten nazioarteko proiektuetan jarraitu izan dira (Grysay, 1998, 1999).

Gomendioak (Ikus. eranskina) 22 puntuz osaturik daude. Horiek, egokitzapen prozesuaren ostean fruitu onuragarria ziurtatzeko orientabidea, eta lorturiko emaitzen interpretazio eta erabilpen zuzenentzako gida eskaintzen diete testen eraikitzaile/egokitzaileei. Gomendioak lau arlo desberdinetan sailkatuak ematen zaizkigu: testuingurua, eraketa/egokitzapena, aplikazioa eta puntuazioen interpretazioa.

Laburbilduz, gomendioek hiru arlo orokor betetzen dituztela esango genuke, testaren kanpokoa, testaren barnekoa eta erabilpenarekin loturikoak. Hauek lehen aipatu alborapen mailarekin erlazionatuko genituzke; hurrenez hurren, konstruktua, itema eta metodoa.

- a) *Kanpo arloak*: Bi dira arlo honetako gomendioak, C1-C2. Hauek lehen aipaturiko konstruktua baliokidetasunarekin loturik daude; probaren aplikazioen ingurune soziokulturalari so egitea gomendatzen dute, konstruktuen arteko balizko baliokidetasuna aztertuz.
- b) *Barne arloak*: Hamar gomendio biltzen ditu. Proben egokitzapenaren helburu den konstruktua modu berdintsutan neurtu behar dela azpimarratzen du baina kultura bakoitzaren berezitasunetara egokituz. Hori aurrera eraman ahal izateko, beharrezkoa da itzultzaileen aukeraketa fina. Hauek hizkuntzen menarapenaz gain, jokoan dauden kulturak eta itemen erredakziorako oinarrizko arauak ezagutu beharko lituzkete. Hara hemen, non ikusten den egokitzapenaren alderdi linguistikoa, kulturala eta psikometrikoa. Horretaz gain, eta egokitzapen linguistikoa baliokidetasun metrikoa bermatzen ez duenez, lorturiko bertsioaren ezaugarri psikometrikoak aztertu behar dira, fidagarritasuna eta baliagarritasuna bermatzeko. Honen barnean, berebiziko garrantzia du, itemaren funtzionamendu diferentzialaren kontzeptuak.
- c) *Erabilpen arloak*: Bi sailetan banaturiko hamar gomendioak, aplikatzailea eta proba burutzen duten subjektuen arteko *rapport*-arekin eta puntuazioen interpretazioarekin loturik daude. Egokituriko bertsioen erabilpen zuzena eta etikoa bermatzeko idatziriko gomendioak dira.

Gomendio bakoitzaren garrantzia egokitzapenaren iturri eta xede diren hizkuntza-kulturen arteko diferentzi-berdintasunen menpe dago. Izan ere, zenbat eta elementu komun gehiago, sozialak, politikoak, ekonomikoak edo linguistikoa, orduan eta eragin ahulagoa izango dute probaren kanpo alderdiak egokitzapenean. Aplikatzaileaz edo puntuazioen interpretazioaz

gogoeta berdinak egin daitezke. Oro har esan genezake, egokituriko bertsioan oinarrituriko inferentziak justifikatuko dituzten ebidentziak ezinbestekoak direla beti.

4. ITEMEN BALIOKIDETASUN METRIKOA

Baliokidetasun mailetatik eta gomendio guztietatik, lan honi dagozkienetan, baliokidetasun metrikoarekin loturiko aspektuak dira bereziki interesatzen zaizkigunak. Izan ere, nahiz testen erabilera egokia izan, euskaraturiko proben baliagarritasunaren azterketan ezinbesteko urratsa baita.

Item batek bi taldetan dituen barne ezaugarrien parekotasuna ziurtatzeko, horien arteko baliokidetasuna aztertu behar da. Baina zer dira item balioki-deak? Noiz betetzen da baliokidetasun metrikoaren baldintza? Bi item psikometrikoki baliokideak direla esateko, biek populazio desberdinetan jokabide berdinak neurtzeaz gain, gaitasun maila berdineko subjektuengan erantzun jakin bat probabilitate berdinarekin sortarazi beharko lukete. Hau da, itema eta horrek neurtu nahi duen arrazkoaren arteko erlazioa berdina denean bi taldetan.

Hori ebaluatzeko bitarteko kualitatiboak eta kuantitatiboak ditugu. Lehenek, *juizio prozeduraz* ere ezagunak, batez ere alderdi linguistikoa zainduko dute; itemek barneratzen dituzten edukien arteko parekotasunean erreparatuz. Bigarrenak prozedura enpirikoak dira. Hauek itemen arteko baliokidetasuna ebaluatzeko zenbakizko indikatzaileez baliatzen dira. Juizio prozeduren artean eta egokitzapenari dagozkionetan, ikertzaileak hiru aukera ditu: atzerakako-itzulpena, aurreteta eta deszentralizazioa.

- a) *Atzerakako-itzulpena (Back-translation)* (Brislin, 1970): Prozedurarik zaharrenetakoa izanik (lehen aipamenak 60. hamarkadakoak dira (Fink, 1963; Sinaiko, 1963; Werner eta Campbell, 1970), erabiliaren izaten darraio. Metodo iteratibo honen lehen aldian, sorburu bertsioa xede hizkuntzara itzultzen da. Ondoren, lehen itzulpenarekin inolako zerikusirik izan ez du(t)en subjektua(e)k egokituriko bertsioa jatorrizko hizkuntzara darama(te). Pauso horien emaitzaz bi proba izango genituzke sorburu hizkuntzan, jatorrizkoa eta atzerakakoa, eta xede hizkuntzan berriz bat, egokiturikoa. Metodoa sorburu hizkuntzako bi bertsioak erkatzean asentatzen da. Izan ere, horien arteko balizko desberdintasunak aztertuz baliokidetzen dira sorburu eta xede probak.
- b) *Aurreteta*: Bigarren hizkuntzara egokituriko proba, xede taldeko lagin bati egin arazten zaio, bertakoei banan-banan item bakoitzaren esanahiaz itaunduz. Portzentaia handi baten erantzunak arrazoitsutzat hartzea nahikoa baldintza da baliokidetasunarentzat (Mitchel, 1966; Schuman, 1966).
- c) *Deszentralizazioa* (Werner eta Campbell, 1970): Prozedura honek itemak barnera dezaken berezitasun kulturala ezabatu nahi du, bigarren

hizkuntzan horren kontzeptu baliokidea aurkitu ahal izateko. Jatorri hizkuntzako edukiak xede hizkuntzan parekorik ez balu, nola edo hala hori neutralizatzen da antzekoa den batez ordezkatzuz. Sorturiko kontzeptu berria da, bigarren hizkuntzara egokitzen dena.

Egokitzapenaren kalitatea bermatu ondoren, itemen balizko alborapena enpirikoki arakatzeko metodologia itemaren funtzionamendu diferentzialaz baliatzen da. Item batek funtzionamendu diferentziala du, baldin eta talde diferentzetatik datozenek, nahiz gaitasun maila berdina izan, itema ongi erantzuteko dituzten probabilitateak desberdinak direnean (Elosua eta López, 1999b). Honek esan nahi du, itemaren erantzun zuzena ez dagoela proposaturiko xedeaz bakarrik mugatua, aitzitik badagoela beste faktoreren bat horretan eragiten duena.

Egun, ikertzaileak hainbat metodo desberdin eskura ditzake funtzionamendu diferentziala aurkitzeko (ikus. Camilli eta Shepard, 1994; Elosua, López eta Torres, 1999; Holland eta Wainer, 1993). Horien artean, Mantel-Haenszel izenekoa nagusitu da ikerketa aplikatuan. Prozedura honek bi talde desberdinetan, gaitasun maila berdineko subjektuen erantzun zuzen eta okerrak erkatzen ditu item bakoitzean. Funtzionamendu diferentzialaren ezean, erantzun eredu horiek berdinak izango dira. Egokitzapenaren alorrean aldentzen diren taldeak, probaren hizkuntzak zehazten ditu. Jatorri eta xede hizkuntzan proba eginarazi ondoren lorturiko datuen erkaketa da helburua.

5. BALIZKO ALBORAPENAREN ITURRIAK

Probak egokitzen direnean sortzen den funtzionamendu diferentziala detektatzea garrantzizkoa da, baina inportanteagoa iruditzen zaigu hortik ateratzen diren ondorioak. Hau da, funtzionamendu diferentzialaren balizko arazoak aurkitzea. Izan ere, egokitzapenarekin loturiko funtzionamendu diferentzialaren iturriak bilatuz bakarrik auresan eta murriztu daitezke arazoak izan ditzaketen itemak. Hori da lan honen muinetakoa. Gaztelerazko proba psikopedagogiakoak euskaratzean aurki genituzkeen balizko arriskuak antzematea. Funtzionamendu diferentzialaren zergatiak erabat zehaztu ezin badira ere, egindako lanen analisisiek zenbait argibide eman diezagukete, zeinen bidez eta beti ere ekarpen gehiagorekin, gorputz teoriko bat osatzen joan gintezken.

Egokitzapenarekin loturiko funtzionamendu diferentzialaren arazoietan sakonduz, lau iturri nagusi berezituko genituzke, nahiz askotan zaila den horien artean bereiztea, batak bestea ekartzen baitu; hala nola, itzulpenaren eskasia, hizkuntzen arteko desberdintasun gramatikalak, semantikoak eta berezitasun kulturala:

1. *Itzulpena*. Iturri honek itzulpen prozesuak berak barnera ditzaken erroreak, edo zuzentasun ezak hartzen ditu.

2. *Desberdintasun gramatikalak*. Hizkuntzen egitura gramatikalaren eraginez item baten egokitzapenak, horren ezaugarri psikometrikoak aldarazi ditzake, hizkuntza batean bestean baino errazago/zailago bihurtuz. Izan ere, hizkuntzaren berariazko ezaugarriak berak itemaren ebazpenean bide berriak zabaltzeaz gain itemaren formatua ere aldatu dezake. Beraz diferentzi *morfosintaktikoak* eta *formatuzkoak* bereiztuko genituzke.
3. *Baliokidetasun semantikoa*, hitzaren azpian dagoen kontzeptuaren ulermenari dagokio; Izan ere, hitz berdinak familiartasun, konkretzio maila eta balentzi afektibo diferenteak izan baititzake. Horiek guztiak direla eta, itzulpen linguistikoak baliokidetasun psikometrikoan eragiten du.
4. *Berezitasun kulturala*. Lehen aipaturiko *etic* eta *emic* bereizketari dagokio. Badaude kulturatik kulturara zuzenean eraman ezin diren kontzeptuak. Horren barne eta ikuspuntu zabal batetik aztertuz tokian tokiko ohiturak, egitura politikoak, ezaugarri geografikoak edo kultura errepresentazioak (sinboloak ere) sartuko genituzke. Tokian toki azertu beharko da gailentasun kultural nabaria duten itemek derrigorrezkoak diren proposaturiko xedeak heltzeko. Askotan interresgarria izaten baita kultura bakoitzeko bereziak diren itemak sortu eta mantentzea.

Metodoa

Horiek guztiak argitzeko «Bateria de Aptitudes diferenciales y generales E» (Yuste, 1988) osatzen duten bi azpiproba euskaratu eta azertu ditugu; «Habilidad mental general» eta «Comprensión verbal». Guztira, formatu dikotomikoa duten 70 item dira. Helburua aipaturiko alborapen iturrien adibideak eskaintzea da.

PARTAIDEAK

Lagina 9 eta 11 urte bitarteko 667 hurrez osatu dugu. Gaztelera-ko proba $N_G = 368$ partaidek betetzen dute ($\bar{X}_G = 50$; $S_G^2 = 54$) eta euskarazkoa berriz $N_E = 299$ ($\bar{X}_E = 38,35$; $S_E^2 = 45$). Bi batazbestekoen artean dagoen diferentzia esanguratsua da ($t = 21,088$; $p = 0,0001$); gaztelera-koen exekuzio orokorra euskarazkoena baino hobea da.

ANALISIA OROKORRAK

Probaren ezaugarri orokorretan, ez ditugu diferentzi esanguratsuak aurkitu bi bertsioen artean. Cronbach-en alpharen bidez estimaturiko fidagarritasun koefizienteak baliokidetzat har daitezke ($\alpha_G = 0,77$; $\alpha_E = 0,72$; Horien

diferentzia aztertzen duen Feld-ten estatistikoa $W = 1,21$, F banaketaren balio kritikoen barne dago $F_{0,975 (367, 298)} = 0,80$; $F_{0,025 (367, 298)} = 1,24$). *DIM-TEST* (Nandakumar eta Yu, 1996) prozedura erabiliz dimentsioanizkotasunaren hipotesia baztertu dugu bi kasutan, proba bakoitzaren azpian dimentsio nagusi bat dagoela onartuz ($T_G = 0,842$; $p = 0,199$; $T_E = -0,397$; $p = 0,654$). Beraz, konstrukto baliokidetasunaren hipotesia onartuz, baliokidetasun metrikoa arakatzeari ekiten diogu.

Itemen analisia. Funtzionamendu diferentziala

Funtzionamendu diferentziala aurkitzeko Mantel-Haenszel estatistikoaz lagundu gara. Lanaren zehar agertzen diren balioak MH-Deltarenak dira. Estatistiko hau 0-ra hurbiltzen denean funtzionamendu diferentzialik ez dagoenaren seinaleztat hartzen da. Balio negatiboak funtzionamendu diferentziala gaztelerazkoen aldekoa dela adierazten du; eta kantitate positiboak itemak euskarazko bertsioaren alde jokatzeko duela islatzen du.

Proba zeregin desberdinez osaturik dagoenez, interesgarria litzake funtzionamendu diferentziala eta ataza motaren arteko erlazioa aztertzea. Horretarako itemak 1. taulak agertzen dituen sailetan banatu ditugu. Bertan ataza bakoitzean aurkitu den IFD kopurua irakur daiteke.

1. taula

Ataza eta funtzionamendu diferentziala

Ataza mota	Itemak	I.F.D.	%	Gazteleraz	%	Euskaraz	%
Sinonimoak	6	5	83,3	4	80	1	25
Antonimoak	4	3	75	2	66,6	1	33,3
Erabilerak komunena	2	1	50	1	100		
Analogiak	13	7	53,8	5	71,42	2	28,57
Definizioak	5	3	60	3	100		
Hitzak ordenatu							
esaldi bat osatzeko	6	3	50	2	66,6	1	33,3
Hitz zerrenda baten							
ezaugarria bilatu	8	2	25			2	100
Ezaugarri baten							
konstantzia	9	2	22,2	2	100		
Zenbakizko serie							
logikoak	7	5	71,42			5	100
Problemak	7	3	42,85			3	100
Toki-denborazko							
problemak	3	1	33,33			1	100
Guztira	70	35	50	19	57,57	14	42,42

Item guztiek ez dute baliokidetasun metrikoaren baldintza betetzen; izan ere erdiak funtzionamendu diferentziala agertzen baitu. Hau da, gazteleraz eta euskarazko itemen barne ezaugarriak ez dira berdinak. Horretan sakonduz funtzionamendu diferentzian norabidea bereiztu dugu. Azkeneko zutabetan agertzen da informazio hori. Adibidez, sinonimoak deritzan atazan 5 itemek funtzionamendu diferentziala dute; horietatik %80ak gaztelerazko laginaren alde jokatzen du, eta bakar batek (%25) proba euskaraz burutzen dutenen alde.

Itemaren funtzionamendu diferentzialaren norabidea eta ataza motaren arteko erlazioa esanguratsua da ($\chi^2 = 20,64$; $p \leq 0,024$). Badago lotura funtzionamendu diferentzialaren norabidea eta item motaren artean. Erlazio hori es-tuagoa bihurtzen da itemen sailkapena zabaltzen bada. Ahozko gaitasuna eta zenbakizkoa eskatzen dutenen itemak bereiztuz, ondokoak lortuko genituzke:

2. taula

Ahozko eta zenbakizko itemen funtzionamendu diferentziala

Item mota	Kopurua	I.F.D.	Portzentaia	Gazteleraz	%	Euskaraz	%
Ahozkoak	43	26	60,46	19	73	7	26,9
Zenbakizkoak	17	9	52,94	0		9	100
Guztira	70	35	50	19	54,28	16	45,71

Taula honetan argiago azaltzen da item mota eta funtzionamendu diferentzialaren norabidearen arteko erlazioa (Kontingentzi koefizientea = 0,540: $p < 0,0001$). Zenbakizko itemetatik funtzionamendu diferentzial guztiak euskarazko bertsioaren alde daude. Hau harrigarria izan daiteke itemen edukia ikusirik, ez baitute gehienek ahozko osagairik (*¿Qué número sigue? 70-60-50-40*). Emaitza hauek interpretatzeko, besteak beste, elebidunei egokitzen zaien zalutasun kognitiboaren hipotesia aztertu beharko genuke (Cummins eta Gulutsan, 1974). Lan honi dagozkionetan ordea, itzulpenari loturiko arazoak interesatzen zaizkigunez, ahozko itemen azterketari ekingo diogu.

Ahozko item guztietatik %60ak funtzionamendu diferentziala du. Horietatik %73 gaztelerazko bertsioaren alde eta %26,9 euskarazkoaren alde. Horietan oinarrituz, ekarri ditugu ondoko adibideak.

a) ITZULPEN ARAZOAK

Itzulpenaren kalitatea aztertzeke berariazko proba sortu ondoren, hiru itzultzaileei gazteleraz eta euskarazko itemen arteko berdintasuna ebaluatzeke eskatzen zaie. Horretarako item bakoitzari 5 puntuzko eskala bat egokitzen zaio, non partaide bakoitzak itzulpenaren egokitasuna ebaluatuko duen; bai enuntziatuari bai eskainitako aukera bakoitzari dagozkionetan.

Ebaluatu ezazu itzulpenaren kalitatea 1etik 5erako eskalan.

5. Baliokidetasun osoa
4. Itzulpen onargarria
3. Nolabaiteko desberdintasuna
2. Itzulpen arazoak
1. Baliokidetasun gabezi nabariak

Oro har item guztietan lortu diren balioak altuak izan arren, (batzubes-tekoa 4,8koa da) badago item bat guztira (enuntziatua eta aukerak) 25eko balioa gainditzen ez duena. Ondokoa hain zuzen,

VACA es a MUGIR como ELEFANTE a.....

- a) *Rugir* c) *Trotar* e) *Berrear*
b) *Barritar* d) *Pezuñas*

BEHIA BAURREZ egitearekiko, LEHOIA hautetariko baterako dena da:

- a) *Orro egin* c) *Takatata ibili* e) *Irrintzi egin*
b) *Bee egin* d) *Apoak*

Galdera hau gaztelerazko bertsioan %71,20ek ongi erantzuten duen bi-tartean, euskarazkoan %5ak bakarrik heltzen dio. Lortutako funtzionamen-du diferentzialaren balioa MH-delta = 47,06koa da.

Item hau euskaratzean *BARRITAR* ahotsarentzat elementu baliokiderik aurkitu ez zenez, enuntziatua eta aukerak aldatu egin ziren. Aukeraturikoak ez dira behar bada, egokienak.

Horretaz gain, enuntziatua bera ere zalantzan jarriko genuke, zailagoa baita euskaraturiko bertsioaren ulermena. Egokiagoa litzake beste enuntzia-tu sinpleagoa bilatzea, esaterako. «Behiak baurre egiten du eta lehoiak...». Honekin, ataza askoz ere errazago ulertzen da, eta erantzun aukeretan «egin» azaltzea saihesten da.

b) GRAMATIKA DESBERDINTASUNAK

b.1. Morfosintaktikoak

Desberdintasun morfosintaktikoen barnean bi adibide ekarri ditugu. Bata atzizikiekin erlazonaturikoa eta bestea berriz esaldien antolakuntzarekin.

Atzizkien arazoa

¿Qué palabra sobra? CEREZO-GUINDO-PERAL-MATORRAL-NARANJO-MANZANO

- a) *Cerezo* c) *Guindo* e) *Naranja*
b) *Manzano* d) *Peral* f) *Matorral*

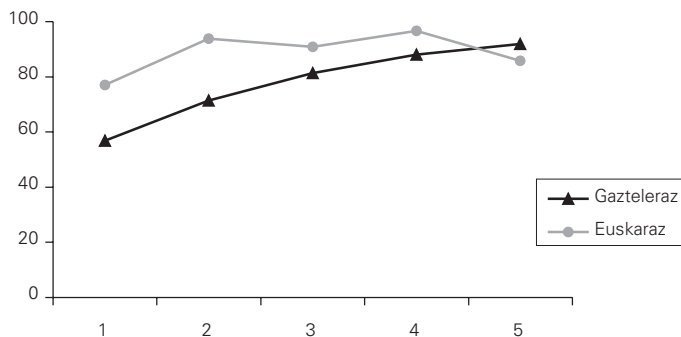
Zer hitz dago soberan? GERIZONDOA-BASAGERIZONDOA-UDAREONDOA-SASTRAKA-LARANJONDOA-SAGARRONDOA

- a) *Gerizondoa* c) *Basagerizondoa* e) *Laranjondoa*
 b) *Sagarrondoa* d) *Udareondoa* f) *Sastraka*

Ikasleak kontzeptuak sailkatuko dituen irizpidea bilatu behar du ataza honetan. Euskarazko bertsioan ordea lana errazten zaio ikasleari. Erantzun zuzena aukeratzekoan —*ondo* atzizkiak argibide orientagarria ematen diolako, hau da, edukiaz gain alderdi gramatikalean ere diferentziak daude aukera zuzena eta okerren artean. Item honen erantzun zuzenen portzentaia gaztelerazko bertsioan %85,33a den bitartean, euskarazkoan %86koa da. Diferentzi horiek esanguratsuak ez badirudite ere, kontuan hartu behar da, funtzionamendu diferentziala ez dela erantzun zuzenen arteko diferentzia. Kontuz. Funtzionamendu diferentziala aztertzeke, erantzun ereduak mailaz-maila aztertzen dira. Gure kasuan MH-deltaren balioa 3,14koa da.

Hizkuntzen arteko diferentziak argiago ikusteko, puntuazio osoaren arabera bost sailetan banatu ditugu haurrak. Hurrengo taulak, mailaz maila erantzun zuzenen portzentajeak jasotzen ditu. Irudiak berriz, zailtasun indize baldintzatu horien adierazpen grafikoa islatzen du, non, mailaz maila hizkuntzen arteko diferentziak antzeman daitezken.

	Mailak				
	1	2	3	4	5
Gazteleraz	56,3	71,9	81,3	87,9	91,1
Euskaraz	76,9	93,8	90,8	96,6	85,7



Sintaxi arazoak

Hizkuntza bakoitzaren esaldien antolakuntza sistema berariazkoak ere funtzionamendu diferentzialaren iturri izan daitezke. Azterturiko zenbait

itemek haurri eskatzen dioten lana, enuntziatuan agertzen diren hitzen ordena aldatu eta balizko aukeren artean esaldi zuzena bukatzeko bat hautatzea da. Item hauek euskaratzerakoan bi dira gaztelerazkoarekiko dituzten diferentziak; batetik ez da enuntziatuan hitz kopuru berdina agertzen, eta bestetik esaldia osatzeko aukeratu beharreko hitza ez da esaldiaren bukaeran geratuko. Honen eraginez, itemaren ebazpen zuzenak eskatzen dituen lanak areagotu egiten dira euskaraturiko proban.

Ordena las palabras y contesta la frase correcta: DE-UNA-CONJUNTO-PALABRA-ES-UN

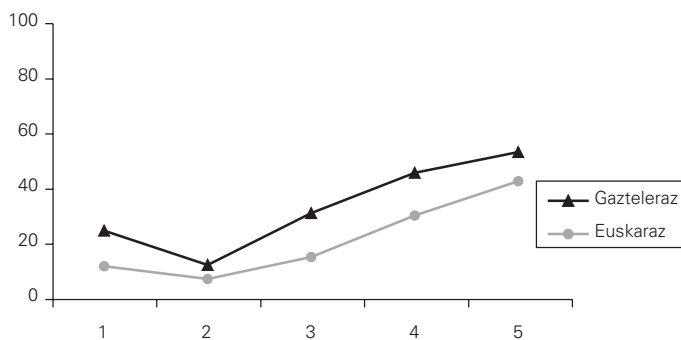
- a) Frases c) Signos e) Oraciones
 b) Palabras d) Vocabulario f) Letras

Ordena itzazu hitzok eta aukeratu erantzun zuzena: BAT-MULTZO-HITZ-BAT-DA

- a) Esaldia c) Zeinu e) Perpaus
 b) Hitza d) Hiztegi f) Letra

Item honen zailtasun indizea gaztelerazko bertsioan 0,43koa den bitartean, euskarazkoan 0,11ra jaisten da. Funtzionamendu diferentzialaren indizea den MHDeltaren balioa -22,91koa da. Erantzun proportzioak maila maila agertuz, ondoko taula eta irudiak osatuko genituzke.

	Mailak				
	1	2	3	4	5
Gazteleraz	25	12,5	31,3	45,9	53,5
Euskaraz	12	7,4	15,4	30,4	42,9



Baina item mota honek ez du beti euskarazkoen aurka jokatzeko. Ataza berdina eskatzen duen hurrengo honetan, funtzionamendu diferentziala an-

tzeman badugu ere, hori euskarazko bertsioaren alde dago. Arrazoa izena eta izenlagunaren arteko komunztadura beharrea aurkitzen dugu, zeinek balizko aukerak murriztu egiten dituen.

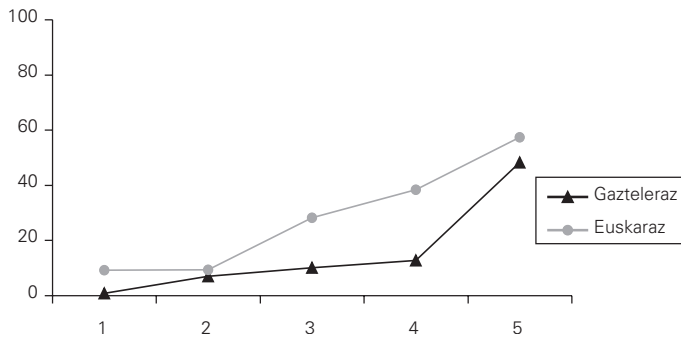
Ordena las palabras y contesta la frase correcta: SON-Y-BLANCO-PALABRAS-NEGRO

- a) *De distinta raza* c) *Contradictorias* e) *Del lenguaje*
 b) *Sinonimas* d) *La noche y el día* f) *Opuestas*

Ordena itzazu hitzik eta aukeratu erantzun zuzena: DIRA-ETA-ZURI-HITZAK-BELTZ

- a) *Arraza desberdinetako* c) *Kontraesankor* e) *Hizkuntzako*
 b) *Sinonimoak* d) *Gaua eta Eguna* f) *Aurkako*

	Mailak				
	1	2	3	4	5
Gazteleraz	—	6,3	9,4	12,1	47,8
Euskaraz	8,5	8,6	27,7	37,9	57,12



Item honen zailtasun indizeak gazteleraz eta euskaraz 0,25 eta 0,16 dira. Ikusten denez badirudi kontraesana dagoela erantzun zuzenen proportzio orokorrak eta funtzionamendu diferentzialaren artean. Horregatik derri-gorrezkoa da hori ebaluatzeko indikatzaile baldintzatuak erabiltzea. Funtzionamendu diferentzialaren MHDELTA-ren balioa 2,16koa da.

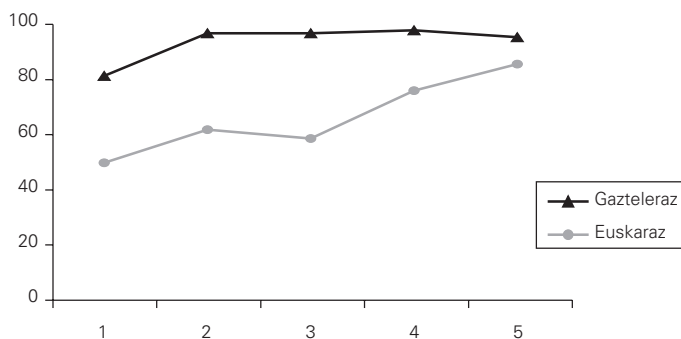
b.2. *Formatu aldaketa*

Formatuaz hitz egiterakoan, esaldien luzera, hitz kopurua, enuntziatuaren hitz garrantzizkoenen arteko distantzia, eta antzerako aspektuetaz ari gara. Esaterako, enuntziatua eta balizko erantzun guztien arteko komunzta-

Jarraian dauden hitzetatik zer da ELEKTRATRESNA?

- a) *Hozkailua* c) *Aulkiak* e) *Fotomakina*
 b) *Autoa* d) *Mahai bat*

	Mailak				
	1	2	3	4	5
Gaztelera	81,3	96,9	96,9	98	95,5
Euskaraz	49,6	61,7	58,5	75,9	85,7



Item honen erantzun zuzenen proportzioak gaztelera eta euskaraz 0,96 eta 0,58koak dira. Funtzionamendu diferentzialaren balioa berriz -5,79. Mailaz mailako azterketan argi ikusten da gaztelera haurrentzat nahiz gaitasun maila baxua izan, itema erraza den. Euskarazkoentzat aldiz, gaitasun maila gorenetan bakarrik heltzen da erantzun zuzenen portzentaia %80ra.

c.2. Polisemia

Itzulpenaren eraginez polisemia duten hitzak azaldu daitezke, itemaren ebazpena zailduz.

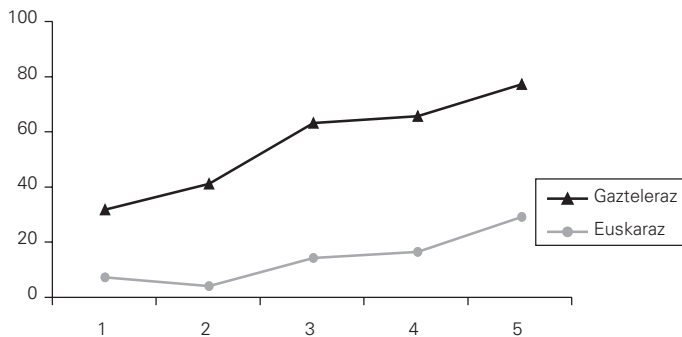
ANDAR es a PIES como TIRAR a

- a) *Hombres* c) *Manos* e) *Basura*
 b) *Piernas* d) *Dedos*

IBILI OINETARAKO, BOTA hauetariko baterako dena da:

- a) *Gizonak* b) *Eskuak* e) *Zaborra*
 b) *Hankak* d) *Behatzak*

	Mailak				
	1	2	3	4	5
Gazteleraz	31,3	40,6	62,5	64,9	76,4
Euskaraz	6,8	3,7	13,8	16	28,6



Zailtasun indizeak gazteleraz eta euskaraz 0,65 eta 0,07koak dira. MH-deltaren balioa 14,21koa da. Emaitzen arrazoiak bilatzerakoan BOTA hitzaren polisemia aipatuko genuke. Batetik «tirar» aditzaren sinonimotzat har daiteke, baina bestetik itemaren edukiak ez du baztertzen «BOTA» zapata motatzat hartzea. «Jaurti» edo «Jaurtiki» egokiagoak lirateke adibide honetan.

6. ONDORIOAK

Proben euskaratze egokiak beraz, ondoko pausoak jarraitu behar dituelakoan gaude.

1. *Egokituriko probaren aukera.* Egokitzapen zuzenaren prozesua erraza eta merkea ez denez, lehen puntua balizko testen artean psikometrikoki egokiena hautatzea litzake. Gutxienezko fidagarritasun eta baliagarritasun baldintzak betetzen ez dituen jatorrizko test batetik egokituriko bertsio zuzena lortzea alferrikako xedea da.

Egun, gazteleraz komertzializaturiko proben ebaluazio sistematikorik ez badago ere, COP-ek (*Colegio Oficial de Psicólogos*) helburu hori beteiko duen proiektuan dihardu. Horretarako, testen ebaluazio sistematikoa gauzatzeko eredu ingelesa eta belgikarrari darraio (Bartram, Burke, Kandola, Lindley, Marxhall eta Rasch, 1997; Evers, 1996) CET probaren bidez (*Cuestionario para la evaluación de los tests*; Prieto eta Muñiz, 2000).

2. *Probaren egokitzapena.* Atal honetan, funtzionamendu diferentzialaren iturriak zaindu behar dira bereziki. Kontutan hartzekoak dira probaren azken hartzaileen ezaugarriak, eta horiei egokitu bai sintaxia bai lexikoa.

Ikusitako puntuen arabera arazoak izan ditzaketen itemak aurreikustea funtsezkoa da. Hala jardunez, balizko arazoak dituztenekiko item berriak sor daitezke. Hauen azterketa kuantitatiboak lagunduko du horien egokitasunari buruzko ondorioak ateratzen.

Prozesuaren aldi honetan, komenigarria litzake hitzen dimentsio semantikoak aztertzen dituzten ikerketa normatiboen laguntza izatea. Hauek, besteak beste, hizkuntzen arteko afinitate indizeak emango lizkigukete, zeinetan oinarritu gintezken egokitzapenean. Horien gabezien aurrean, derrigorrezkoa da ebaluatu nahi den jakintza eremuan adituen laguntza. Helburua baliokidetasun literala gainditzea litzake, komunztadura semantikoa bilatuz. Arau normatibo zehatzak, hitzen maiztasun tasetan oinarriturikoak, besteak beste, finkatzen ez diren bitartean egokitzapen psikologikoak linguistikoki zuzenak izanagatik ez dute beti baliokidetasun semantikoa lortuko, eta ondorioz, ezta metrikoa ere.

3. *Probaren aplikazioa*. Xede populaziotik lagin egokia hautatu eta egokituriko proba eginarazten zaio.

4. *Fidagarritasunaren azterketa*. Probaren alderdi formalen arteko baliokidetasuna aztertu beharko litzake lehenik. Horretarako badaude tinkotasun koefizienteak bi populaziotan hartzen dituen balioen arteko berdintasuna aztertzen dituzten hipotesi testak (Feldt, 1969).

5. *Egiturazko baliokidetasuna*. Testaren barne egitura bi taldetan berdina dela ziurtatzea ezinbestekoa da. Horretarako ezkutuko egitura aztertzen dituzten ereduak erabili ohi dira, faktore analisisa nahiz egitura ekuazioak. Behin konstruktuen arteko berdintasunaren hipotesia deuseztatu ezin dugula, hurrengo pausoarekin jarraituko genuke.

6. *Baliokidetasun metrikoa*. *Itemaren funtzionamendu diferentzialaren detekzioa*. Itemaren funtzionamendu diferentzialaren detekzioaren ostean, horren arrazoiak aurkitu behar dira, alborapena dagoen ala ez dagoen ebaluatuz.

7. *Puntuazioen berdinketa*. Pauso horiek guztiak ongi bete arren, zenbait itemek funtzionamendu diferentziala izan dezakete. Horien iturria berezitasun kulturala denean, aztertu egin beharko da item horiek mantentzea komenigarria den finkaturiko xedeak betetzeko. Hala bada, kulturalki egokituriko probez hitz egiten dugu. Horien aurrean, puntuazioak aldendu nahi direnetan, psikometriak badu tresneria egokia probak berdintzeko; horretarako nahikoa da zenbait item komunak izatea.

Jasotze-data: 2003-10-28

Onartze-data: 2003-11-10

Abstract

This study illustrates the problems related to the adaptation of psycho-pedagogic tests from both a general perspective and as applied to the Spanish-Basque situation. From the first point of view,

the different levels of equivalence which may be found in the adaptation of the tests: construct, metric and related to the method, are described, with special attention being paid to the second. The methods of judgement and the differential functioning of the item as methods of analysis of the metric equivalence are described. The rules for the adaptation of the tests accepted by the International Test Commission are set out. In relation to the adaptation of Spanish-Basque tests, the different potential sources of differential functioning of the items are systemised, differentiating between four possible causes.; problems of translation, lack of grammatical similarity and semantics between languages, and cultural relevance. Within the second, special attention is paid to the syntactic structures and the differences of format. The article sets out an empirical example from each one of these potential sources of bias, based on the results obtained following the adaptation of a general aptitude test to Basque.

Keywords: *Evaluation. Test. Equivalence. Differential Working. Bias.*

Este trabajo muestra los problemas relacionados con la adaptación de tests psicopedagógicos, tanto desde una perspectiva general como aplicada al entorno castellano-euskera. Desde el primer punto de vista, se describen los distintos niveles de equivalencia que pueden encontrarse en la adaptación de tests: constructo, métrica y relacionada con el método, haciendo especial hincapié en la segunda. Se describen los métodos de juicio y el funcionamiento diferencial del ítem como procedimientos de análisis de la equivalencia métrica. Se exponen las normas para la adaptación de tests aceptadas por la International Test Commission. Con relación a la adaptación de tests castellano-euskera, se sistematizan las diferentes fuentes potenciales de funcionamiento diferencial de los ítems, diferenciándose entre cuatro posibles causas; problemas de traducción, ausencia de igualdad gramatical y semántica entre idiomas, y relevancia cultural. Dentro de la segunda, se presta especial atención a las estructuras morfosintácticas y a las diferencias de formato. El artículo expone un ejemplo empírico de cada una de estas fuentes potenciales de sesgo, basándose en los resultados obtenidos tras la adaptación al euskera una prueba de aptitud general.

Palabras clave: *Evaluación. Test. Equivalencia. Funcionamiento Diferencial. Sesgo.*

Ce travail met en évidence les problèmes liés à l'adaptation de tests psychopédagogiques, aussi bien depuis une perspective générale que d'une perspective appliquée à l'environnement castillan-euskera. Depuis la perspective générale les différents niveaux d'équivalence que l'on peut trouver dans l'adaptation de tests sont décrits: construct, métrique et rapport avec la méthode, tout en insistant particulièrement

sur la deuxième. Les méthodes de jugement et le fonctionnement différentiel de l'item en tant que processus d'analyse de l'équivalence métrique sont également décrits. Les normes pour l'adaptation de tests acceptés par l'International Test Commission sont exposées. Quant à l'adaptation de tests castillan-euskera, les différentes sources potentielles de fonctionnement différentiel des items sont systématisées, se séparant en quatre causes possibles; les problèmes de traduction, le manque d'équivalence grammaticale et sémantique entre les langues, et la pertinence culturelle. Dans la deuxième, les structures morphosyntaxiques et les différences de format sont particulièrement intéressantes. L'article expose un exemple empirique de chacune de ces sources potentielles de déviation, se basant sur les résultats obtenus suite à l'adaptation en euskera d'une épreuve d'aptitude générale.

Mots clé: Évaluation. Test. Équivalence. Fonctionnement Différentiel. Biais.

BIBLIOGRAFIA

- BARTRAM, D.; BURKE, E.; KANDOLA, R.; LINDLEY, P.; MARSHALL, L. eta RASCH, P. (1997): *Review of tests of ability and aptitude (level A) for use in occupational setting*. Leicester: BPS Books.
- BRISLIN, R.W. (1970): «Back-translation for cross-cultural research». *Journal of cross-cultural psychology*, 1(3), 185-216.
- CAMILLI, G. eta SHEPARD, L.A. (1994): *Methods for identifying biased test items*. London: Sage.
- ELOSUA, P. (2003): «Sobre la validez de los tests». *Psicothema* 15(2), 315-321.
- ELOSUA, P. eta LÓPEZ, A. (1999a): «Funcionamiento diferencial de los ítems y sesgo en la adaptación de dos pruebas verbales». *Psicológica*, 20, 23-40.
- ELOSUA, P. eta LÓPEZ, A. (1999b): «Neurketa tresna psikopedagogikoen euskaraztea eta itemaren funtzionamendu diferentziala». *Tantak*, 21, 45-58.
- ELOSUA, P.; LÓPEZ, A. eta EGAÑA, J. (2000): «Idioma de aplicación y rendimiento en una prueba de comprensión verbal». *Psicothema*, 12(2), 201-206.
- ELOSUA, P.; LÓPEZ, A.; EGAÑA, J.; ARTAMENDI, J.A. eta YENES, F. (2000): «Funcionamiento diferencial de los ítems en la aplicación de pruebas psicológicas en entornos bilingües». *Revista de Metodología de las Ciencias del Comportamiento*, 2(1), 17-33.
- ELOSUA, P.; LÓPEZ, A. eta TORRES, A. (1999): «Adaptación al euskera de una prueba de inteligencia verbal». *Psicothema*, 11(1), 151-161.
- EVERS, A. (1996): «Regulations concerning test qualifications and test use in The Netherlands». *European Journal of Psychological Assessment*, 12, 153-159.
- FELDT, L.S. (1969): «A test of the hypothesis that Cronbach's Alpha or Kuder-Richardson coefficient twenty is the same for two test». *Psychometrika*, 34, 363-373.
- FINK, R. (1963): «Interviewer training and supervision in a survey of Laos». *International Social Science Journal*, 15, 21-34.
- GRYSAY, A. (1998): *Instructions for the translation of the PISA material (OECD/PISA report)*. Melbourne, Australia: Australian Council for Educational Research.

- GRYSAY, A. (1999): *Report on the development of the French source version of the PISA test material (OECD/PISA report)*. Melbourne, Australia: Australian Council for Educational Research.
- HAMBLETON, R.K. (2001): «The next generation of the ITC test translation and adaptation guidelines». *European Journal of Psychological Assessment*, 17(3), 164-172.
- HAMBLETON, R.K. eta PATSULA, L. (1999): «Increasing the validity of Adapted tests: Myths to be avoided and guidelines for improving test adaptation practices». *Journal of Applied Testing Technology*, 1(1).
- HOLLAND, P.W. eta WAINER, H. (Ar.). (1993): *Differential item functioning*. Hillsdale: Lawrence Erlbaum Associates.
- HUI, C.H. eta TRIANDIS, H.C. (1989): «Effects of culture and response format on extreme response style». *Journal of Cross-Cultural Psychology*, 18(2), 115-142.
- JENSEN, A.R. (1980): *Bias in mental testing*. New York: Free Press.
- MITCHELL, R. (1966): *The problems and possibilities of measuring social attitudes in african social surveys*. Paper presented at The Ninth Annual Meeting of The African Studies Association, Bloomington, Indiana.
- NANDAKUMAR, R. eta YU, F. (1996): «Empirical validation of DIMTEST on nonnormal ability distributions». *Journal of educational measurement*, 33(3), 355-368.
- PRIETO, G. eta MUÑIZ, J. (2000): «Un modelo para evaluar la calidad de los tests utilizados en España». *Papeles del Psicólogo*, 77, 65-75.
- SCHUMAN, H. (1966): «The random probe: A technique for evaluating the quality of closed questions». *American sociological review*, 31, 218-222.
- SUPER, C.M. (1983): «Cultural variation in the meaning and uses of children's "intelligence"». Non J.B. DEREGOWSKIETA, R.C. ANNIS (Arg.): *Explorations in cross-cultural psychology* (orr. 199-212). Lisse, Netherlands: Swets.
- VAN DE VIJVER, F.J.R. eta LEUNG, K. (1997): *Methods and data analysis for cross-cultural research*. Newbury Park, CA: Sage.
- VAN DE VIJVER, F.J.R. eta POORTINGA, Y.H. (1997): «Towards an integrated analysis of bias in cross-cultural assessment». *European Journal of Psychological Assessment*, 13, 29-37.
- VAN DE VIJVER, F.J.R. eta TANZER, N.K. (1997): «Bias and equivalence in cross-cultural assessment: An overview». *European Review of Applied Psychology*, 47(4), 263-279.
- WERNER, O. eta CAMPBELL, D. (1970): «Translating, working through interpreters, and the problem of decentering». Non R. NAROLLETA, R. COHEN (Arg.): *A handbook of method in cultural anthropology* (orr. 398-420). New York: Natural History Press.
- YUSTE, C. (1988): *BADYG-E*. Madrid: Ciencias de la Educación Preescolar y Especial.

ERANSKINA

Testen egokitzapenerako gomendioak

Testuingurua

C.1. Ikasketako helburu nagusietarako garrantzitsuak ez diren ezberdintasun kulturalen efektuak, ahal den heinean, gutxitu beharko lirateke.

C.2. Interes gune diren populazioetan, konstruktuen gainjartze maila ebaluatu beharko litzateke.

Testaren eraketa eta egokitzapena

D.1. Testen sortzaileek egokitzen prozesuan, proben bertsio egokituak helburutzat izango dituzten populazioen ezberdintasun linguistikoak eta kulturalak kontutan hartu direla ziurtatu beharko lukete.

D.2. Testen sortzaileek/argitaratzaileek, argibideetan, item beraietan eta testaren manualean erabilitako hizkuntza, testa zuzenduta dagoen populazio kultural eta idiomatiko guztientzako, egokia dela ziurtatzen duten datuak eskaini beharko lituzkete.

D.3. Testen sortzaileek/argitaratzaileek hautatutako ebaluaketa teknikak, itemen formatuak, testen arauak eta prozedurak zuzenduta dauden populazio guztiei ezagunak egiten zaizkieneko frogak aurkeztu beharko lituzkete.

D.4. Testen sortzaileek/argitaratzaileek itemen edukiak eta estimuluen materialak zuzenduta dauden populazio guztiei ezagunak egiten zaizkieneko frogak aurkeztu beharko lituzkete.

D.5. Testen sortzaileek/argitaratzaileek justifikazio arrazional sistematikoa, bai linguistikoa, bai psikologikoa, aurkeztu beharko lukete egokitzapen prozesuaren zehaztasuna hobetzeko; era berean, hizkuntza ezberdinetan egindako bertsio ezberdin guztien ekibalentziaren inguruko datuak bildu beharko lituzkete.

D.6. Testen sortzaileek/argitaratzaileek, datuen bilketarako diseinuak, testaren bertsio idiomatiko ezberdinetako itemen arteko baliokidetasuna zehazteko aproposak diren teknika estatistiko egokien erabilera ahalbidetzen duela ziurtatu beharko lukete.

D.7. Testen sortzaileek/argitaratzaileek egokiak diren teknika estatistikoak erabili beharko lituzkete 1) test baten bertsio ezberdinen arteko ekibalentzia zehazteko eta 2) testa zuzenduta dagoen populazioaren batetarako osagai arazotsuak edo ez-egokiak izan daitezkeen testaren alderdiak identifikatzeko.

D.8. Testen sortzaileek/argitaratzaileek egokitutako testa, zuzenduta dagoen populazio-helburu guztietan, baliotasunaren ebaluazioaren inguruko informazioa eskaini beharko lukete.

D.9. Testen sortzaileek/argitaratzaileek testaren ekibalentziaren inguruko datu estatistikoak eman beharko lituzkete, zuzenduta dauden populazio guztien arabera.

D.10. Eskala komuna eratzen ari denean edo populazio ezberdinak erkatzen direnean, populazio ezberdin horietara zuzendutako bertsioetan baliokideak ez diren galderak ez dira erabili behar. Hala ere, populazio bakoitzaren puntuazioen (isolatuki) eduki baliotasuna indartzeko baliagarriak izan daitezke.

Aplikazioa

A.1. Testen eratzailleak eta aplikatzaileek ager daitezkeen arazoak aurreikusten, eta horiek saihesteko neurri egokiak hartzen, material eta argibide egokiak prestatzen, saiatu beharko lirateke.

A.2. Testak aplikatzen dituztenak, estimuluetarako erabilitako materialekin, aplikazio prozedurekin eta erantzun erarekin erlazionatuta dauden hainbat faktoreekiko sentikorrak izan beharko lirateke, horiek puntuazioetatik ateratako inferentzien baliotasuna gutxitu baitezakete.

A.3. Testaren aplikazioan eragina duten ingurunearen alderdiak, testa zuzenduta dagoen populazio guztietarako, ahalik eta antzekoen mantendu beharko lirateke.

A.4. Testaren aplikaziorako argibideek, jatorrizko hizkuntzan eta helburukoan, desiratzen ez diren aldakuntza iturrien eragina gutxitu behar dute.

A.5. Testaren manualak, testuinguru berri batetan erabiltzean, berrikusi beharko liratekeen testa beraren eta bere aplikazioaren alderdi guztiak zehaztu beharko lituzke.

A.6. Aplikatzaileak ezin du interferitu, hark azterketariengan izan dezakeen eragina murriztu behar delarik. Testaren manulean deskribatutako arau zehatzak hitzez hitz jarraitu behar dira.

Dokumentazioa/Puntuazioen interpretazioa

I.1. Test bat beste populazio batetan erabiltzeko egokitzen denean, aldaketen inguruko dokumentazioa eskaini behar da, baita bertsioren arteko baliotasunaren inguruko datuak ere.

I.2. Testa aplikatu zitzaieneko laginek lortutako puntuazioen arteko ezberdintasunak ezin dira zuzenean horrela hartuak izan. Ikertzaileak ezberdintasunak beste eremu enpirikoekin substantziatearen ardura du.

I.3. Populazioen arteko konparaketak puntuazioak adierazten diren eskalarako zehaztutako inbariantzaren mailan egin daitezke bakarrik.

I.4. Testaren eratzailleak, testuinguru soziokultural eta ekologiko ezberdinek, populazioek testean erakutsitako errendimenduaren gain eragina izateko duten forma ezberdinen inguruko informazio zehatza eskaini beharko luke eta emaitzen interpretazioan efektu hauek kontutan izateko prozedurak sujeritu beharko lituzke.