

# LIJEČNIČKI VJESNIK

## GLASILO ZBORA LIJEČNIKA HRVATSKE

THE JOURNAL  
OF THE MEDICAL  
ASSOCIATION  
OF CROATIA

ЖУРНАЛ ОБЩЕСТВА  
ВРАЧЕЙ ХОРВАТИИ

ORGANE OFFICIEL  
DE LA SOCIÉTÉ  
MÉDICALE  
DE CROATIE

FACHZEITSCHRIFT  
DER  
ÄRZTEGESELLSCHAFT  
KROATIENS

ZAGREB

God. 88  
SRPANJ  
1966

### CONTENTS

#### LEADING ARTICLE

*Checking the Knowledge — Examiner or Test? (Summary in French)* Z. Bujas . . . . . 755

#### TRANSACTIONS

*Group Therapy of Patients with Chronic Diseases in General Medicine (Summary in English)* Ž. Kulčar, M. Šešo, V. Majnarić . . . . . 769

*The Role of Vagotomy in the Treatment of Peptic Ulcer (Summary in English)* M. Štulhofer, Gj. Vince, K. Župančić . . . . . 781

#### REVIEW

*Partial Ruptures of Muscles in Young People — Soldiers and Sportsmen (Summary in German)* R. Kandel, P. Keros . . . . . 791

*A Case of Primary Lung Lymphogranulomatosis (Summary in English)* M. Ribić, V. Pešut, R. Hirtzler . . . . . 813

#### REFRESHER COURSE

*Current Use of Digitalization in Practice (Summary in English)* Lj. Barić . . . . . 821

#### FROM OUR MEDICAL PAST

*A Contribution to the History of Education in Forensic Medicine in Zagreb a Hundred Years Ago (Summary in French)* M. D. Grmek . . . . . 831

#### EDITORIALS

*Psychology of Group Work* D. Blažević . . . . . 837  
*The Cardiac Pacemaker with a Biologic (Autogenous) Source of Electric Energy* Z. Ašperger . . . . . 839

WORLD MEDICINE IN THE LAST MONTH . . . . . 848

ABSTRACTS FROM THESES . . . . . 865

PROCEEDINGS AND OTHER ACTIVITIES OF THE SOCIETY . . . . . 877

ABSTRACTS . . . . . 881

BOOK REVIEWS . . . . . 892

OBITUARY . . . . . 894

OUR MEDICAL JOURNALS . . . . . 897

NEWS . . . . . 903

IN OTIO . . . . . 907

## PROVJERAVANJE ZNANJA — ISPITIVAČ ILI TEST?

ZORAN BUJAS

*Iz Filozofskog fakulteta Sveučilišta u Zagrebu*

Provjeravanje znanja po svojoj osnovnoj strukturi odgovara vrlo primitivnom psihološkom eksperimentu. Na sličan način kao što se pri psihološkom pokusu izlaže ispitanika određenim podražajima, koji treba da izazovu stano-vitu vrstu psihičke aktivnosti, koja će se očitovati u reakcijama, tako za vri-jeme ispita, ispitivač svojim pitanjima pobuđuje misaono-mnemičke procese ispitanika, čije sadržaje ispitanik objektivira u svojim odgovorima. Kod psiho-loškog pokusa ispitivač, mjereći objektivne karakteristike podražajne situacije i registrirajući vrstu i veličinu reakcija, indirektno mjeri i veličinu psihičke aktivnosti koja je situacijom izazvana i koja se u ponašanju očitovala. Tako nastoji postupati i nastavnik pri procjenjivanju znanja učenika, kada na osno-vi pitanja koja je postavio i odgovora koje je dobio sankcionira nekom ocje-nom kvalitet i ekstenzitet kandidatova znanja.

Uza spomenute sličnosti postoje i određene razlike. Mjerenje, tj. određi-vanje neke pojave prema aspektu veličine, uključuje pojavu i predmet mje-renja, mjerni instrumenat, tehniku mjerenja i samog mjerioca. Pri određiva-nju znanja na ispitu postoji ta osobitost da je onaj koji mjerenje vrši ujedno i instrument mjerenja. Ličnost nastavnika, kao izvor pitanja i kao mjerni instrument odgovora, posjeduje svoje individualne i promjenljive karakteri-stike koje u značajnoj mjeri utječu na rezultate.

Budući da su prilike, u kojima se vrši procjenjivanje znanja, samo djelo-mično standardizirane, to se ocjenjivanje po svojoj strukturi najviše približuje psihološkom postupku skaliranja. Princip je tog postupka jednostavan. Odre-đena psihička veličina definira se na što je moguće jednoznačniji način, a onda se na kontinuumu koji predstavlja tu veličinu, traži od ocjenjivača da označi mjesto koje prema njegovoj procjeni zauzima ta psihička veličina kod određenog ispitanika. Kao uporišne tačke uzduž kontinuumu služe ili kvalita-tivne kategorije ili brojevi koji se odnose na te kategorije. U prvom slučaju je to kvalitativna intervalna skala, a u drugom kvantitativna intervalna skala.

Ocjena, koja se daje nekom kandidatu na ispitu na fakultetu, ima u prvom redu dijagnostičko značenje, a u nešto manjoj mjeri i prognostičko značenje. Dijagnostička funkcija ocjene je da izrazi stupanj znanja kandidata

u času ispita i to bez obzira kako i u kojim je prilikama kandidat stekao to znanje. Prognostička pak funkcija ocjene sankcionira ili daje mogućnost daljnjeg napredovanja u studiju ili sankcionira osposobljenost kandidata za rad u struci.

Kako je ocjenjivanje na fakultetu u prvom redu pokušaj indirektnog mjerenja ili kategoriziranja znanja, to je opravdano promatrati vrijednost tog postupka pod vidom psihometrijske metrologije. Da bi pod tim vidom neko mjerenje psihičkih veličina moglo uopće biti valjano, ono mora posjedovati u određenom stupnju i neke međusobno povezane metrijske karakteristike. Među tim karakteristikama najvažnije su: objektivnost, pouzdanost, osjetljivost i tačnost.

### OBJEKTIVNOST OCJENA

Za neko mjerenje kažemo da je objektivno ako dobiveni rezultati zavise isključivo od veličine pojave koju mjerimo, a ne zavise od onoga koji mjerenje vrši.

U konkretnom slučaju ocjenjivanja na ispitu, mogli bismo reći da je ono objektivno ukoliko bi ocjena bila isključivo određena odgovorima ispitanika, a potpuno nezavisna od onoga tko ispituje i ocjenjuje.

Objektivnost nekog mjerenja ili procjenjivanja određuje se obično tako da se usporede rezultati do kojih su došli različiti mjerioci (ocjenjivači) mjereći istu veličinu ili niz stalnih veličina. Stupanj objektivnosti nekog mjerenja izražava se koeficijentom korelacije između rezultata do kojih su došli različiti ocjenjivači mjereći iste predmete. Na osnovi tog koeficijenta može se onda utvrditi koliko u dobivenim rezultatima sudjeluju konstantni faktori, koji su tipični za veličinu predmeta mjerenja, a koliko promjenljivi faktori koji ukazuju na veličinu subjektivnih faktora onih, koji su mjerenje izvršili. Konkretno, pri ocjenama koje se daju na tradicionalnim ispitima, kontrola se vrši tako da se uspoređuju ocjene koje su različiti, a formalno podjednako kompetentni ocjenjivači, dali istim pismenim ili usmenim odgovorima kandidata.

Provjeravanja te vrste pokazala su na slabu objektivnost školskih ocjena i to bez obzira radi li se o pismenim ili usmenim odgovorima. Tako npr. pri nekom ispitivanju u našim srednjim školama, dane su iste školske zadaće na nezavisnu ocjenu kompetentnim ocjenjivačima. Za svaki predmet, iz kojeg se pisala zadaća, uzeta su po četiri profesora-ocjenjivača. Usporedba ocjena koje su oni dali, pokazala je da je suglasnost u ocjeni izuzetak, a da ocjene iste zadaće mogu ponekad ići od jednog ekstrema do drugoga, tj. od nedovoljnog do odličnog<sup>1</sup>. Pri nekom drugom ispitivanju, ovaj put objektivnost ocjenjivanja usmenih odgovora od 45 učenika iz fizike, nađena su slična neslaganja<sup>2</sup>.

Dakako ni objektivnost ocjena na fakultetima nije bolja. Za ilustraciju mogu poslužiti podaci jednog danas već klasičnog ispitivanja koje su izvršili *Laugier* i *Weinberg* u Parizu. U ono vrijeme studenti fiziologije polagali su ispit u obliku klauzurne radnje, a njihove pismene sastavke ocjenjivala su tri profesora, među koje su se takvi sastavci nasumce dijelili. Pri spomenutom provjeravanju, nakon što je redovno ocjenjivanje završeno i kandidati obaviješteni o ocjeni, zamoljeni su nastavnici da naknadno ocjene i one sastave koje

prethodno nisu ocijenili. U cjelini samo bi 27% istih kandidata prošlo, a 19% palo prema suglasnim ocjenama sve trojice ocjenjivača. Kod 54% kandidata postojala su neslaganja, tj. već prema tome kojem bi profesoru njihov rad došao na ocjenu oni bi bili prošli ili pali<sup>3</sup>.

Velik broj naših i tuđih provjeravanja, kako u srednjim školama tako i na fakultetima, pokazao je da se koeficijenti korelacije između ocjena što ih nezavisno daju kompetentni ocjenjivači istim odgovorima, kreću oko 0,70 ili još manje. Kako kvadrat koeficijenta korelacije pokazuje na postotak zajedničkih faktora znači da u ocjeni gotovo podjednako sudjeluje ispitanik sa svojim odgovorima koliko i subjektivni faktor ocjenjivača. Ili još jasnije: za kandidata je gotovo podjednako važno što je napisao ili odgovorio na ispitu koliko i koji će nastavnik te odgovore provjeriti i ocijeniti.

Spomenuti podaci odnose se na slaganje odnosno neslaganje koje je promatrano na čitavoj ljestvici ocjena. Ipak neslaganje je najveće za srednje ocjene dok je nešto manje, iako ne mnogo manje, za ekstremne ocjene: nedovoljno i odlično.

### POUZDANOST OCJENA

Za neko mjerenje ili ocjenjivanje kažemo da je pouzdano ako pri uzastopnim mjerenjima iste veličine dobivamo jednake vrijednosti.

U konkretnom slučaju ocjenjivanja na ispitu, ocjene bi bile pouzdane kad bi *isti* ocjenjivač, ocjenjujući ponovno *iste* pismene ili usmene odgovore, oba puta dao jednake ocjene.

Pouzdanost nekog mjerenja je dakako vrlo važna metrijska karakteristika, jer ako mjereći više puta isti predmet dolazimo do različitih rezultata onda takvo mjerenje ne može biti valjano.

Provjeravanje te metrijske karakteristike izvršeno je na pismenim odgovorima i na usmenim odgovorima, koji su snimljeni na magnetofonu. Kao mjera pouzdanosti služe koeficijenti autokorelacije, tj. korelacije između ocjena koje je u dva navrata dao isti ocjenjivač istim odgovorima. Ispitivanja pouzdanosti pokazala su da se ocjenjivači, u određenom vremenskom razmaku, često veoma slabo sami sa sobom slažu.

Tako npr. kod ocjenjivanja istih školskih zadaća u razmaku od godinu dana, nije izuzetan slučaj da isti nastavnik, koji je prvi put ocijenio neku zadaću s nedovoljno drugi put toj istoj zadaći dade ocjenu vrlo dobro, i obrnuto<sup>4</sup>. Gotovo jednako velike divergencije nađene su i pri uzastopnom ocjenjivanju usmenih odgovora. Posebno zabrinjuju velike razlike u proporciji prolaznih i neprolaznih ocjena. Broj nedovoljnih ocjena, npr. kod istih nastavnika za iste odgovore varirao je u jednom slučaju od 46% više nego prvi put do 57% manje nego prvi put. Pritom je zanimljivo, da i uz značajno povećan broj nedovoljnih ocjena svi oni koji su prvi put dobili tu ocjenu, ipak nisu i drugi put tako ocijenjeni.

Ocjene istog ocjenjivača, dane u dva navrata istim odgovorima, bolje se ipak između sebe slažu, nego ocjene različitih ocjenjivača. To je i razumljivo, jer kad isti ocjenjivači ocjenjuju iste odgovore, onda oba puta djeluju uz kvalitet odgovora i neki slični subjektivni faktori ocjenjivača. Autokorelacija je izraz objektivnih i nekih relativno stabilnih subjektivnih faktora.

## OSJETLJIVOST OCJENJIVAČA

Pri kvalitetnom mjerenju tražimo da mjerni instrument bude i osjetljiv, tj. da s pomoću njega možemo utvrditi i male razlike u veličini pojave koja je predmet mjerenja. Ali jasno je da je osjetljivost nekog mjernog instrumenta samo onda odlika ako je ona opravdana valjanošću instrumenta.

Kod tradicionalnog ocjenjivanja redovito susrećemo anomaliju da ocjenjivači nastoje razlikovati i ono što ne mogu ispravno diferencirati. Drugim riječima, njihova je osjetljivost samo subjektivna i neopravdana pa je dalji izvor pogreške umjesto da bude pozitivna mjerna karakteristika.

Ako isti nastavnik može u razmaku od godine dana istom odgovoru dati ocjenu koja se od predhodne razlikuje i za tri pune jedinice, onda je dakako i ljestvica od pet ocjena preširoka, a da ne govorimo o opravdanosti među-ocjena unutar takve ljestvice.

## TAČNOST OCJENE

Kad neko mjerenje ili procjenjivanje nije objektivno ni pouzdano ono dakako ne može biti ni tačno.

Za neko mjerenje kažemo da je tačno ako se dobivene vrijednosti podudaraju sa stvarnom veličinom predmeta mjerenja.

U slučaju ocjenjivanja, ocjena bi bila tačna kad bi ona odgovarala stvarnom znanju učenika. Pritom treba naglasiti da se tačnost ocjene može promatrati na dva načina. Prvo kao relativna mjera znanja ispitanika unutar jednog razreda ili godišta, drugo kao opće valjana mjera znanja bez obzira na stupanj i raspon znanja konkretne skupine ispitanika. U prvom slučaju ocjene bi bile tačne kad bi reflektirale barem rang koji zauzima ispitanik unutar skupine u kojoj se nalazi (ordinalna procjena). Naprotiv, prema drugom, opravdanim zahtjevu, tačna ocjena bi bila ona koja ispravno izražava općenito stupanj znanja do kojeg je došao neki ispitanik; stupanj znanja koji je određen prema opće valjanim i stabilnim normama (kardinalna procjena).

U kojoj su mjeri ocjene tačne može se pokušati odrediti na dva načina. Jedan je da se ocjene, koje daju pojedini ocjenjivači istim pismenim ili usmenim odgovorima, usporede sa srednjim ocjenama (koje je veći broj ocjenjivača dao tim istim odgovorima). Pritom je pretpostavka da se u prosječnoj ocjeni poništavanju različite individualne i slučajne osobine ocjenjivača, tako da prosječna ocjena odražuje ono što je jedino zajedničko pri mjerenju, a to je sam odgovor. Drugi je način da se dane ocjene usporede s nekim drugim »valjanijim« kriterijem aktualnog znanja ispitanika. U tu se svrhu, kako ćemo kasnije vidjeti, redovito upotrebljavaju rezultati u testovima znanja.

Naša provjeravanja tačnosti ocjena u srednjoj školi, s pomoću kriterija prosječne ocjene, pokazala su da moguća pogreška koja se vezuje uz ocjene iznosi približno jednu jedinicu<sup>4</sup>. Praktički to znači da »prava« ocjena nekog odgovora može teoretski biti svaka vrijednost koja se nalazi u intervalu: dana ocjena  $\pm 1$ .

U nekom drugom našem ispitivanju tačnosti srednjoškolskih ocjena, ovaj put s pomoću testova znanja, utvrdili smo da nastavnici pri davanju ocjena ne mjere znanje prema nekim opće valjanim kriterijima, nego da u stvari

ranguju svoje učenike prema njihovom relativnom znanju unutar razreda u kojem se nalaze. Pritom oni učenici koji su najbolji u svom razredu redovno dobivaju najbolje ocjene, a najslabiji u svom razredu dobivaju slabe ocjene i to u oba slučaja bez obzira koliko iz tih predmeta znaju učenici u drugim razredima istovrsnih škola. Iz toga slijedi da nominalno ista ocjena može odgovarati veoma različitom znanju, bar onom znanju koje dolazi do izražaja u rezultatima u testu znanja. U općenito boljem razredu nominalno iste ocjene odgovaraju mnogo većem znanju nego u općenito slabijem razredu.

Ta često ekstremna nepouzdanost ocjene, kao opće valjane mjere znanja, dolazi jasno do izražaja ako usporedimo kako se distribuiraju ocjene u različitim razredima, a kako rezultati u testu znanja iz istog predmeta. U nekom našem ispitivanju te vrste primijenili smo četiri testa za ispitivanje znanja i to iz matematike, materinjeg jezika, biologije i fizike i to na 580 učenika u 39 završnih razreda gimnazija u različitim republikama naše zemlje<sup>5</sup>. Na osnovi prosječnog rezultata u testu znanja mogli smo podijeliti razrede na one koji su iznad općeg prosjeka i na one koji su ispod općeg prosjeka. Distribucije rezultata se u testu takvih razreda dakako značajno razlikuju, naprotiv distribucije njihovih ocjena praktički su identične. Ta pojava vrijedi čak i za slučaj kad usporedimo ocjene i rezultate u testu između dva najbolja i dva najslabija razreda u čitavoj skupini. Pri takvu uspoređivanju može se naći npr. da najlošiji ispitanik u najboljim razredima ima još uvijek bolji rezultat u testu od najboljeg učenika u najslabijem razredu. Ali usprkos toga najbolji učenik u općenito slabom razredu dobit će ocjenu odlično, dok će najslabiji učenik u najboljim razredima, premda zna više, imati ocjenu nedovoljno!

Gotovo potpuna nezavisnost ocjene i znanja, koje je određeno s pomoću testova znanja u općenitom mjerilu, dolazi do izražaja u koeficijentima korelacije između prosječnih ocjena razredâ i prosječnih uspjeha razredâ u testu. Te su korelacije bile: 0,26 za matematiku, 0,17 za biologiju, 0,03 za materinski jezik i -0,13 za fiziku! Naprotiv *unutar* istog razreda postoji mnogo bolje slaganje između ranga učenika po ocjeni i njegova ranga po rezultatu testa znanja. U spomenutom slučaju te su rang korelacije iznosile: 0,54 za matematiku, 0,36 za biologiju, 0,62 za materinski jezik i 0,43 za fiziku. Dakle, dok ocjenjivač još nekako uspijeva da u svojim ocjenama kako tako razlikuje stupanj znanja učenika unutar određene skupine, on gotovo uopće nije u mogućnosti da dade opće valjane procjene znanja. Kad pak postoje, kako je to slučaj u našoj zemlji, vrlo velike razlike u kvalitetu nominalno jednakih škola, tad nam klasifikacija po rangu unutar pojedinih razreda kao kriterij ocjenjivanja, nikako ne garantira da jednaka ocjena odražava bar približno slično znanje.

To gotovo potpuno pomanjkanje stabilnih i opće vrijednih standarda pri davanju ocjena ima teške praktičke posljedice, jer ista ocjena, pa ma gdje da je ona postignuta daje u društvu jednaka prava, kao npr. pri namještenju, pri prijelazu u neku višu školu, pri konkursu za upis na sveučilište itd.

Opći zaključak iz mnogih, kako tuđih tako i naših ispitivanja, jest da transverzalno ocjenjivanje usmenih ili pismenih odgovora nije nažalost ni objektivno, ni pouzdano, ni tačno.

Može se opravdano pretpostaviti da su ocjene, koje dobiva učenik u osnovnoj i srednjoj školi, a koje se temelje na češćem ispitivanju i dužem opažanju, nešto pouzdanije i tačnije od jednokratno danih ocjena. Ali na fakul-

tetima je longitudinalno provjeravanje i ocjenjivanje znanja prije izuzetak nego pravilo. Na sveučilištu je često prvi ispitni susret između nastavnika i studenta i posljednji susret. Ako se pak ti susreti ponavljaju onda je to više zbog nedovoljne ocjene koju je ocjenjivač dao nego iz potrebe da se umnože informacije na kojima će se osnivati ocjena.

Drugi, praktički zanimljiv podatak, koji smo našli jest da ni prekratko, ali ni predugo ispitivanje ne utječu pozitivno na valjanost ocjene<sup>6</sup>.

Kako je poznato, propisi o trajanju ispita ne osnivaju se na rezultatima nekih istraživanja, nego na impresionizmu koje se opravdava iskustvom. Ali propisi i kad postoje redovito se ne poštvaju. Nastavnici već prema broju kandidata, časovitom vremenu kojim raspoložu, kvalitetu odgovora ispitanika i svojim ličnim karakteristikama, ispituju duže ili kraće. Razlike u tipičnom trajanju ispita kod različitih nastavnika mogu katkada poprimiti iznenađujuće razmjere. Tako npr. na jednom našem fakultetu, profesori iste struke, na ispitu jednakog ranga, ispituju u prosjeku jednog kandidata od nekoliko minuta do nekoliko sati.

U toku ispita ispitivač nastoji doći do određenog kvantuma informacija o znanju kandidata, a dužina ispita je determinirana u prvom redu od broja informacija koje ispitivač smatra dovoljnim da donese svoju procjenu. Sigurno je da taj kritički broj informacija zavisi i od kvaliteta i jednoznačnosti informacija, ali je isto tako tačno da on zavisi i od ličnih osobina ispitivača, u svakom slučaju više nego od opravdane sigurnosti.

Protivno od pravog mjerenja, kod kojeg su procjene parametara to tačnije što je uzorak podataka veći, ocjenjivača čini se zbunjuje velik broj informacija, jer porastom broja informacija raste i vjerojatnost da sve informacije neće biti istog smjera. U prilikama ispita, ocjenjivač se ponaša kako to teorija informacija i predviđa, tj. tačnost se procjene najprije povećava kad se malom broju informacija pridruži nekoliko novih, ali tačnost se opet smanjuje kod velikog broja informacija zbog neizbježive redundance informacija.

Spoznaja da su tradicionalne ocjene znanja sa statističkog gledišta slaba mjera stvarnog znanja učenika, dovela je do niza istraživanja koja su imala za cilj da utvrde uzroke te činjenice.

Kako je kod ocjenjivanja ocjenjivač ujedno i mjerni instrument to bi teoretski na ocjene mogla djelovati tri glavna faktora i to: osobitosti predmeta mjerenja, sredstva kojim mjerenje vršimo i tehnika mjerenja. Premda je detaljna analiza različitih izvora pogreške koji su vezani uz te faktore važna za poboljšanje procesa tradicionalnog ocjenjivanja, ja ću se, zbog ograničenog prostora, na tome samo ukratko zadržati.

Pogledajmo prvo glavne poteškoće na koje nailazi ocjenjivanje, a koje su u vezi s *predmetom mjerenja*.

Svako mjerenje i procjenjivanje, u pogledu predmeta koji se mjeri, zahtijeva: 1. da ono što se mjeri ili procjenjuje bude jednoznačno definirano; 2. da ono što se mjeri ili procjenjuje odgovara jedinstvenom tipu aktiviteta, tj. da ne uključuje više različitih momenata koji mogu varirati nezavisno, i 3. da su reakcije koje se mjere ili procjenjuju reprezentativni i valjani uzorak predmeta mjerenja. Pod vidom tih zahtjeva ocjenjivanje znanja nailazi na velike poteškoće.

Ono što se definira kao znanje posjeduje svoj kvantitativni, kvalitativni i upotrebnii aspekt, a ti aspekti ne idu uvijek uporedo. Netko npr. može pam-

titi velik broj činjenica, principa i zakona, a da pritom slabo može kombinirati podatke i smisleno koristiti upamćene principe i zakone. Drugi opet može na teoretskom planu i s malim brojem činjenica uspješno i s razumijevanjem baratati, a da svoje znanje slabo u praksi može pretvoriti u prikladnu akciju, itd.

Sve ako i apstrahiramo od nekih drugih momenata, koji uza znanje mogu impresionirati ocjenjivača, kao što su to npr. marljivost i disciplina, već ta polivalentnost aspekata znanja nužno će dovesti do različitih ocjena, već prema tome kojem aspektu znanja pridaje ocjenjivač najveću važnost.

Drugi momenat, koji zavisi od predmeta mjerenja, a koji također interferira s tačnošću procjena, je često nereprezentativnost reakcija u kojima se očituje znanje kandidata i često slaba vrijednost odgovora kao simptoma znanja.

Pri ocjenjivanju, znanje kandidata nije dostupno direktno nego samo indirektno, tj. ono što se ocjenjuje to su odgovori, a ne neposredno ono što kandidat zna. Jasno je da će od toga koliko su konkretni odgovori ispitanika reprezentativni za njegovo znanje i koliko su u skladu s tim znanjem, zavisiti i valjanost ocjene. Premda nema sumnje da između znanja i kvaliteta odgovora postoji pozitivna korelacija, ipak ta korelacija nije potpuna. Osim toga i kad se u konkretnom odgovoru valjano očituje znanje kandidata iz područja na koje se odnosi konkretno pitanje, ostaje problem koliko je opravdano iz nužno ograničenog broja pitanja i odgovora zaključivati na znanje iz čitavog predmeta koji se ocjenjuje.

Nadalje, svaki odgovor ima svoju sadržajnu i svoju formalnu stranu pa dok sadržaj odgovora još i zavisi od znanja, na oblikovanje odgovora — uza znanje — djeluju i određene vještine. Budući da se ocjenjuje odgovor, razumljivo je da će doći do određenog precjenjivanja znanja onih koji se lako izražavaju, a podcjenjivanja onih koji teško oblikuju ono što znaju. Neadekvatnost i neodređenost odgovora sile nastavnika na interpretacije, a to subjektivno različito dekodiranje dovodi i do različitih procjena.

Konačno na odgovor djeluju i neki drugi faktori, koji su tek djelomično ili nikako u vezi sa znanjem, npr. emotivna stabilnost kandidata, njegova opća snalažljivost itd.

Još su brojniji izvori pogreške koji zavise od *nastavnika* kao mjernog sredstva.

Kako je poznato svaki mjerni instrument posjeduje određenu skalu i indikator, čiji položaj ukazuje na veličinu pojave koju mjeri. Pritom je nužno da kriterij prema kojem se mijenja položaj indikatora na skali bude definiran i stabilan.

Pri ocjenjivanju objektivni kriterij, prema kojemu bi trebalo davati ocjene, sastojao bi se u programu predmeta, koji se ispituje. Ali dok programi često vrlo iscrpno sadrže ono što kandidat treba da nauči, oni u pravilu ne sadrže nikakve propise koje osnovne dijelove građe treba kandidat da nauči za prolaznu ocjenu i koliko toga dodatno treba da zna pa da dobije ocjenu određenog stupnja iznad prolazne ocjene. U praksi, programi predstavljaju redovno maksimalne zahtjeve, a kako nema objektivnog kriterija, to svaki ispitivač može, u granicama tih maksimalnih zahtjeva, postavljati svoje vlastite kriterije. Drugim riječima, ocjenjivač je nalik nekakvom instrumentu koji bi sam definirao kriterije svojih registracija. Posljedica je toga da ocjenjivači,



iako mjere nominalno istu pojavu (znanje) i upotrebljavaju jedinice istog imena (ocjene), čine to na skalama različitih vrijednosti<sup>7</sup>. Nema sumnje da je najvažniji uzrok netačnost i nekomparabilnost ocjena upravo ta neodređenost kriterija ocjenjivanja.

Ispitivanja, što sve utječe na visinu kriterija prema kojem nastavnici daju ocjene, pokazala su da su među relativno stabilnim faktorima najvažniji: lična jednadžba ocjenjivača i njegova stručna sprema.

Lična jednadžba se obično definira kao relativno stabilna ocjena ocjenjivača koja se očituje u karakterističnom pomaku ocjena u smjeru ekstrema na kontinuumu blagost-strogost. Ali osim takve za kandidate najvažnije karakteristike ocjenjivača, distribucije ocjena, koje daju različiti ocjenjivači, često imaju i drugi oblik kao leptokurtičan, platikurtičan, bimodalan, itd. Te distribucije svjedoče da su razlike u kriteriju ocjenjivanja, koje su u vezi s ličnošću ocjenjivača, mnogo dublje i raznovrsnije nego što se to obično drži.

U pogledu stručne spreme nastavnika čini se da što je ona bolja to je u prosjeku i nešto viši kriterij ocjenjivanja. Naprotiv, ni dob ni dužina staža nastavnika ne utječu značajno na visinu kriterija<sup>8</sup>.

Ocjenjivač nije samo »instrument« koji sam određuje svoj kriterij ocjenjivanja, nego ga on i neprestano mijenja. Osim slučajnih faktora, koji djeluju dvosmjerno na promjene kriterija, kao što su to raspoloženje, zdravlje, umor itd., postoje neki faktori koji u predvidivom smjeru utječu na kriterij nastavnika. Među tim faktorima najvažniji su već spomenuto prilagođivanje kriterija ocjenjivanja konkretnom znanju kandidata i tzv. halo-efekt.

Kad nastavnik pristupa ispitivanju on redovno ima na početku neki kriterij, rezultat iskustva i njegovih ličnih osobina. Ali odgovori njegovih ispitanika mogu u toku samog ispitivanja znatno izmijeniti početni kriterij. Poznato je tako da se nakon nekoliko dobrih ispita kriterij ocjenjivanja redovito diže, i obrnuto, da niz loših kandidata snizuje kriterij prema kojem se daju ocjene.

Halo-efekt je nažalost česta pojava koja se sastoji u neopravdanoj generalizaciji nekog općeg suda na pojedinačne sudove. Kod ocjenjivanja halo-efekt se očituje u prilagođivanju ocjena iz jednog predmeta ocjenama koje je kandidat dobio iz drugih predmeta. Ocjenjivač kao da vlastitu nesigurnost umiruje povodeći se za jednako tako nesigurnim ocjenama drugih ocjenjivača.

Konačno i u raznolikosti u *tehnika*ma ispitivanja i ocjenjivanja kriju se određeni izvori pogrešaka.

Bitno je za svako pravo mjerenje da se ono vrši u standardiziranim prilikama na način za koji je utvrđeno da je u tu svrhu najbolji. Naprotiv, u praksi ispitivanja i ocjenjivanja, nailazimo na veliku šarolikost pristupa. Tako postoje pasivni i aktivni ispitivači, ispitivači koji postavljaju velik broj užih pitanja i oni koji postavljaju manji broj širih pitanja, ispitivači koji ispituju dugo i oni koji ispituju kratko, ispitivači koji posebno ocjenjuju svaki odgovor i oni koji daju svoju ocjenu na osnovu opće impresije itd.

Posebnu poteškoću čini nužno ograničeni broj pitanja koje nastavnik može postaviti u toku ispita. Na taj način kandidati se stavljaju pred nejednake zadatke, a ocjena često zavisi od slučaja koja je pitanja kandidat dobio. Reprezentativnost pak uzorka pitanja je problem, za koji ispitivači i ne znaju.

Ukratko: pri ocjenjivanju ne postoje nikakvi ujednačeni ni stabilni kriteriji, znanje se provjerava različitim zadacima i to skupom, a ne uzonkom pitanja, tehnika ispitivanja i ocjenjivanja je različita, a to se sve nužno negativno odrazuje na pouzdanost i valjanost ocjene kao mjere znanja.

Etiologija poteškoća, na koje nailazi i najsavjesniji ocjenjivač, može poslužiti u terapiji našeg tradicionalnog načina ocjenjivanja. Ali kako to nije zadatak ovog prikaza, navest ću samo taksativno glavne korektivne mjere. Među njima su najvažnije: 1. informiranje nastavnika o različitim izvorima pogrešaka pri ocjenjivanju i uvježbavanje nastavnika u toj funkciji važnoj za pojedinca i društvo; 2. propisivanje najbolje opće tehnike ispitivanja i ocjenjivanja; 3. uvođenje pravog komisijskog ispitivanja i ocjenjivanja gdje god je moguće; 4. smanjenje opsega ljestvice ocjena; 5. kontrola distribucije ocjena i korigiranje lične jednadžbe; 6. češće provjeravanje znanja na vježbama, seminarima, pa čak i na predavanjima. Skupljanje longitudinalnih informacija i korištenje u nastavi i ocjenjivanju dvostrane povratne informacije, tj. povratnu informaciju studentu o njegovu znanju i povratnu informaciju nastavniku o tome kako je organizirao i proveo svoju nastavu; 7. izrada unutar svakog predmeta nekih bar približnih normi za ocjenjivanje.

Porazna kritika, sa statističkog gledišta, objektivnosti, valjanosti i komparabilnosti tradicionalnog načina ocjenjivanja, dovela je do pokušaja da se radikalno izmijeni način provjeravanja i ocjenjivanja znanja. Taj pokušaj sastojao se u zamjeni ispitivača *testom znanja*.

Prema definiciji psihološki test je standardizirani postupak kojim se izazivaju manifestacije neke psihičke osobine i mjere te manifestacije na osnovi usporedbe individualnog rezultata s prosječnim vrijednostima uzorka u koji po svojim bitnim obilježjima ide ispitanik.

Kako je poznato, testovi znanja sastoje se iz većeg broja posebno sastavljenih zadataka i pitanja, koja se odnose na sadržaje iz nekog područja znanja. Postoje tako testovi koji su namijenjeni provjeravanju znanja iz različitih nastavnih predmeta, zatim testovi kojima se provjerava stupanj u kojem su usvojene različite vještine i profesionalna sprema, do testova za ispitivanje opće obrazovanosti i kulture.

Testovima znanja nastoji se utvrditi u kojoj je mjeri neki ispitanik stekao određena znanja ili vještine. Testovi znanja razlikuju se od testova sposobnosti po tome što se s pomoću njih nastoje provjeriti ne toliko opće intelektualne mogućnosti ispitanika, koliko rezultati specifičnog učenja ili treninga. Osim toga druga je bitna razlika što se testovima znanja mjeri veličina na koju se može utjecati, tj. koja je i kod istog ispitanika u vremenu promjenljiva.

Od tradicionalnog provjeravanja znanja, testovi znanja razlikuju se po tome: 1. što oni u cijelosti isključuju djelovanje subjektivnog faktora ocjenjivača, 2. što su prilike ispita i ocjenjivanja standardizirane, i 3. što se vrednovanje rezultata može provesti prema stalnom unaprijed dogovorenom kriteriju, koji je jednak za sve ispitanike pa ma gdje se oni nalazili.

U testu znanja subjektivni faktor ocjenjivača isključen je sastavom zadatka, koji su tako oblikovani da je odgovor ispitanika uvijek jednoznačan, tj. ne dopušta različitu interpretaciju ocjenjivača. To se postiže ili ograničavanjem odgovora na samo jednu riječ, broj ili neki drugi podatak, odnosno

predlaganjem, uza svaki zadatak, većeg broja odgovora, među kojima ispitanik treba da odabere samo jedan i to onaj koji je po njegovu znanju najtačniji.

Pri upotrebi testova znanja prilike su ispita standardizirane, jer su svi ispitanici stavljeni pred jednake zadatke, a njihovi odgovori ocijenjeni po istom ključu. Osim toga, djelovanje slučajnog izbora pitanja i njihove nereprezentativnosti uklonjeno je, jer test redovno sadrži velik broj pitanja, kojima se može prekriti čitava građa, koja je predmet ispitivanja. Konačno, jer su svi kandidati ispitani istim zadacima, a njihovi odgovori ocijenjeni prema istom ključu, moguće je njihove rezultate vrednovati prema unapred dogovorenom općenitom i stabilnom kriteriju.

Na prvi pogled testovi znanja rješavaju tako sve glavne poteškoće s kojima se sudara tradicionalno ocjenjivanje. Testovi su znanja objektivni u primjeni i korekciji, oni su na kraći rok visoko pouzdani, zbog velikog broja zadataka su reprezentativni i posjeduju značajnu osjetljivost, i što je veoma važno, oni omogućuju da ista ocjena, pa ma gdje postignuta, reflektira i jednak uspjeh u testu.

Određene poteškoće nastaju istom onda kad pristupimo izradi testova znanja i provjeravanju njihove valjanosti u svrhu *individualne dijagnoze znanja*.

Konstruirati neki upotrebljivi test znanja nije lak posao i taj rad zahtijeva usku suradnju predmetnog stručnjaka i psihometričara. Osim toga, svaki test znanja mora prethodno proći kroz duga ispitivanja u toku kojih se za svaki pojedini zadatak posebno određuje diskriminativna valjanost, relativna težina, jednoznačnost, samostalnost, a kod zadataka višestrukog izbora i atraktivnost pojedinih distraktora. Na osnovu takvog testiranja testova s pomoću ispitanika, mijenjaju se zadaci ondje gdje je to potrebno i rekonstruira se test prije nego što se s pomoću njega počnu testirati ispitanici. Uz to postoji i čitav niz posebnih pravila kojih se treba držati pri konstrukciji zadataka različitog tipa. To spominjem samo zbog toga što je u posljednje vrijeme kod nas zavladała prava pošast sastavljanja kojekakvih zadataka više ili manje objektivnog tipa, koji se, bez ikakve osnove, proglašuju testovima znanja. Ti paratestovi su samo testovi neznanja tih psihometrijski nevinih sastavljača.

Ali i kod dobro sastavljenih testova znanja ostaje otvoreno pitanje njihove valjanosti. Govoreći jezikom psihometrije, valjanost većine testova znanja je samo aprioristička, tj. sadržajna valjanost. Zbog toga što se zadaci u testu znanja odnose na sadržaj programa iz nekog predmeta, a priori se pretpostavlja da odgovori na takve zadatke ukazuju na stupanj znanja ispitanika iz predmeta ispitivanja.

Valjanost nekog testa, kao sredstva za dijagnozu znanja, mogli bismo odrediti tek kad bismo imali, nezavisno od rezultata u testu, tačne podatke o stvarnom znanju ispitanika. Kako s takvim podacima ne raspoložemo to se aprioristička valjanost katkada kuša dopuniti indirektnom empiričkom validacijom. Pritom se kao kriterij valjanosti upotrebljava stupanj napretka u testu u toku studija (što je uglavnom opravdano samo kod vještina), ili slaganje s tradicionalnim školskim ocjenama.

Kako smo vidjeli, uspoređivanja slaganja ocjena s rezultatima u testu znanja unutar jednog razreda, pokazalo je da se te korelacije kreću oko 0,50. Razumije se, kad već znamo da ocjene nisu valjana mjera znanja, to relativno

slabo slaganje nije nešto što bi jednoznačno govorilo protiv valjanosti testova znanja. Ali te su korelacije toliko niske da ih je teško pripisati samo nepouzdanosti ocjena. Postoje indikacije da je slabo slaganje između rezultata u testu i ocjena istih učenika uvjetovano i razlikama u predmetu koje mjeri test znanja s jedne strane, a nastavnik s druge strane.

Poznato je tako da je relativno lako sastavljati zadatke za testiranje u kojem su opsegu usvojeni različiti podaci i činjenice, dok je veoma teško sastaviti takve zadatke kojima se testira razumijevanje, kritičko prosuđivanje, kombiniranje i primjena općih principa i zakona. Drugim riječima, dok testovi znanja uspješno mjere tzv. faktografsko znanje, oni su manje valjani pri određivanju instrumentalnog i sintetičkog znanja. To je uostalom i shvatljivo jer test kao instrumenat, koji je konstruiran na osnovu statističkih podataka, nužno i registrira u prvom redu ono što je više ili manje zajedničko svim ispitanicima, a znatno slabije odražava kvalitativne razlike koje postoje među njima. U poznavanju golih podataka i činjenica mogu među ispitanicima postojati kvantitativne razlike, dok u pogledu kombiniranja, razumijevanja i korištenja podataka, principa i zakona, postoje velike kvalitativne diferencije, već i zbog toga što se ispitanici razlikuju po svojoj općoj kulturi, svome predznanju, svojim interesima i svojim intelektualnim mogućnostima.

Od mnoštva varijeteta znanja, testovi znanja prvenstveno mjere faktografski ili materijalni aspekt, a znatno slabije znanje u užem smislu, koje se sastoji u sistemu spoznaja, svojevrsnoj strukturi osmišljenih fakata, a koje predstavlja fleksibilni instrument za korištenje iskustva. Ti individualno različiti kvalitativni aspekti znanja nesumnjivo pretpostavljaju i faktografsko znanje, ali nisu s njim u potpunoj korelaciji. U tom pogledu nastavnik kao mjerni instrument uza sve svoje krupne metrijske nedostatke, ima odliku da se može plastično prilagoditi specifičkim karakteristikama određenog varijeteta znanja. Da je to uistinu tako dokazuje i raznolikost načina kojima pristupa nastavnik ispitivanju znanja različitih kandidata, nastojeći da unutar individualnog iskustva i pristupa nađe mjerilo za procjenu. Tako će npr. ispitivač, ako ga tim putem vode odgovori, s jednim kandidatom pretežno diskutirati, drugog će kandidata stavljati pred probleme, koji se mogu riješiti kombiniranjem parcijalnih znanja, od trećeg će kandidata, ako ne ide drugačije, tražiti reprodukciju činjenica i zakona, itd. Na taj način, standardiziranje pitanja i prilika ispita, što predstavlja nesumnjivu prednost pri statističkom provjeravanju znanja kao jedinstvene kategorije, ima određene nedostatke pri provjeravanju različitih aspekata znanja kod pojedinaca. Razumije se, i u testove znanja nastoje se unijeti neki tipovi zadataka, koji stavljaju u funkciju složenije i kreativnije aspekte znanja, ali usprkos tih nastojanja, današnji testovi znanja pretežno su upravljani na pretpostavke pravom znanju.

Daljna razlika, koja bi mogla biti odgovorna za divergencije između rezultata u testovima znanja i ocjena nastavnika, leži u različitosti vrednovanja odgovora. Tipično je za testove znanja, da se građa mrvči na takve čestice da je znanje u odnosu na te čestice alternativnog tipa po principu »sve ili ništa«. U svakom zadatku, koji je uključen u neki test znanja, samo je jedan odgovor ispravan, dok se sa svim ostalim odgovorima postupa kao jednako pogrešnim. Taj stav, koji je inače rezultat opravdanog nastojanja da odgovori ispitanika budu jednoznačni, u određenom je sukobu s realnošću u kojoj kompleksno »znanje« nije gotovo nikada alternativnog tipa već se distribuira na rasponu

od potpuno tačnog znanja, preko približnog znanja, neznanja, antiznanja sve do teške ignoracije i apsurdnosti. Naprotiv, nastavnik pri svom ocjenjivanju, uzima u obzir ne samo sasvim tačne odgovore, nego i približne odgovore, a penalizira, i to često oštro, antiznanje i apsurdnosti. I u tom pogledu, da nije drugih nedostataka, nastavnik bi bio bliže pravom predmetu mjerenja od testa. Mi smo, u toku protekle godine, pristupili konstrukciji testova znanja novog tipa, u kojima se upravo tom nijansiranju vrijednosti odgovora nastojalo dati značajno mjesto. Umjesto da tražimo od ispitanika da samo označi odgovor, koji je po njegovu mišljenju najtačniji, njegov je zadatak da među predloženim odgovorima pronađe i onaj koji po vrijednosti dolazi odmah nakon najboljeg odgovora, kao i onaj koji je najmanje tačan. U drugoj varijanti, s tipom zadataka višestrukog izbora, čiji su odgovori skalirani po vrijednosti, vodili smo računa uz pravo znanje i o približnom znanju, a penalizirali smo negativnim bodovima antiznanje i apsurdnosti. Zanimljivo je da se rezultati u tako konstruiranim testovima bolje slažu s ocjenama nastavnika nego što je to slučaj s testovima znanja klasičnog tipa<sup>9</sup>.

Konačno, u ovom nabrajanju ograničenja u upotrebi testova znanja, treba spomenuti i nezgodnu reperkusiju koju mogu imati testovi znanja kako na nastavu tako i na učenje. Standardizacija testova znanja, koja u statističkom pristupu predstavlja odliku, može dovesti i do određene standardizacije instruiranja, pogotovu ako se na osnovu rezultata, koje postižu učenici u testu, ne procjenjuje samo njihovo znanje, nego i uspjeh nastavnika. Standardizacija nastave, pogotovu na Sveučilištu, nužno bi interferirala upravo s onim što je karakteristično za tu nastavu, a to je podsticanje na kreativnost, originalnost i kritičnost. Još teže posljedice može imati isključiva upotreba testova znanja na način učenja studenata. Kako je poznato, većina studenata uči u prvom redu za ocjenu i zbog toga, kad se ocjene daju samo na osnovu rezultata u testu, nužno će i učenje biti upravljeno na onaj tip znanja koji honorira test. Uz to će standardizirani testovi dovesti neizbježno i do standardizacije i uniformnosti znanja, koje će se od studenta do studenta razlikovati u prvom redu po svom kvantitetu, a manje po originalnosti i kvalitetu organizacije.

Da rezimiram: nastavnici, kao mjerni instrumenti znanja, imaju mnogo nedostataka. U našim i tuđim istraživanjima ti su nedostaci nešto dramatiizirani, jer se provjeravanje uglavnom vršilo statistički na ocjenama, koje su dane odgovorima, a ne ispitanicima. Ali, s druge strane njihov plastičan individualni pristup ima nesumnjivih prednosti pri procjenjivanju individualno različitih oblika sintetičkog i instrumentalnog znanja. Testovi znanja opet, na svom današnjem razvojnom stupnju, mogu vrlo dobro poslužiti pri utvrđivanju prosječnog znanja pojedinih razreda, kvaliteta pojedinih škola, primjerenosti nastavnih planova i programa, te pri analizi različitih faktora koji utječu na valjanost tradicionalnih školskih ocjena. Osim toga, oni su veoma korisni pri utvrđivanju i kompariranju stupnja faktografskog znanja ispitanika. Ali, nasuprot njihove upotrebljivosti pri otkrivanju onoga što je tipično i zajedničko velikom broju učenika, njihova je dijagnostička, a pogotovu prognostička valjanost ograničena kad se radi o pojedincima, a ne skupinama.

Individualni pristupi nastavnika i statistički pristup testa ispitivanju i ocjenjivanju, u mnogočemu su različiti i ti se pristupi ne isključuju nego se dopunjavaju. Zbog toga napredak u ispitivanju i ocjenjivanju ne sastoji se u napu-

štanju jednog ili drugog oblika pristupa, nego u njihovoj sintezi. Testove znanja treba upotrebljavati, ali tako dobiveni podaci su samo djelomične informacije o znanju kandidata, koje informacije treba dopuniti i interpretirati na osnovu informacija do kojih može doći nastavnik jedino u direktnom ispitnom kontaktu s ispitanikom.

Prema tome odgovor na retoričko pitanje u naslovu ovog prikaza »Ispitivač ili test?« glasi — *i ispitivač i test.*

#### L I T E R A T U R A

1. *Bujas, R., Bujas, Z. i Blašković, J.*: Napredak, 82:3, 1941. — 2. *Bujas, Z. i Sorokin, B.*: Acta Inst. Psychol. Univ. zagrab., br. 28, 1961. — 3. *Laugier, H. Piéron, H. Piéron, M., Toulouse, E. i Weinberg, D.*: Etudes docimologiques sur le perfectionnement des examens et concours, Paris, 1935. — 4. *Bujas, Z.*: Napredak, 82:4, 1941. — 5. *Roth, N. i Bujas, Z.*: Trav. hum., 22:1, 1959. — 6. *Bujas, Z. i Grgin, T.*: Acta Inst. Psychol. Univ. zagrab., br. 29, 1961. — 7. *Bujas, Z.*: Testovi znanja i mogućnost njihove upotrebe u školskoj praksi, Zagreb, 1943. — 8. *Zorman, L.*: Doba poučavanja in veljavnosti školskih ocen, disertacija, 1964. — 9. *Bujas, Z.*: Trav. hum., 28:3, 1965.

#### R É S U M É

L'ÉVALUATION DU SAVOIR — L'EXAMINATEUR OU LE TEST?

*Z. Bujas*

L'analyse critique des examens traditionnels a montré que les notes manquent d'objectivité et de fidélité et qu'en conséquence les notations ont une validité assez médiocre en tant que mesure de la connaissance des examinés.

D'autre part les tests de connaissances, objectifs et très valides en ce qui concerne la connaissance des faits, sont trop rigides pour déceler valablement les différentes variétés du savoir.

L'approche clinique et souple d'examineur à la qualité spécifique du savoir d'examiné et l'approche objective et statistique du test de connaissances, sont deux modalités qui possèdent certains avantages et certains désavantages. Ainsi l'amélioration de l'évaluation du savoir peut s'obtenir seulement en combinaison de ces deux sources d'informations.

*Faculté de Lettres*

*Reçu le 13. juillet 1966*

*Zoran Bujas, doktor filozofije, redovni profesor psihologije na Filozofskom fakultetu u Zagrebu, dopisni član Jugoslavenske akademije znanosti i umjetnosti.*