

I ● DIO

Izrada, evaluacija i dokumentacija
testova

1. VALJANOST

Uvodna objašnjenja

Valjanost se odnosi na stupanj u kojem podaci i teorija podržavaju interpretacije testovnih rezultata koje proizlaze iz predloženih upotreba testova. Valjanost je stoga najvažnije pitanje u razvoju i evaluaciji testova. Proces validacije uključuje prikupljanje podataka koji pružaju čvrstu znanstvenu osnovicu zamišljenih interpretacija testovnih rezultata. Ono što se u ovom kontekstu vrednuje upravo su interpretacije nužne za predložene upotrebe, a ne sam test. Kada se testovni rezultati koriste ili tumače na više načina, svako od tih tumačenja mora se posebno validirati.

Validacija logički počinje eksplicitnim određenjem interpretacije testovnih rezultata koje treba izložiti zajedno s obrazloženjem njezine relevantnosti za predloženu upotrebu testa. Interpretacija se odnosi na konstrukt ili pojam koji se testom želi mjeriti. Primjeri konstrukata su matematičko postignuće, učinak kompjutorskog tehničara, depresija, samopoštovanje i sl. Da bi se opravdao razvoj testa, predložena interpretacija se obrazlaže opisivanjem njezina obuhvata i širine te označavanjem aspekata konstrukta koji se trebaju zahvatiti. Detaljni opis racionalne treba sadržavati konceptualni okvir testa te prikaz znanja, vještina, sposobnosti, procesa ili karakteristika koje se žele procjenjivati. Konceptualni okvir treba pokazati kako se ta reprezentacija konstrukta razlikuje od drugih konstrukata i kako ona treba biti povezana s drugim varijablama.

Konceptualni okvir djelomično je određen i načinom na koji će se upotrebljavati testovni rezultati. Na primjer, test matematičkog postignuća može služiti za razvrstavanje polaznika u odgovarajuće obrazovne programe, za dobivanje školske diplome ili kao dio fakultetskog prijemnog ispita. Svaka od ovih različitih upotreba podrazumijeva i ponešto različito shvaćanje rezultata u testu matematičkog postignuća: da će polaznik imati koristi od određene nastave, da je ovladao određenim programom, odnosno da će vjerojatno biti uspješan na fakultetskoj razini rada. Slično, test samopoštovanja može se upotrebljavati u psihološkom savjetovanju, prilikom odlučivanja o zapošljavanju ili za temeljnu znanstvenu svrhu proučavanja konstrukta samopoštovanja. Svaka od navedenih potencijalnih upotreba determinira specifični okvir i interpretacije testovnih rezultata, a također ima implikacija i na izradu testa te njegovu evaluaciju.

Validacija se može promatrati kao prikupljanje uvjerljive znanstvene argumentacije koja opravdava planiranu interpretaciju testovnih rezultata i njihovu relevantnost za predviđene svrhe. Konceptualni okvir upućuje na vrste podataka koje se mogu prikupljati da bi se ocijenile planirane interpretacije u svjetlu svrhe testiranja. Kako validacija napreduje i prikupljaju se novi podaci o značenju testovnih rezultata, može se pokazati potreba za mijenjanjem testa, njegova konceptualnog okvira, pa čak i konstrukta koji se želi mjeriti.

Velika raznolikost testova i okolnosti njihove upotrebe razlozi su zbog kojih su neke vrste podataka u određenom slučaju ključne, dok su druge manje korisne. Odluka o tome koji su tipovi podataka važni za neku validaciju može se pojasniti razvojem skupa pretpostavki koje opravdavaju predloženu interpretaciju za određenu svrhu testiranja. Na primjer, ako se test matematičkog postignuća koristi za procjenu spremnosti polaznika za sudjelovanje u višoj razini programa, potrebno je podacima potkrijepiti sljedeće pretpostavke: (a) da su neke vještine preduvjet za višu razinu programa, (b) da je sadržaj testa adekvatan za procjenu tih vještina, (c) da se testovni rezultati mogu generalizirati na relevantne skupove zadataka, (d) da testovni rezultati nisu pretjerano opterećeni utjecajem “pomoćnih” varijabli kao što je npr. vještina pisanja, (e) da se uspjeh u višoj razini programa može valjano procijeniti i (f) da će ispitanici s višim rezultatima u testu biti uspješniji u višoj razini programa od ispitanika s nižim rezultatima. Primjeri pretpostavki u drugim kontekstima testiranja mogu, na primjer, uključivati očekivanje da ispitanici s višim pokazateljima opće anksioznosti doživljavaju značajnu tjeskobu u određenom rasponu situacija, da su rezultati djece na testu inteligencije čvrsto povezani s njihovim školskim uspjehom, ili da neki sklopovi rezultata u neuropsihološkoj bateriji testova upućuju na oštećenje uzrokovano ozljedom mozga. Validacijski proces napreduje u funkciji artikuliranja takvih pretpostavki i prikupljanjem podataka za njihovu provjeru.

Proces formuliranja očekivanja koja proizlaze iz zamišljene interpretacije testa može se olakšati razmatranjem alternativnih hipoteza koje pobijaju prvotna tumačenja. Korisno je također uzeti u obzir poglede različitih zainteresiranih strana, postojeća iskustva sa sličnim testovima i kontekstima kao i očekivane posljedice predložene upotrebe testa. Uvjerljive alternativne pretpostavke često se mogu izvesti razmatranjem pitanja mjeri li test manje ili više od pretpostavljenog konstrukta. Takvi problemi mogu upućivati na *podzastupljenost konstrukta* i *konstruktno irelevantne izvore varijance*.

Podzastupljenost konstrukta tiče se stupnja u kojem testom nisu obuhvaćeni njegovi važni aspekti. To ima za posljedicu sužavanje značenja testovnih rezultata jer testom nisu adekvatno obuhvaćeni potrebni sadržaji, aktivirani odgovarajući

psihički procesi ili zahvaćeni neki za konstrukt važni načini odgovaranja. Uzmimo za primjer test čitanja namijenjen mjerenju sposobnosti dječjeg čitanja i tumačenja priča s razumijevanjem. Određeni test može biti slaba reprezentacija ovog konstrukta zbog toga što ne sadrži dovoljno raznolike odlomke ili ako su u njemu izostavljeni uobičajeni tekstovi. Drugi primjer je test anksioznosti kojim su obuhvaćene samo fiziološke, ali ne i emocionalne, kognitivne ili situacijske komponente.

Konstruktno irelevantna varijanca posljedica je stupnja u kojem su testovni rezultati pod utjecajem procesa koji nisu dio mjenog konstrukta. Testovni rezultati mogu sustavno biti pod stanovitim utjecajem čimbenika koji nisu dio konstrukta. U slučaju testa čitanja, konstruktno irelevantne komponente mogu proizlaziti iz emocionalnih reakcija na testovni sadržaj, iskustva u čitanju odlomaka u testu ili vještina pisanja potrebnih za sastavljanje odgovora. Ovisno o definiciji konstrukta, irelevantne sastavnice mogu biti poznavanje rječnika kao i brzina čitanja. U testu anksioznosti izvorom konstruktno irelevantne varijance može se smatrati i izbjegavanje davanja simptomatski indikativnih odgovora.

Gotovo svi testovi izostavljaju neke elemente za koje neki potencijalni korisnici drže da ih treba mjeriti i uključuju druge koje ovi smatraju neprikladnim. Validacija uključuje pažljivo razmatranje svih mogućih poremetnji u značenju testovnih rezultata koje mogu biti posljedica neadekvatne reprezentacije konstrukta, ali i aspekata mjerenja kao što su oblik testa, uvjeti primjene ili jezična razina, koji mogu bitno ograničiti ili opravdati njihovu interpretaciju. To znači da proces validacije može dovesti do izmjena testa, njegova konceptualnog okvira ili i jednog i drugog. Dakako, izmijenjeni test može zahtijevati i novu validaciju.

Kada su pretpostavke koje opravdavaju planirane interpretacije formulirane, validacija se u svrhu njihove evaluacije nastavlja prikupljanjem empirijskih podataka, proučavanjem relevantne literature i/ili provođenjem logičke analize. Empirijski podaci mogu biti lokalnog karaktera (dobiveni u kontekstima u kojima će test biti korišten), a mogu potjecati i iz sličnih primjena u drugim okruženjima. Korištenje postojećih podataka za slične testove i okruženja može unaprijediti kvalitetu argumentacije valjanosti, naročito kada su neposredni validacijski podaci ograničeni.

Budući da argumentacija valjanosti najčešće ovisi o većem broju pretpostavki, snažna podrška jednoj ni na koji način ne otklanja potrebu dokazivanja drugih. Na primjer, visoka povezanost prediktora i kriterija za neki posao nije dovoljna za opravdanje upotrebe testa u selekciji bez pažljive provjere značenja i adekvatnosti kriterijske mjere. Odluke o tome koji oblik podataka može najbolje opravdati zamišljenu interpretaciju i upotrebu testa trebaju biti vođene stručnim mišljenjem. Kao i u svim drugim znanstvenim postupcima, primarna je kvaliteta podataka. Nekoliko čvrstih dokaza određene pretpostavke bolji su od brojnih podataka upitne kakvoće.

Validacija je zajednička odgovornost autora i korisnika testa. Autor je odgovoran za pribavljanje relevantnih podataka i obrazloženje predložene upotrebe testa. Korisnik testa ima konačnu odgovornost za procjenu upotrebljivosti testa u okolnostima u kojima ga treba primijeniti. Kada se upotreba testa razlikuje od one koju je opravdao njegov autor, korisnik snosi posebnu odgovornost za validaciju. Posebni standardi primjenjuju se na validacijski proces u kojem postoji podjela odgovornosti. Treba imati na umu da važne doprinose u validacijskom procesu pružaju i drugi istraživači, izvješćujući o nalazima koji su povezani sa značenjem testovnih rezultata.

Izvori validacijskih podataka

U sljedećim odjeljcima ukratko su opisani različiti izvori podataka koji se mogu iskoristiti u evaluaciji predloženih tumačenja testovnih rezultata za određene svrhe. Takvi izvori podataka mogu osvijetliti različite aspekte valjanosti, ali oni ne znače i posebne tipove valjanosti. Valjanost je jedinstven koncept. Ona je stupanj u kojem svi akumulirani podaci podržavaju iskazanu interpretaciju testovnih rezultata za određenu svrhu. Slično *Standardima* iz 1985., ovo izdanje govori o vrstama validacijskih podataka, a ne o posebnim tipovima valjanosti. Da bi se naglasilo ovo razlikovanje, tekst koji slijedi ne koristi tradicionalno nazivlje (tj. upotrebu termina kao što su sadržajna ili prediktivna valjanost). Rječnik na kraju ove knjige sadrži i definicije tradicionalnih naziva kojima se objašnjava razlika između njihova prijašnjeg (tradicionalnog) i sadašnjeg značenja.

Podaci utemeljeni na sadržaju testa

Važni validacijski podaci mogu se dobiti analizom odnosa između testovnog sadržaja i konstrukta koji se želi mjeriti. Testovni sadržaj tiče se obuhvaćenih tema, formulacija i oblika zadataka, zadaća ispitanika ili pitanja u testu kao i uputa za primjenu i ocjenjivanje. Autori testova često rade prema specifikaciji područja sadržaja. Specifikacija sadržaja je detaljni opis, koji često uključuje razdiobu područja i vrsta čestica. Podaci utemeljeni na testovnom sadržaju mogu uključivati logičke ili empirijske analize adekvatnosti reprezentacije područja sadržaja testa i relevantnosti područja za deklarirane interpretacije testovnih rezultata. Ovi podaci također mogu biti rezultat prosudbi stručnjaka o odnosu između dijelova testa i konstrukta. Na primjer, pri izradi testa namijenjenog licenciranju za određenu struku mogu se specificirati njezini glavni aspekti te od stručnjaka u tom zanimanju tražiti da pridruže testovne zadatke odgovarajućim kategorijama. Oni, ili drugi kvalificirani

stručnjaci, tada mogu prosuđivati i reprezentativnost skupa odabranih zadataka. Katkada se konstruiraju i pravila ili algoritmi za biranje ili generiranje zadataka koji se sustavno razlikuju po sadržajnim svojstvima, utvrđenim specifikacijama.

Neki testovi temelje se na sustavnim opažanjima ponašanja. Na primjer, popis zadaća obuhvaćenih nekim poslom može se dobiti opažanjem ponašanja na poslu, zajedno s prosudbama kvalificiranih stručnjaka. Mišljenja stručnjaka mogu se upotrijebiti za procjenu relativne važnosti, "kritičnosti" i/ili frekvencije različitih zadaća. Test uzorka posla tada se može izraditi na temelju slučajnog ili stratificiranog uzorka zadaća visoko procijenjenih po spomenutim svojstvima i primijeniti u standardiziranim uvjetima izvan samog posla.

Prikladnost nekog područja sadržaja povezana je sa specifičnim zaključcima koji se trebaju izvesti iz testovnih rezultata. Stoga, ako se razmišlja o upotrebi testa za svrhu za koju on prvotno nije bio namijenjen, posebno je važno ocijeniti primjerenost izvornog područja sadržaja novoj upotrebi. Primjerice, pri evaluaciji obrazovnog programa testovi trebaju adekvatno pokrivati sadržaje kojima je u nastavnom planu poklonjena slaba ili nikakva pažnja, kao i one koji su bili više obuhvaćeni nastavom. U takvim prilikama osobe zadužene za donošenje odluka mogu evaluirati postignuće s obzirom na zanemarena kao i na obrađena područja. S druge strane, kada se testira stupanj u kojem je student ovladao programom koji je prošao, pri čemu se želi doći do informacija potrebnih za odlučivanje o njegovu napredovanju ili diplomiranju, okvir elaboracije područja sadržaja adekvatno je ograničiti na ono što je on imao prilike naučiti u završenom programu.

Podaci o sadržaju mogu se dijelom upotrijebiti i za pojašnjavanje pitanja o razlikama u značenju ili tumačenju testovnih rezultata u relevantnim podskupovima ispitanika. Od posebne je važnosti mogućnost da slaba reprezentacija konstrukta ili konstruktno irelevantni činitelji favoriziraju ili hendikepiraju jednu ili više skupina ispitanika. Temeljito proučavanje konstrukta i područja sadržaja od strane različitih stručnjaka može uputiti na moguće irelevantne izvore težine (ili lakoće) koji zahtijevaju dodatno istraživanje.

Podaci utemeljeni na procesu odgovaranja

Teorijske i empirijske analize procesa odgovaranja mogu pružiti podatke o slaganju konstrukta s prirodom učinka ili odgovora koje su ispitanici doista koristili. Na primjer, ako je test namijenjen procjeni matematičkog rezoniranja, važno je provjeriti da li ispitanici doista rezoniraju na datom materijalu ili možda samo slijede uobičajene algoritme. Slično, skala namijenjena procjeni stupnja ekstraverzije ili introverzije ne bi trebala biti pod većim utjecajem društvene poželjnosti.

Podaci utemeljeni na procesu odgovaranja općenito dolaze iz individualnih odgovora. Odgovori ispitanika na pitanja o njihovim strategijama rada ili odgovaranja na pojedine zadatke mogu pružiti podatke koji obogaćuju definiciju konstrukta. Na primjer, praćenje nastanka odgovora u zadacima pisanja kroz sukcesivne etape ili elektronički zabilježene izmjene također daju podatke o aktiviranim procesima. Dokumentiranje drugih aspekata učinka kao što su pokreti očiju ili vrijeme reakcije također može biti relevantno za neke konstrukte. Zaključci o procesima na kojima se temelji učinak u testu mogu se zasnivati i na analizi odnosa među dijelovima testa te odnosa između testa i drugih varijabli. Ustanove li se velike individualne razlike u procesima, korisno je ponovo razmotriti neke aspekte testa.

Podaci o procesima aktiviranim testom mogu pripomoći razumijevanju razlika u značenju ili tumačenju testovnih rezultata kod relevantnih podskupina ispitanika. Istraživanja procesa koja uključuju ispitanike iz različitih skupina može omogućiti stjecanje uvida u stupanj u kojem sposobnosti ili pomoćne vještine, irelevantne za konstrukt, diferencijalno djeluju na njihov učinak u testu.

Proučavanja procesa odgovaranja nisu ograničena samo na ispitanike. U njima često sudjeluju opažači ili procjenjivači koji bilježe i/ili ocjenjuju učinak ispitanika. U takvim slučajevima relevantni validacijski podaci uključuju stupanj u kojem su procesi koje primjenjuju opažači ili procjenjivači sukladni s predviđenom interpretacijom rezultata. Na primjer, ako se očekuje da suci pri ocjenjivanju učinka ispitanika primjenjuju određene kriterije, važno je ustanoviti jesu li oni to doista činili i u kojem stupnju na njihove procjene djeluju činitelji koji su irelevantni za predviđene interpretacije. Stoga validacija može obuhvatiti i empirijska proučavanja procesa koje pri registraciji i evaluaciji ispitanikovih odgovora upotrebljavaju opažači ili procjenjivači, s posebnim naglaskom na primjerenosti tih procesa predviđenoj interpretaciji ili definiciji konstrukta.

Podaci utemeljeni na unutrašnjoj strukturi

Analize unutrašnje strukture testa mogu pokazati koliko odnosi između testovnih zadataka i komponenata testa odgovaraju konstrukt na kojem je utemeljena interpretacija testovnih rezultata. Konceptualni okvir testa može implicirati samo jednu dimenziju ponašanja ili pretpostaviti nekoliko homogenih, ali odijeljenih komponenata. Na primjer, mjera poteškoća na upitniku zdravstvenog stanja može se odnositi i na fizičko i na emocionalno zdravlje. Za valjanost je relevantan stupanj u kojem odnosi među česticama potkrepljuju određenja konceptualnog okvira.

Specifične vrste analiza i njihova interpretacija ovise o tome kako će se test koristiti. Na primjer, ako određena aplikacija pretpostavlja seriju testovnih kom-

ponenata rastuće težine, potrebni su empirijski podaci koji će omogućiti prosudbu ostvarenosti takve pretpostavke. Teorija koja pretpostavlja jednodimenzionalnost tražit će informacije o homogenosti zadataka. U tom slučaju povezanost među zadacima također znači procjenu pouzdanosti rezultata, ali takav indeks je neprikladan za testove kompleksnije unutrašnje strukture.

Neka proučavanja unutrašnje strukture služe ispitivanju eventualno različitog funkcioniranja zadataka u različitim prepoznatljivim skupovima ispitanika. O različitom funkcioniranju zadataka govori se kada različite skupine ispitanika slične opće sposobnosti ili sličnog statusa na nekom prikladnom kriteriju, imaju u prosjeku sustavno različite odgovore na određeni zadatak. Ovo pitanje raspravljeno je u poglavljima 3 i 7. Međutim, različito funkcioniranje zadataka nije uvijek pogreška ili slabost. Podskupovi zadataka koji imaju neko zajedničko svojstvo (npr. specifični sadržaj, način prezentacije) mogu funkcionirati različito u različitim grupama slično ocijenjenih ispitanika. To upućuje na multidimenzionalnost koja može biti neočekivana ili pak predviđena konceptualnim okvirom testa.

Podaci utemeljeni na odnosima s drugim varijablama

Analize povezanosti testovnih rezultata s vanjskim varijablama važan su izvor validacijskih podataka. Vanjske varijable mogu biti mjere nekih kriterija za koje se očekuje da bi ih promatrani test mogao predviđati, drugi testovi za koje se pretpostavlja da mjere iste konstrukte i testovi koji mjere povezane ili sasvim različite konstrukte. Mjere koje su operativno neovisne o testovnim rezultatima, kao što su kriteriji učinka, često se koriste u kontekstu zapošljavanja. Kategorijalne varijable, uključujući varijable grupne pripadnosti, relevantne su kada se na temelju teorijske osnovice testa očekuju ili ne očekuju međugrupne razlike. Na temelju podataka o povezanosti s drugim varijablama propituje se stupanj u kojem su takve povezanosti sukladne s konstruktom koji je u osnovi predloženih interpretacija testa.

Konvergentni i diskriminantni podaci. Povezanosti testovnih rezultata i drugih mjera namijenjenih procjeni sličnih konstrukata pružaju konvergentne podatke, dok povezanosti testovnih rezultata s mjerama različitih konstrukata daju diskriminantne podatke. Na primjer, unutar istog teorijskog okvira, može se očekivati da rezultati u testu razumijevanja pročitano, koji je sastavljen od zadataka tipa višestrukog izbora, visoko koreliraju (konvergentni podaci) s mjerama razumijevanja pročitano izvedenim drugačijim metodama, npr. zadacima esej-tipa; obrnuto se može očekivati (niže korelacije, diskriminantni podaci) za mjere drugih vještina, npr. logičkog rezoniranja. Povezanosti među različitim metodama mje-

renja konstrukta mogu biti od naročite pomoći u “izoštavanju” i osmišljavanju značenja i interpretacije rezultata.

Podaci o odnosima s drugim varijablama mogu potjecati iz eksperimentalnih kao i iz korelacijskih proučavanja. Istraživanja se, primjerice, mogu organizirati da bi se ispitalo mijenjaju li se rezultati u nekoj mjeri anksioznosti u funkciji određenog psihološkog tretmana, ili, razlikuju li rezultati u testu akademskog postignuća podučavane i nepodučavane skupine ispitanika. Ako se poboljšavanje učinka u testu u funkciji kratkotrajnog instruiranja tretira kao ugrožavanje valjanosti, korisno je istražiti razlikuju li se instruirana i neinstruirana skupina po svojim rezultatima.

Povezanosti testa i kriterija. Podaci o odnosu testovnih rezultata i relevantnih kriterija mogu se izraziti na različite načine, ali temeljno pitanje uvijek glasi: koliko precizno testovni rezultati predviđaju kriterijski učinak? Traženi stupanj preciznosti nužno ovisi o svrsi za koju se test koristi.

Kriterijska varijabla je mjera nekog atributa ili ishoda važnog korisniku testa, koji može biti npr. rukovoditelj u školskom sustavu, menadžer u poduzeću ili klijent. Izbor kriterija i mjernog postupka za dobivanje kriterijskih rezultata od izuzetne su važnosti. Vrijednost proučavanja odnosa testa i kriterija ovisi o relevantnosti, pouzdanosti i valjanosti interpretacije utemeljene na kriterijskoj mjeri u datoj testovnoj aplikaciji.

Za evaluaciju test-kriterij povezanosti tradicionalno se razlikuju dva nacrt-a koji se najčešće nazivaju prognostički (prediktivni) i dijagnostički (istodobni). Prognostičko istraživanje pokazuje s kojom se preciznošću na temelju testa mogu predviđati poslije dobiveni kriterijski rezultati. U dijagnostičkim istraživanjima prediktorski i kriterijski podaci dobivaju se u približno isto vrijeme. Kada se predviđanje doista i čini, kao što je to slučaj u obrazovanju, zapošljavanju ili planiranju rehabilitacijskih programa, prognostička istraživanja trebaju uključiti vremensku razliku i druge karakteristike konkretne situacije. Dijagnostički podaci, koji izbjegavaju vremenske promjene, posebno su korisni za psihodijagnostičke testove ili ispitivanje alternativnih mjera nekog konstrukta. Općenito, izbor istraživačkog nacrt-a ovisi o prethodnim podacima koji pokazuju daju li prognostička i dijagnostička istraživanja u promatranom području iste ili različite rezultate.

Testovni rezultati katkad se koriste za pridruživanje ispitanika različitim tretmanima kao što su npr. različiti poslovi unutar ustanove, s ciljem da takva klasifikacija bude korisna i za ustanovu i za pojedinca. U ovom kontekstu potrebni su podaci koji omogućuju prosudbu adekvatnosti upotrebe testa za klasifikacijske svrhe. Klasifikacijske odluke opravdavaju se podacima koji pokazuju da je povezanost testovnih rezultata s kriterijem učinka različita u različitim tretmanima. Moguće je da testovi budu visoko prediktivni za učinak u različitim programima ili poslovima,

ali da nema informacija nužnih za komparativnu prosudbu njihove efikasnosti u klasifikaciji u različite tretmane. Općenito, pravila odlučivanja u selekciji ili klasifikaciji ovise i o broju osoba koje se trebaju prihvatiti ili broju ispitanika koji se mogu smjestiti u alternativne klasifikacijske kategorije.

Podaci o odnosima s drugim varijablama koriste se također za ispitivanje diferencijalne predikcije. Na primjer, nalaz da odnos testovnih rezultata i relevantnog kriterija varira od jedne do druge skupine može indicirati da značenje testovnih rezultata, možda zbog slabe reprezentacije konstrukta ili konstruktno irelevantnih komponenata, nije isto za pripadnike različitih grupa. Međutim, razlika može također upućivati na različito značenje kriterija u različitim grupama. Razlike u povezanostima testa i kriterija mogu biti i posljedica pogreške mjerenja, posebno kada se grupne središnje vrijednosti razlikuju, tako da takve razlike ne upućuju nužno i na razlike u značenju rezultata.

Generalizacija valjanosti. Važno pitanje u obrazovanju i zapošljavanju odnosi se na mogućnost poopćavanja validacijskih podataka, utemeljenih na odnosu test-kriterij, na nove prilike bez dodatnih istraživanja valjanosti u tim novim situacijama. Kada se test koristi za prognozu istog ili sličnog kriterija (npr. učinka u određenom poslu) u različito vrijeme ili na različitim mjestima, tipično se nalazi da test-kriterij korelacije znatno variraju. U prošlosti se iz toga izvodila obveza redovite lokalne validacije. Međutim, novije meta-analitičke studije pokazale su da se dobar dio tog varijabiliteta u nekim područjima može objasniti statističkim artefaktima kao što su fluktuacije u uzorkovanju i varijacije raspona testovnih rezultata te pouzdanosti kriterijskih podataka. Kada se eliminiraju utjecaji ovakvih činitelja, preostali varijabilitet koeficijenata valjanosti može biti relativno malen. Stoga statistički sažeci ranijih validacijskih studija u sličnim situacijama mogu biti vrlo korisni u procjeni test-kriterij povezanosti u novim prilikama. Ovakvi postupci odnose se na proučavanje generalizacije valjanosti.

U nekim okolnostima postoji čvrsta osnovica za korištenje generalizacije valjanosti. To može biti slučaj kada je meta-analitička baza podataka velika, kada meta-analitički podaci adekvatno reprezentiraju situaciju na koju se želi generalizirati i kada korekcija statističkih artefakata proizvodi jasne i dosljedne sklopove validacijskih podataka. U takvim okolnostima informacijska vrijednost lokalne validacijske studije može biti relativno ograničena. U drugim prilikama prepreke za korištenje generaliziranih podataka mogu biti znatno veće. Meta-analitička baza može biti mala, nalazi manje konzistentni, a nova situacija imati svojstva bitno različita od onih sadržanih u bazi podataka. U tom će slučaju situacijski specifični podaci biti od relativno većeg značenja. Iako istraživanja generalizacije valjanosti pokazuju da rezultati jednog lokalnog validacijskog postupka mogu biti sasvim neprecizni, po-

stoje situacije u kojima i jedno, pažljivo provedeno istraživanje, s adekvatno velikim uzorkom, daje dovoljno podataka za opravdavanje upotrebe testa u novim okolnostima. Ovo upućuje na važnost pažljivog ispitivanja komparativne informacijske vrijednosti lokalnih u odnosu na meta-analitičke studije.

U provedbi ispitivanja mogućnosti generalizacije validacijskih podataka, uključena istraživanja mogu varirati prema nekoliko situacijskih obilježja. Neka od glavnih su (a) razlike u načinu mjerenja prediktorskog konstrukta, (b) vrsta posla ili programa, (c) vrsta korištene kriterijske mjere, (d) vrsta ispitanika i (e) vrijeme u kojem je istraživanje provedeno. U svakom istraživanju generalizacije mogu postojati varijacije u bilo kojem broju navedenih obilježja, a njihov glavni cilj je empirijski odrediti veličinu utjecaja takvih obilježja na vrijednost dobivenih test-kriterij korelacija.

Stupanj u kojem se prognostički ili dijagnostički podaci o generalizaciji valjanosti mogu upotrijebiti u novim situacijama velikim dijelom je funkcija prikupljenih istraživanja. Iako generalizirane informacije često mogu potkrijepiti procjenu valjanosti u novoj situaciji, njezina prihvatljivost ograničena je količinom dostupnih podataka.

Prethodna rasprava usmjerena je na upotrebu kumulativnih baza podataka za procjenu povezanosti prediktora i kriterija. Meta-analitičke tehnike mogu se također koristiti za sažimanje drugih oblika podataka relevantnih za druge vrste zaključaka koje netko želi izvesti iz testovnih rezultata, kao što su efekti instruiranja i efekti nekih izmjena u uvjetima testiranja uvedenih radi prilagodbe ispitanicima s invaliditetom.

Podaci utemeljeni na posljedicama testiranja

Pitanje uključivanja planiranih i neplaniranih posljedica testiranja u poimanje valjanosti posljednjih je godina privuklo određenu pažnju. Podaci o nekim posljedicama mogu utjecati na zaključke o valjanosti. Ovdje, međutim, treba razlikovati podatke koji su neposredno relevantni za valjanost od onih koji mogu djelovati na odluke u socijalnoj politici, ali se ipak nalaze izvan okvira valjanosti.

Razlikovanje pitanja valjanosti i socijalne politike postaje posebno važno kada se uoče različite posljedice upotrebe testa za različite društvene skupine. Na primjer, znatnu pažnju privukle su posljedice grupnih razlika u testovnim rezultatima na odabir i napredovanje namještenika, smještaj djece u specijalne obrazovne programe i sužavanje školskih programa da bi se isključili sadržaji koji nisu procjenjivani. Iako informacije o posljedicama testiranja mogu utjecati na odluke o upotrebi testa, takve posljedice same po sebi ne negiraju valjanost planiranih interpretacija.

Zaključci o valjanosti ili ne-valjanosti u svjetlu posljedica testiranja ovise o detaljnijem uvidu u izvore tih posljedica.

Razmotrimo kao primjer nalaz o različitim proporcijama izbora članova različitih skupina na temelju rezultata u selekcijskom testu. Ako su te razlike nastale jedino zbog nejednakih distribucija vještina koje se mjere testom, te ako su te vještine važne za učinak na poslu, tada nalaz o grupnim razlikama per se ne upućuje na poremetnju valjanosti testom predviđenih zaključaka. Ako se, međutim, testom mjere razlike u vještinama koje nisu povezane s učinkom na poslu (npr. složeni test čitanja za posao koji zahtijeva samo minimalnu funkcionalnu pismenost), ili ako su razlike nastale zbog osjetljivosti testa na neka svojstva ispitanika koja nisu dio testovnog konstrukta, tada valjanost testa može biti upitna, čak i kada njegovi rezultati pozitivno koreliraju s nekom mjerom uspjeha na poslu. Stoga podaci o posljedicama mogu biti neposredno relevantni za valjanost ako se one mogu pripisati izvorima poremetnje kao što su slaba reprezentacija konstrukta ili prisutnost za njega irelevantnih komponenata. Podaci o posljedicama koji se ne mogu tumačiti na ovaj način, koji u biti znače valjane razlike u učinku, ključni su za donošenje društvenih odluka, ali se nalaze izvan okvira valjanosti testa.

Testovi se obično primjenjuju uz očekivanje da će se određena korist postići upravo planiranom upotrebom testovnih rezultata. Neki od mnogih mogućih dobitaka su odabir učinkovitih terapijskih tretmana, raspodjela zaposlenika na odgovarajuće poslove, sprečavanje zapošljavanja nekvalificiranih pojedinaca ili poboljšavanje metoda podučavanja. Temeljna svrha validacije je provjera mogućnosti ostvarivanja takvih specifičnih dobitaka. Stoga, u slučaju korištenja testa za klasifikacijske svrhe, validacija treba pokazati da alternativne klasifikacije imaju i različite posljedice za pojedince i organizacije. Kada je riječ o testiranju pri zapošljavanju, ako izdavač testa tvrdi da će njegova primjena dovesti do smanjenja troškova obuke zaposlenika, poboljšanja radne učinkovitosti ili postizanja neke druge dobiti, tada validacija treba biti utemeljena na podacima koji potkrepljuju tu tvrdnju.

Katkada se tvrdnje o koristi testiranja odnose na dobitke koji nisu neposredna posljedica upotrebe testovnih rezultata. Na primjer, pedagoški testovi mogu se zagovarati uvjerenjem da će njihova upotreba povećati motivaciju učenika ili studenata ili potaknuti promjene nastavnih metoda naglašavanjem odgovornosti nastavnika za postizanje ciljeva podučavanja. Ako su takve tvrdnje od ključne važnosti za obrazloženje testiranja, neposredno ispitivanje posljedica testiranja ima čak i veću važnost. U takvim slučajevima validacijski proces treba se utemeljiti na podacima koji pokazuju da se predviđene koristi od testiranja mogu i ostvariti.

Integracija validacijskih podataka

Cjelovito obrazloženje valjanosti integrira različite skupove podataka u koherentno objašnjenje stupnja u kojem postojeći podaci i teorija podržavaju predloženo korištenje testovnih rezultata za određene svrhe. Ono obuhvaća podatke prikupljene novim istraživanjima kao i one koji su prije objavljeni. Validacijski podaci mogu pokazati potrebu dotjerivanja definicije konstrukta, sugerirati promjenu sadržaja testa ili nekog drugog aspekta procesa testiranja kao i određena dodatna istraživanja.

Konačno, valjanost predloženog tumačenja testovnih rezultata počiva na svim dostupnim podacima relevantnim za tehničku kvalitetu sustava testiranja. Ona uključuje podatke prikupljene u izradi testa, odgovarajuću pouzdanost, upute za primjenu i ocjenjivanje, preciznu osnovicu za ocjenu, izjednačavanje i normiranje testovnih rezultata i odgovarajuću pažnju posvećenu nepristranosti za sve ispitanike prema načelima opisanim u sljedećim poglavljima *Standarda*.

Standard 1.1

Za svako preporučeno tumačenje i upotrebu testovnih rezultata potrebno je navesti obrazloženje koje uključuje sažetak empirijskih i teorijskih argumenata na kojima počiva predložena upotreba ili interpretacija.

Komentar: U obrazloženju upotrebe ili interpretacije trebaju biti navedene sve okolnosti koje je nužno istražiti da bi se opravdalo predloženo korištenje testa. Sažetak bi trebao biti kombinacija logičke analize i empirijskih podataka koji podržavaju racionalu testa. Podaci mogu dolaziti iz lokalnih proučavanja, provedenih u okolnostima u kojima će test biti korišten, iz prijašnjih specifičnih studija ili opsežnih statističkih sinteza dostupnih podataka koji zadovoljavaju jasno postavljene kriterije. Nijedna od ovih vrsta podataka nije načelno bolja od drugih; kvaliteta i relevantnost podataka za predloženu upotrebu testa svojstva su koja određuju njihovu vrijednost. Navođenje bilo kakvih podataka treba dati odgovarajuću težinu svim relevantnim nalazima, uključujući i one koji nisu sukladni s predloženom upotrebom testa. Autori testa odgovorni su za prikupljanje podataka koji opravdavaju njihove preporuke, ali je prosudba kvalitete validacijskih podataka i njihove relevantnosti u lokalnim prilikama odgovornost koja pripada korisniku testa.

Standard 1.2

Autor testa dužan je jasno pokazati kako se testovni rezultati trebaju interpretirati i upotrebljavati. Populacija (populacije) kojoj je test namijenjen treba se precizno odrediti, a konstrukt čijoj je procjeni test namijenjen jasno opisati.

Komentar: Odredbe o valjanosti trebaju se odnositi na pojedine interpretacije i upotrebe testa. Neispravno je upotrebljavati opću frazu “valjanost testa”. Nije-dan test nije valjan za sve svrhe ili u svim situacijama. Svaka preporučena svrha ili interpretacija zahtijeva validaciju i treba imati jasnu specifikaciju populacije kojoj je test namijenjen, konstrukta koji se njime mjeri te načina i okolnosti primjene njegovih rezultata.

Standard 1.3

Ako valjanost neke uobičajene ili vjerojatne interpretacije nije istražena, ili je mogućnost nekog tumačenja u neskladu s dostupnim podacima, ta činjenica mora biti jasno istaknuta, a potencijalni korisnici upozoreni na njegovu neutemeljenost.

Komentar: Ako prethodna iskustva pokazuju da je test neprikladan za donošenje nekih vrsta odluka, za takve situacije potrebna su specifična upozorenja. S druge strane, budući da ne postoje dvije iste situacije, neke generalizacije su uvijek nužne. Stoga se procjena stupnja u kojem je upotreba testa opravdana validacijskim podacima treba izvesti na temelju stručnog mišljenja.

Standard 1.4

Ako se test koristi na način koji nije validiran, dužnost je korisnika da to opravda, ako je to nužno i prikupljanjem novih podataka.

Komentar: Stupanj u kojem su postojeći validacijski podaci primjenjivi na novu situaciju te koji su eventualno novi podaci potrebni, prosuđuje se na temelju stručnog mišljenja. Količina i vrsta novih podataka mogu ovisiti o iskustvu sa sličnim prijašnjim upotrebama ili interpretacijama testa te o količini, kvaliteti i relevantnosti postojećih informacija.

Standard 1.5

Strukturu svakog uzorka ispitanika čiji su rezultati poslužili za dobivanje validacijskih podataka treba opisati onoliko detaljno koliko je to svrsishodno, uključujući i glavne relevantne sociodemografske i razvojne karakteristike.

Komentar: Statistički nalazi mogu ovisiti o čimbenicima povezanim s validacijskim uzorkom. Kada se uzorkom namjerava predstaviti neka populacija, ona se treba opisati uz odgovarajući osvrt na sve sistematske faktore koji mogu ograničiti reprezentativnost uzorka. Faktori za koje je opravdano pretpostaviti da mogu djelovati na rezultate jesu spontana selekcija, osipanje uzorka, poznavanje jezika, stupanj invalidnosti, kriteriji isključivanja i drugi. Ako su npr. ispitanici u validacijskoj studiji pacijenti, tada su važne njihove dijagnoze, kao i druge karakteristike, npr.

ozbiljnost poremećaja. Za testove koji se koriste u industriji relevantne informacije mogu biti radni status (npr. kandidati za posao prema zaposlenima), opća razina iskustva, naobrazba te spolna i etnička struktura. Za testove koji se primjenjuju u obrazovnom okruženju važne informacije mogu biti prethodna naobrazba, razvojna razina, karakteristike lokalne zajednice, upisna politika obrazovne ustanove, a također i spolna i etnička struktura uzorka. U nekim slučajevima ograničenja koja proizlaze iz zaštite privatnosti mogu biti prepreka za dobivanje takvih informacija.

Standard 1.6

Kada se validacija testa dijelom oslanja na prikladnost testovnog sadržaja, potrebno je opisati i obrazložiti postupke njegove specifikacije i generiranja u odnosu na konstrukt ili područje koje se želi mjeriti ili obuhvatiti. Ako definicija odabranog sadržaja uključuje kriterije kao što su važnost, čestina ili nužnost, takvi se kriteriji također trebaju detaljno objasniti i obrazložiti.

Komentar: Na primjer, autor testa može opisati logičku strukturu koja povezuje čestice testa s područjem njegova sadržaja, ilustrirajući relevantnost svakog zadatka i reprezentativnost cijelog skupa čestica. Također, mogu se istaknuti i oni dijelovi područja sadržaja testa koji nisu obuhvaćeni testovnim zadacima.

Standard 1.7

Kada se validacija testa dijelom oslanja na mišljenja ili odluke stručnjaka, opažaca ili procjenjivača, potrebno je u cijelosti opisati postupke njihova odabira i dobivanja njihovih mišljenja ili procjena. Potrebno je navesti i njihove kvalifikacije i iskustvo. Opis postupaka treba sadržavati informacije o svim uputama i uvježbavanju procjenjivača, jesu li svoje odluke donosili nezavisno te podatke o stupnju njihova međusobnog slaganja. Ako ovi sudionici surađuju ili razmjenjuju informacije, potrebno je objasniti postupke kojima oni mogu utjecati jedni na druge.

Komentar: Sustavno prikupljanje sudova ili mišljenja može se odvijati u više etapa izrade testa (npr. u pribavljanju sudova stručnjaka o prikladnosti sadržaja ili njegovoj reprezentativnosti), u oblikovanju pravila ili standarda za tumačenje rezultata (npr. u postavljanju graničnih rezultata) ili u ocjenjivanju odgovora (npr. procjena odgovora na otvorena pitanja). Kada se takvi postupci primjenjuju, kvaliteta dobivenih sudova važna je za validaciju testa. Može biti sasvim prikladno imati stručnjake koji rade zajedno i odluke donose konsenzusom, ali se u takvim slučajevima njihovi sudovi ne mogu tretirati kao statistički nezavisne jedinice.

Standard 1.8

Ako logička osnovica neke primjene testa ili tumačenja njegovih rezultata ovisi o pretpostavkama o psihičkim procesima ili kognitivnim operacijama ispitanika, potrebno je pružiti teorijske ili empirijske dokaze koji podržavaju takve pretpostavke. Kada su tvrdnje o procesima koje upotrebljavaju opažači ili ocjenjivači dio argumentacije valjanosti testa, njih je također potrebno obrazložiti na sličan način.

Komentar: Ako se specifikacijom testa navode procesi koji se žele procjenjivati, potrebno je dokazati da njegovi zadaci doista obuhvaćaju upravo te procese.

Standard 1.9

Ako se za test tvrdi da je suštinski neovisan o uvježbavanju ili podučavanju, potrebno je dokumentirati neosjetljivost učinka na promjenu takvih oblika uputa.

Komentar: Materijali koji služe za tumačenje testovnih rezultata trebaju sadržavati podatke o stupnju u kojem se može očekivati poboljšanje rezultata u funkciji uvježbavanja ili podučavanja. Također, materijali koji su pisani za ispitanike trebaju ponuditi informacije o vrijednosti aktivnosti pripremanja za testiranje, uključujući i podučavanje.

Standard 1.10

Kada se sugerira interpretacija učinka u pojedinim zadacima ili manjim podskupovima zadataka, potrebno je navesti podatke i logičko obrazloženje koji opravdavaju takav postupak. Kada je interpretacija pojedinih odgovora izgledna, ali ne i preporučena od strane autora testa, korisnika treba upozoriti na njezinu neutemeljenost.

Komentar: Upute koje se daju korisnicima testa moraju biti dostatne da bi mogli prosuditi stupanj povjerenja zajamčen za bilo koju upotrebu ili interpretaciju koju preporučuje autor testa. Testovni priručnici i izvješća o rezultatima trebaju obeshrabrivati interpretacije informacija koje mogu biti opterećene velikom pogreškom. Ovo je posebno važno kada se sugerira interpretacija učinka u pojedinim zadacima, manjim podskupovima zadataka ili subtestovima.

Standard 1.11

Ako racionala upotrebe ili interpretacije ovisi o pretpostavkama o odnosima dijelova testa, potrebno je navesti podatke koji se odnose na njegovu unutrašnju strukturu.

Komentar: Na primjer, za test se može zahtijevati da bude u osnovi jednodimenzionalan. Takav zahtjev treba podržati multivarijantnim statističkim analizama (kao

što je faktorska analiza) koje pokazuju da je varijabilitet rezultata objašnjiv jednom glavnom dimenzijom mnogo veći od varijabiliteta pripisivog bilo kojoj drugoj prepoznatljivoj dimenziji. Kada test nudi više od jednog rezultata, na temelju odnosa među tim rezultatima treba pokazati sukladnost s konstruktom ili konstruktima koji se žele procjenjivati.

Standard 1.12

Kada se za test predlaže interpretacija rezultata u njegovim dijelovima, razlika rezultata ili profila, potrebno je navesti logičku osnovicu i podatke koji opravdavaju takvu interpretaciju. Tamo gdje se predlaže korištenje kompozitnih rezultata, potrebno je navesti temelj i racionalu formiranja kompozita.

Komentar: Kada se testom dobiva više od jednog rezultata, potrebno je pokazati njihovu distinktivnost, a na temelju odnosa tih rezultata sukladnost s konstruktima koji se žele procjenjivati. Nadalje, podaci o valjanosti interpretacije dvaju odvojenih rezultata ne opravdavaju nužno i interpretaciju razlika među njima. Racionala i podaci moraju se odnositi neposredno na specifični rezultat ili kombinaciju rezultata koja se želi tumačiti ili koristiti.

Standard 1.13

Kada podaci o valjanosti uključuju statističke analize testovnih rezultata, samih ili zajedno s drugim varijablama, uvjete u kojima su podaci prikupljeni treba opisati dovoljno detaljno da bi korisnici mogli prosuditi relevantnost statističkih nalaza za svoje lokalne uvjete. Potrebno je posvetiti pažnju svim aspektima prikupljanja validacijskih podataka za koje se razložno drži da se razlikuju od uvjeta tipičnog operativnog testiranja, za koje je opravdano pretpostaviti da mogu utjecati na učinak u testu.

Komentar: Takvi uvjeti mogu uključivati (ali se ne ograničavaju samo na njih) sljedeće: motivacija ispitanika ili prethodna priprema, obavješćivanje ispitanika o njihovim rezultatima, vrijeme za odgovaranje i drugi uvjeti primjene testa, uvježbavanje ispitivača i njihove druge karakteristike, vrijeme između prikupljanja podataka na različitim mjerama ili uvjeti koji se mogu promijeniti nakon što su dobiveni validacijski podaci.

Standard 1.14

Kada validacijski podaci uključuju empirijske analize testovnih odgovora zajedno s podacima izraženim drugim varijablama, potrebno je obrazložiti izbor tih dodatnih varijabli. Tamo gdje je to prikladno i moguće potrebno je prikazati ili

citirati podatke o konstruktima koje one predstavljaju kao i one koji se odnose na njihova tehnička svojstva. Potrebno je obratiti pažnju na svaki mogući izvor ovisnosti (ili odsutnosti neovisnosti) među varijablama osim onih koji proizlaze iz međuovisnosti konstrukata koje predstavljaju.

Komentar: Sklopovi povezanosti između i unutar rezultata na proučavanom instrumentu i drugim varijablama trebaju biti sukladni teorijskim očekivanjima. Dodatne varijable mogu biti demografske karakteristike, pokazatelji tretmanskih uvjeta ili rezultati na drugim mjerama. Oni mogu uključivati odabrane mjere istog konstrukta, ili se mogu odnositi na druge oline. Pouzdanost rezultata na takvim drugim mjerama i valjanost njihovih interpretacija važan su dio validacijskih podataka proučavanog testa. Ako takve varijable uključuju kompozitne rezultate, njihova konstrukcija treba biti objašnjena. Osim razmatranja svojstava svake pojedinačne varijable, važno je spriječiti pogrešne interpretacije koje proizlaze iz spurioznih izvora ovisnosti među mjerama, uključujući korelirane pogreške ili zajedničku varijancu proizvedenu identičnom mjernom metodom ili zajedničkim elementima.

Standard 1.15

Kada se tvrdi da neka razina uratka u testu predviđa adekvatan ili neadekvatan kriterijski učinak, potrebno je navesti podatke o razini učinka u kriteriju za date razine testovnih rezultata.

Komentar: Regresijske jednadžbe korisnije su od koeficijenata korelacije koji su općenito nedostatni za potpuno opisivanje veza između testova i drugih varijabli. Potrebno je navesti aritmetičke sredine, standardne devijacije i druge sumarne statističke pokazatelje, kao i podatke o distribucijama kriterijskog učinka za fiksne testovne rezultate. Pokazatelji opće povezanosti među varijablama trebaju biti dopunjeni informacijama o obliku povezanosti i njezinu varijabilitetu u različitim rasponima testovnih rezultata. Treba imati na umu da podaci prikupljeni na ispitanicima koji na jednoj ili više mjera pripadaju ekstremnim grupama obično ne mogu pružiti adekvatne informacije o stupnju povezanosti.

Standard 1.16

Kada se validacija temelji na podacima o povezanosti testovnih rezultata s jednom ili više kriterijskih varijabli, potrebno je navesti informacije o njihovoj prikladnosti i tehničkoj kvaliteti.

Komentar: Opis svake kriterijske varijable treba sadržavati podatke o njezinoj pouzdanosti, stupnju u kojem predstavlja ciljani konstrukt (npr. učinak na poslu) i stupnju u kojem je određena drugim izvorima varijabiliteta. Posebnu pažnju tre-

ba posvetiti čimbenicima za koje prethodna istraživanja pokazuju da bi mogli biti izvor pozitivne ili negativne pristranosti u procjeni kriterija kod nekih prepoznatljivih grupa.

Standard 1.17

Ako se testovni rezultati koriste za predviđanje nekog ishoda ili kriterija zajedno s drugim varijablama, regresijske (ili ekvivalentne) analize trebaju osim testovnih rezultata uključiti i te dodatne relevantne varijable.

Komentar: Općenito, ako je dostupno nekoliko prediktora neke kriterijske varijable, njihova optimalna kombinacija ne može se odrediti samo na temelju povezanosti svakog pojedinačnog prediktora i kriterija. Često je informativno procijeniti porast u točnosti predviđanja koji se može očekivati kada se svaka varijabla, uključujući i testovni rezultat, pridruži svim ostalim prediktorima. Analize koje uključuju veći broj prediktora treba, kada je to moguće, provjeriti križnom validacijom ili nekim ekvivalentnim postupkom, a u opisivanju njihovih rezultata potrebno je navesti preciznost procjena regresijskih koeficijenata.

Standard 1.18

Kada su u statističkim analizama korištene korekcije kao što su one koje se odnose na sužavanje raspona rezultata ili nepouzdanost, potrebno je navesti i korigirane i nekorigirane koeficijente, upotrijebljeni postupak i sve statističke vrijednosti temeljem kojih su korekcije izvršene.

Komentar: Korelacija između dviju varijabli (kao što su testovni i kriterijski rezultati) ovisi o rasponu vrijednosti u svakoj od njih. Na primjer, testovni rezultati i kriterijske vrijednosti selekcioniranih kandidata u pravilu će imati manji raspon nego što je to slučaj s cijelom skupinom kandidata. Postoje statističke metode koje omogućavaju "popravku" procjene korelacije u smislu da ona bolje odražava populacijsku vrijednost nego što to čini procjena dobivena na selekcioniranom uzorku. Ovakve korekcije često su korisne, kao npr. kod usporedbe rezultata dobivenih u različitim situacijama. Navođenje korigirane korelacije treba biti popraćeno opisom metode i statističkih vrijednosti korištenih u njezinoj primjeni.

Standard 1.19

Ako se test preporučuje za pridruživanje osoba alternativnim tretmanima ili je moguće da će biti upotrijebljen za tu svrhu, te ako se ishodi tih tretmana mogu uspoređivati po nekom zajedničkom kriteriju, potrebno je, kad god je to moguće, pružiti podatke za različite ishode.

Komentar: Ako se test koristi za klasifikaciju u alternativne stručne, terapijske ili obrazovne programe, nedovoljno je samo pokazati da on predviđa ishode tretmana. Valjanosti postupka klasifikacije moguće je potkrijepiti potvrdom korisnosti testa u određivanju koje će osobe različito profitirati u jednom ili drugom tretmanu. Katkada će se tretmanske kategorije morati spajati da bi se dobilo dovoljno jedinica za statističke analize. Međutim, treba imati na umu da ovakva istraživanja mogu biti neprovediva zbog etičkih i zakonskih ograničenja koja zabranjuju diferencijalno pridruživanje ispitanika i sprečavaju formiranje kontrolnih skupina.

Standard 1.20

Kada se za dokazivanje veličine povezanosti testa i kriterija koristi meta-analiza, test i kriterijske varijable u lokalnim prilikama trebaju biti usporedivi s onima koji su uključeni u sintetičku studiju. Ako relevantna istraživanja na vjerodostojan način pokazuju da bilo koje drugo svojstvo primjene testa može utjecati na veličinu povezanosti testa i kriterija, potrebno je ispitati i navesti podatke o podudarnosti takvih svojstava u lokalnim prilikama i meta-analitičkoj studiji. Potrebno je izričito upozoriti na sva značajna odstupanja koja mogu ograničiti primjenjivost meta-analitičkih nalaza u lokalnoj situaciji.

Komentar: Meta-analiza treba obuhvatiti sva dostupna istraživanja koja udovoljavaju eksplicitno postavljenim kriterijima uključivanja. Meta-analitički podaci korišteni u validaciji obično se temelje na određenom broju testova koji mjere iste ili slične konstrukte i kriterijskih mjera koje se također odnose na iste ili slične predmete interesa. Ovakve sintetičke studije mogu biti također ograničene na samo jedan test i samo jedan kriterij. Povezanost između testa i kriterija za svako se uključeno istraživanje izražava u nekoj zajedničkoj metrici, često kao standardizirana vrijednost poznata kao *veličina efekta*. Veličina povezanosti može ovisiti o karakteristikama situacije u kojoj su testovne i kriterijske mjere dobivene (npr. vrste poslova, karakteristike ispitanika, vremenski interval između prikupljanja testovnih i kriterijskih podataka, godina ili dekada u kojoj su podaci prikupljeni). Ako povezanost testa i kriterija varira ovisno o takvim moderatorskim varijablama, tada, koliko to dopušta broj uključenih istraživanja, meta-analiza treba odijeljeno navesti procijenjene veličine efekta za relevantne situacijske karakteristike. Na primjer, to se može postići navođenjem odvojenih distribucija podskupina istraživanja ili procjenom utjecaja situacijskih karakteristika na veličinu efekta.

Standard 1.21

Svaki meta-analitički podatak koji se koristi kao podrška predloženoj upotrebi testa treba biti jasno opisan, uključujući metodološke izbore u prepoznavanju i

označavanju istraživanja, korigiranje artefakata i ispitivanje mogućih moderator-skih varijabli. Potrebno je navesti pretpostavke na temelju kojih je izvedeno korigiranje artefakata kao što su nepouzdanost kriterija i suženje raspona rezultata, a također i jasno naznačiti posljedice takvih pretpostavki.

Komentar: Meta-analize neizbježno podrazumijevaju prosudbe koje se odnose na određeni broj metodoloških rješenja. Osnovica takvih prosudbi treba biti obrazložena. U slučaju da rješenja uključuju stanovit stupanj nesigurnosti (kao npr. kod korekcije artefakata na temelju pretpostavljenih vrijednosti), o njoj treba voditi računa te ispitati i izvijestiti o stupnju u kojem zaključci o valjanosti ovise o tim pretpostavkama.

Standard 1.22

Kada je jasno iskazano ili implicirano da će preporučena upotreba testa dovesti do specifičnog ishoda, potrebno je navesti osnovicu očekivanja takvog ishoda zajedno s relevantnim podacima.

Komentar: Ako se, na primjer, tvrdi da će upotreba datog testa u selekciji namještenika dovesti do smanjenja pogrešaka na poslu ili cijene treninga, potrebno je pružiti podatke koji opravdavaju takvu tvrdnju. Tvrdnje o dobitku od upotrebe testa mogu se obrazlagati logičkim ili teorijskim argumentima kao i empirijskim podacima. Odgovarajuća pažnja treba se posvetiti nalazima u znanstvenoj literaturi koji mogu biti nesukladni s deklariranim očekivanjima.

Standard 1.23

Kada se upotreba testa ili interpretacija rezultata preporučuje na temelju mišljenja da će testiranje ili program testiranja sam po sebi, osim korištenja testovnih rezultata, dovesti i do neke indirektno koristi, potrebno je eksplicitno navesti obrazloženje očekivanja takvog posrednog dobitka. Za ovakva očekivanja potrebno je pružiti logičke ili teorijske i empirijske dokaze. Odgovarajuća pažnja treba se posvetiti svim kontradiktornim nalazima u znanstvenoj literaturi, uključujući i one koji osim predviđenih sugeriraju druge važne indirektno posljedice.

Komentar: Na primjer, neki obrazovni programi testiranja zagovaraju se na temelju očekivanja da će imati povoljan utjecaj na didaktičku praksu ili da bi polaznicima mogli pojasniti vrstu ili razinu postignuća koja se od njih očekuje. U mjeri u kojoj su takva očekivanja uključena u opravdavanje programa testiranja, ona postaju dio argumenata valjanosti upotrebe testa te trebaju biti proučena u sklopu njegove validacije. Potrebno je obratiti pažnju i na dokaze koji proturječe takvim predviđanjima, npr., da neke okolnosti obrazovnog testiranja mogu imati negativne efekte na didaktičku praksu.

Standard 1.24

Kada upotreba testa dovodi do neplaniranih posljedica, potrebno je istražiti je li njihov uzrok osjetljivost testa na svojstva čijoj procjeni nije bio namijenjen ili je riječ o njegovoj nemogućnosti da u cijelosti obuhvati ciljani konstrukt.

Komentar: Valjanost tumačenja testovnih rezultata može biti ograničena prisutnošću irelevantnih komponenata ili slabom reprezentacijom konstrukta. Kada se čini da neplanirane posljedice barem dijelom potječu iz upotrebe jednog ili više testova, posebno je važno provjeriti ne potječu li one iz takvih izvora slabe valjanosti. Iako grupne razlike same po sebi ne dovode u pitanje valjanost predložene interpretacije, one mogu dati veću uvjerljivost drugim hipotezama koje treba istražiti u sklopu validacijskih postupaka.