

Виолета ПЕТРОСКА — БЕШКА

ТЕСТИРАЊЕ НА УВЕРЕНОСТА ВО ТОЧНОСТА НА ИЗБРАНИТЕ ОДГОВОРИ КАЈ ТЕСТОВИТЕ ОД ТИПОТ ПОВЕКЕЧЛЕН ИЗБОР

Класичниот тест од типот повеќеџлен избор го сочинуваат задачи кои се состојат од прашање на кое треба да се одговори и од два или повеќе понудени можни одговори (алтернативи) од кои најчесто само еден е точен. Кога се задава ваков тест, од испитаниците се бара да изберат и означат која од понудените алтернативи ја сметаат за точен одговор на секој од ајтемите. Бодувањето на дадените одговори се врши на два начина, со примена на скала од два или скала од три бода. Скалата од два бода на испитаникот му доделува еден бод ако го избрал точниот одговор, а не му дава ниеден бод ако се определил за еден од неточните одговори или ако не направил никаков избор (не заокружил ниту една од понудените алтернативи). Резултатот од тестот во целина (тест-скорот) се определува како збир од ајтемите на кои е даден точен одговор. Кога пак, се користи скалата од три бода, испитаникот добива еден бод за секој точен одговор, ниеден бод за секој пропуштен одговор, а негативен дел од бодот за секој неточно избран одговор (колкав ќе биде негативниот дел од бодот зависи од бројот на понудените алтернативи). Со најпопуларната постапка од овој вид, тест-скорот се пресметува со примена на таканареченото „бодување по формула“ (formula scoring):

$$X = T - \frac{N}{a - 1}$$

каде „Т“ го претставува бројот на точните одговори, „Н“ бројот на неточните одговори, а „а“ бројот на понудените одговори по ајтеми.

Ваквиот начин на задавање и бодување на тестовите е многу едноставен и економичен: испитаникот може да решава голем број ајтеми за определено време, а испитувачот може

да ги оцени и да ги бодува одговорите за многу кратко време. Меѓутоа, ваквиот начин на тестирање има и недостатоци.

Еден сериозен недостаток е можноста да се погоди точниот одговор и кога не се знае. Кога бодувањето се врши на скала од два бода, испитаникот кој сака максимално да го зголеми својот скор, одговара на сите ајтеми без разлика дали ги знае точните одговори или не. Тогаш кога не знае кој одговор е точен тој погодува — случајно избира еден од понудените одговори знаејќи дека на тој начин може евентуално да го зголеми својот скор, без опасност да го намали. Случајниот избор на точниот одговор доведува до зголемување на варијансата на грешка во тоталниот скор на тестот, што има за последица намалување на валидноста и релијабилноста на применетиот тест. Скалата од три бодови се применува под претпоставка дека кога испитаниците одговараат неточно, го прават тоа случајно избирајќи една од сите понудени алтернативи. Меѓутоа, испитаниците и кога погодуваат, често знаат да елиминираат дел од неточните алтернативи и прават случаен избор меѓу помал број од понудените одговори. Од таа причина, како и заради постоењето на индивидуални разлики во тенденцијата за погодување (некои луѓе секогаш ризикуваат, други не се осмелуваат) и овој начин на бодување придонесува за намалување на релијабилноста и валидноста на добиените скорови.

Друг сериозен недостаток на класичните постапки на задавање и бодување на тестовите од типот повеќечлен избор, е неможноста да го мерат парцијалното знаење на испитаниците. Под парцијално знаење се подразбира она знаење кое му овозможува на испитаникот да ги разликува подобрите неточни алтернативи (оние што се поблиски со точниот одговор) од полошите (подалечните) во случај кога не може да го идентификува точниот одговор. Техниката на бодување, која би го зела во предвид парцијалното знаење на испитаниците, би водела кон порелијабилни и повалидни процени на нивното ниво на постигнување.

Тенденцијата да се надминат недостатоците на традиционалното (конвенционално) тестирање, довела до појава на техниките на тестирање на увереноста кои имаат за цел да овозможат проценување на парцијалното знаење на испитаниците и намалувањани на ефектите на погодувањето кај објективните тестови од типот повеќечлен избор.

Тестирањето на увереноста (confidence testing) се разликува од конвенционалното тестирање и по начинот на кој испитаниците одговараат и по начинот на кој добиените одговори се бодуваат. Тестирањето на увереноста е посебен метод на тестирање, кој подразбира директно или индиректно

припишување на пондери на ајтем-одговорите на начин кој ја одразува увереноста на испитаникот во точноста на избраната алтернатива.

Терминот тестирање на увереноста се однесува на повеќе различни постапки кои бараат од испитаникот да ја означат својата увереност во точноста на понудените одговори на еден ајтем. Сите тие постапки опфаќаат главно два пристапа: пондерирање на увереноста (confidence weighting) и пробабилистичко тестирање (probabilistic testing). Кога се применуваат постапките на пондерирање на увереноста, од испитаникот се бара да означат кој одговор го смета за точен и колку е сигурен во точноста на означениот одговор. Од друга страна, постапките на пробабилистичкото тестирање бараат испитаникот да припише определени пондери на секоја од алтернативите во зависност од увереноста во нивната точност.

Голем број автори сметаат дека тестирањето на увереноста треба да ги вклучи и техниките на таканареченото елиминационо тестирање (elimination testing) кај кое увереноста во точноста на алтернативите се изразува преку идентификување на неточните алтернативи (дистракторите). Кога се применува елиминационо тестирање, од испитаниците се бара да определат кои од понудените одговори се неточни, а нивниот скор се пресметува како збир од точно идентификувани дистрактори.

Пондерирање на увереноста

Ebel (1965 a) го опишува пондерирањето на увереноста како „посебен начин на одговарање на објективните тест ајтеми и како посебен начин на бодување на дадените одговори. Општо кажано, од испитаникот се бара да означат не само кој одговор го смета за точен, туку и колку е уверен во точноста на избраниот одговор. Кога неговите одговори се бодуваат тој добива повеќе бодови за точните одговори дадени со голема увереност, отколку за оние дадени без увереност. Но, за секој неточен одговор, даден со голема увереност, во неговата точност се добива доста висока казна која обесхрабрува несакано присуство на увереност“ (стр. 49).

Иако почетоките на пондерирањето на увереноста датираат уште од почетокот на овој век, историјата на пондерирањето на увереноста како метод за зголемување на количеството информации од објективните тестови, започнала триесеттите години кога неговите постапки почнале да се применуваат со цел да се сведат на минимум ефектите од погодување во тестовите составени од ајтеми од типот двочлен избор (точно-неточно). Во 1932 година Nevner спровела истражување со намера да

утврди до кој степен пондерирањето на увереноста ја зголемува релијабилноста на тестовите од областа на естетското и музичкото резонирање. Таа применила тестови од типот точно-неточно на тој начин што од испитаниците барала да изберат кој од две музички нумери е помелодична и да го означат степенот на увереноста во точноста на своите судови определувајќи се за А — ако не се многу уверени, за В — ако се прилично уверени, или за С — ако не се воопшто уверени. Тест-скоровите на испитаниците биле пресметувани на четири начини: (а) како сума од точните одговори, (б) како разлика меѓу бројот на точните и бројот на неточните одговори в) како сума од пондерирани точни одговори (каде што точниот одговор означен со А носел 3 бодови, означен со В — 2 бода, а со С — еден бод) и (г) како разлика меѓу бројот на пондерирани точни одговори и бројот на пондерирани неточни одговори (каде испитаниците не само што биле наградувани за точните одговори, туку и казнувани за неточните: по два бода за секој неточен одговор означен со А, по еден бод за секој означен со В, а ниеден за оние означени со С). Пред задавањето на тестот биле применети објективни скали за мерење на музичкиот талент и музичката извежбаност на испитаниците, чии мерки послужиле како критериум за определување на валидноста на тестот на музичко резонирање.

Методот на бодување кој ја користел сумата од пондерирани точни одговори за пресметување на тест-скоровите (в), резултирал и во највисока релијабилност (проценета со Спирман-Брауновата постапка изнесувала 0,70 за музичко и 0,80 за естетско разбирање) и во највисока валидност на тестот и музичко разбирање (0,44 за критериумот талент и 0,57 за критериумот извежбаност). Невнер забележала дека субјектите во нејзиниот експеримент ја одобриле можноста да го изразат својот степен на увереност, особено оние кои се чувствувале помалку сигурни во врска со тестот. Таа искажала мислење дека е потребно да се чува во тајност начинот на бодување за да се спречат нечесните испитаници вештачки да ги подигаат своите скорови.

Во 1936 година Soderquist спровел слично истражување користејќи поинаков систем на бодување. Тој систем им овозможувал на испитаниците да вложуваат 4, 3 или 2 бода во одговорот на секој ајтем и на тој начин да ја означуваат својата увереност во точноста на избраниот одговор (еден бод вложувале ако не биле уверени). Пресметувањето на скорот се вршело по формулата бројот на пондерирани точни одговори минус бројот на пондерирани неточни одговори, при што пондерите за неточни одговори биле два пати поголеми од износот на бодовите кои ја означувале увереноста. Овој начин

на бодување бил применет на тест од типот точно-неточно, зададен на испитаници кои биле запознати со начинот на пресметување на скоровите. Тест-скоровите добиени на овој начин биле споредувани со тест-скоровите добиени како разлика меѓу бројот на точни и бројот на неточни одговори во поглед на релијабилноста (користејќи ја Спирман-Брауновата формула). Добиените резултати укажале на повисока релијабилност на методот на пондерирање на увереноста ($0,85 \pm 0,02$ споредено со $0,72 \pm 0,04$ за конвенционалниот метод).

Во 1953 година Dressel и Schmid презеле опширно истражување експериментирајќи со различни типови на ајтеми и системи на бодување, обидувајќи се да ја подобрат дискриминативноста на ајтемите од типот повеќеџен избор без продолжување на времето на тестирање. Тие споредиле конвенционален тест од типот повеќеџен избор со четири различни модификации на тестови: (а) тест на слободен избор (free-choice test) кој се состоел од ајтеми со еден точен одговор на кои испитаниците можеле да одговорат со заокружување на повеќе алтернативи за да бидат сигурни дека не го пропуштиле точниот одговор, (б) тест на степен на сигурност (degree-of-certainty) означат колку се сигурни во точноста на единствениот одговор кој го избрале како точен, (в) тест со повеќе одговори кој се состоел од ајтеми со повеќе точни одговори кои испитаникот требало да ги идентификува и (г) тест со два одговора во кој два од пет понудени одговори за секој ајтем биле точни. Само тестот на слободен избор и тестот на степен на сигурност се однесуваат на тестирање на увереноста и затоа само нивните системи на бодување се посебно прикажани:

Тест на слободен избор:

— ако е избран точниот одговор, а бројот на означени одговори изнесува

1	4
2	3
3	2
4	1
5	0

— ако не е избран точниот одговор, а бројот на означени одговори изнесува

1	—1
2	—2
3	—3
4	—4

Тест на степен на сигурност:		
— ако е избран точниот одговор, а означеното ниво на сигурност одгвара на	ајтем-скорот	изнесува
позитивно		4
прилично сигурно		3
рационално погодување		2
случаен избор		1
— ако не е избран точниот одговор, а означеното ниво на сигурност одгвара на	ајтем-скорот	изнесува
позитивно		—4
прилично сигурно		—3
рационално погодување		—2
случаен избор		—1

Очигледно е дека скоровите добиени со примена на овие две тестовни форми не претставуваат само мерка на знаењето на испитаниците, туку се истовремено и мерка на нивната сигурност во покажаното знаење. Меѓутоа, истражувањето покажало дека двете тестовни форми не укажуваат на еднаков однос помеѓу сигурноста во извршениот избор и покажаното ниво на знаење. Така, иако времето за задавање и решавање на тестот било продолжено за сите испитаници подеднакво, тестот на слободен збор ги диференцирал добрите испитаници (добри во погледот на покажаното ниво на знаење) од просечните и слабите, но не правел разлика помеѓу просечните и слабите испитаници. Од друга страна пак, тестот на степен на сигурност ги диференцирал сите три нивоа на испитаници скоро подеднакво успешно. Понатаму, во тестот на слободен избор, лесните ајтеми (лесни во поглед на сигурноста со која е извршен изборот од страна на испитаниците) биле успешно разделени од просечните и тешките, но просечните не се разликувале од тешките ајтеми. Истовремено, анализата на добиените резултати покажала дека во тестот на степенот на сигурност, просечните и лесни ајтеми не предизвикале различни степени на сигурност во извршениот избор, туку само дека сигурноста во изборот поврзана со тешките ајтеми била значително поголема.

Имајќи ги предвид сите направени споредувања Dressel и Schmid заклучиле дека факторот на сигурност (увереност) мерен на еден начин (со помош на еден метод) не може да се псистовети со факторот на сигурност мерен на друг начин.

Ebel (1965 b) применил поинаква постапка на пондерирање на увереноста обидувајќи се да ја намали грешката во тестирањето која се јавува како резултат на погодување. Тој го модификувал методот на бодување кој го користел уште So-

derquist во 1936 година и го применил исто така на тест од типот точно/неточно. Меѓутоа, наместо да ги ограничи одговорите на две можности — точно или неточно, Ebel предложил по пет алтернативи за секој ајтем даден во облик на тврдење и од испитаниците барал да ја означат точноста или неточноста на секое тврдење преку определување за една од понудените алтернативи. Предложените одговори и системот на бодување го имале следниот облик:

можни одговори	доделени поени		
	точен одговор	неточен одговор	неодговорен ајтем
— тврдењето е веројатно точно	2	-2	
— тврдењето е евентуално точно	1	0	
— не знам			0,5
— тврдењето е евентуално неточно	1	0	
— тврдењето е веројатно неточно	2	-2	

Авторот презентирал податоци за релијабилноста на три стандардизирани тестови на знаење бодувани конвенционално и по наведениот начин на пондерирање на увереноста. Податоците покажуваат дека неговиот начин на бодување ја подобрил релијабилноста на сите три теста: коефициентите на релијабилност за тест—скоровите добиени со конвенционалната постапка на бодување (по Кудер-Ричардсоновата формула) изнесувале 0,574, 0,765 и 0,728, додека коефициентите на релијабилност за тест-скоровите добиени со постапките на пондерирање на увереноста биле повисоки и изнесувале 0,713, 0,828 и 0,821 за истите тестови. Понатаму, Ebel покажал дека релијабилноста постигната во услови на пондерирање на увереноста може да се достигне и во услови на конвенционално бодување ако секој од оригиналните тестови се продолжи 1,84 пати, односно 1,48 пати, односно 1,72 пати (предвидено со Спирмана Брауновата постапка).

Пододнежните истражувања на Ebel (1965 а) не ги потврдиле претходните заклучоци. Добиеениот пораст не бил доволен да ја оправда сложеноста на одговарањето и бодувањето по методите на пондерирање на увереноста. Од тие причини и со цел поуверливо да ги демонстрира ефектите од различните начини на одговарање и бодување, Ebel применил неколку различни математички модели во обработувањето на одговори дадени од една група замислени студенти на еден хипотетички тест. На тој начин, авторот демонстрирал дека тестирањето на увереноста би можело да биде ефикасно кога поспособните испитаници би биле и подискриминативни при изборот на одгово-

рите кои ги даваат со голема увереност. Меѓутоа, експерименталните истражувања укажуваат дека посposобните испитаници не се секогаш и подискриминативни од помалку способните во избирањето на одговорите во чија точност се сигурни.

Hopkins, Hakstian и Hopkins (1973) ги резимираше резултатите од претходните истражувања на тестирањето на увереноста споредувајќи ги мерките на релијабилност добиени со примена на конвенционалните постапки на тестирање. Иако сите наведени студенти пружиле докази дека техниките на пондерирање на увереноста тежнеат да продуцираат порелијабилни скорови од конвенционалните техники за тестирање, Hopkins и соработниците ги критикувале, затоа што „се посветиле скоро исклучиво на откривање на ефектите врз релијабилноста, обично имплицирано претпоставувајќи дека порастот на релијабилноста предизвикува и пораст на валидноста“ (стр. 136). Поттикнути од таа констатација тие спровеле истражување кои имало за цел да провери дали порастот во релијабилноста потекнува од релевантни извори и на тој начин да придонесе за осветлување на прашањето за валидноста на постапките на пондерирање на увереноста.

Субјектите во нивното истражување биле ученици со претходно искуство во пондерирање на увереноста. Од нив се барало да означат дали се многу, средно или малку уверени во точноста на нивните одговори дадени во ајтеми од типот повеќе-член избор, знаејќи дека секој точен одговор во кој се многу уверени заслужува 3 бода, а секој неточен проследен со исто толкава увереност донесува — 3 бода за тој ајтем. Слични постапки биле применувани и кога испитаниците изразувале дека се средно или малку уверени во точноста на избраните одговори при што се доделувале 2 и 1 бод, односно — 2 и — 1 бод.

Пресметаниот коефициент на релијабилност за скоровите добизни со пондерирање на увереноста бил незначително повисок од коефициентот на релијабилност за конвенционално добиените скорови (0,92 према 0,88). Меѓутоа, коефициентот на валидност за скоровите од пондерирањето на увереноста бил понизок од коефициентот на валидност за конвенционалните скорови (0,67 спрема 0,70), но разликата не била статистички значајна. Овој наод сугерира дека порастот во релијабилноста кој се јавува како резултат на примена на техниките на пондерирање на увереноста најверојатно потекнува од ирелевантни извори (како што е стилот на одговарање) кои не само што не ја зголемуваат валидноста, туку напротив, можат и да ја намалат.

Елиминационо тестирање

Елиминационото тестирање е друга постапка на задавање и бодување на објективните тестови која претендира да ги дискриминира нивоата на парцијалното знаење и да ја намали варијансата на грешка во тест—скоровите. Психолошкото образложение на оваа техника лежи во претпоставката дека парцијалното знаење се експонира во препознавање на некои од дистракторите (погрешните алтернативи). Комплетното знаење, знаењето на она што е точно, е еднакво на знаењето на сè што е погрешно. Парцијалното знаење пак, е знаење на дел од она што е погрешно, и е различно од погрешното знаење кое го карактеризира верување дека она што е точно е погрешно. Значи, испитаникот кој има комплетно знаење за даден ајтем може да ја разликува точната алтернатива од погрешните, испитаникот кој има парцијално знаење може да идентификува дел од погрешните алтернативи и кога не знае која алтернатива е точна, а испитаникот кој има погрешно знаење мисли дека од погрешните алтернативи е точна и дека точната алтернатива е погрешна.

Постапките на елиминационото тестирање се применуваат на објективни тестови кои се состојат од ајтеми со една точна алтернатива и повеќе дистрактори. Задавањето и бодувањето на ајтемите треба на испитаниците да им обезбеди место на скалата која се протега од погрешно знаење, преку неколку степени на парцијално знаење до комплетно знаење. Затоа, од испитаниците се бара да одговараат со означување на оние алтернативи кои можат да ги елиминираат како неточни (наместо со означување на точниот одговор). Со цел аа се контролира погодувањето, испитаникот се казнува ако го елиминира точниот одговор.

Овој метод првпат бил предложен од Coombs во 1953 год. Според него, од испитаниците кои го решаваат тестот треба да се бара да ги прецртаат сите алтернативи кои ги сметаат за неточни. Од испитаникот кој го знае точниот одговор се очекува да ги прецрта сите погрешни алтернативи. Испитаникот кој не го знае точниот одговор обично има доволно знаење да забележи дека една или повеќе од погрешните алтернативи се неточни и од него се бара да ги прецрта само оние алтернативи за кои верува дека не се точни без да погодува, затоа што постапката на бодување не оди во прилог на погрешно погодување.

Пондерите користени во постапката на бодување зависеле од бројот на погрешни алтернативи. Секоја точно прецртана погрешна алтернатива вредела 1 бод, но во случај и да

бил и точниот одговор елиминиран, се доделувала казна во вредност од 1—к (каде „к“ е вкупниот број на алтернативи во ајтемот). Како што може да се види од долуприкажаната табела, еден ајтем со четири понудени алтернативи може да обезбеди скала од седум бодови со опсег од — 3 до + 3:

алтернативи	обрасци на можни одговори**						
а*	х	xxx	xxx	х			
б		х	хх	х	х	хх	х
в		х	х х	х	х	х х	х
г		х	хх	х	х	хх	х
бодови	—3	—2	—1	0	+1	+2	+3

+ точна алтернатива

++ со х се означени прецртаните алтернативи

Coombs, Milholland и Womer (1956) спровеле истражување со цел да утврдат какво е влијанието на постапката на тестирање предложена од Coombs врз релијабилноста на тестот. Тие зададе три теста од типот повеќечлен избор (со четири алтернативи) на една група средношколци користејќи (а) конвенционална постапка, (б) постапка на елиминационо тестирање и (в) комбинација од двете постапки. Резултатите покажале дека елиминационото тестирање ја зголемило релијабилноста на тестовите и тоа просечно во висина која одговара на 20% од зголемување на должината на еден конвенционален тест од ист тип. Меѓутоа, зголемувањето на релијабилноста се покажало завсина од тежината на тестот: добиените коефициенти на релијабилност за потешките тестови биле повисоки. Понатаму, авторите утврдиле дека тестовите зададени во услови на елиминационото тестирање се дискриминативни најмалку онолку колку што биле кога се задаваат конвенционално под најоптимални услови.

Во 1962 година Archer извршил споредување на постапката на тестирање со тестови на слободен избор (каде испитаниците ги избираат сите алтернативи за кои сметаат дека се можеби точни, иако само една е навистина точна) и постапката на елиминационо тестирање со конвенционалната постапка на тестирање, користејќи ги релијабилноста и валидноста како критериуми. Со истражувањето било најдено дека и двете мерки добиени со постапките на тестирање на увереноста со значително порелијабилни од конвенционално добиените мерки (елиминационото тестирање се покажало изразито порелијабилно само за потешките ајтеми), но затоа пак, конвенционално добиените мерки кои повисоко корелирале со критериумот (по-

ложбата на ученикот во одделението) се покажале повалидни Исто така, авторот соопштил дека релијабилноста изразена во единици на време потребно за задавање, решавање и бодување на тестовите му дава предност на конвенционалниот метод.

Пробабилистичко тестирање

Постапките на пробабилистичкото тестирање се базираат на барањето испитаникот да го изрази своето знаење во вид на субјективна веројатност која ѝ ја припишува на секоја алтернатива која ја смета за точна. Пробабилистичко тестирање се применува кога од испитаниците се бара „на секоја од понудените алтернативи на еден ајтем од типот повеќечлен избор (кој може, но не мора да има еден точен одговор) да ѝ припаѓат пондер кој ќе ја означува преференцијата или степенот на увереност во нејзината точност“ (Rippey, 1968, стр. 211).

De Finetti (1965) го применил своето знаење за субјективната веројатност на некои проблеми од областа на објективното тестирање. Тој ги посматрал проблемите од гледна точка на испитаниците обидувајќи се да одговори на прашањето како индивидуата се однесува во ситуација кога треба да избере само еден од повеќе понудени одговори. Тој пошол од претпоставката дека испитаникот е заинтересиран да постигне висок скор и дека скорот се определува по правило кое тој го знае, а кое зависи од врската помеѓу неговиот избор и вистинскиот точен одговор. Во случај кога испитаникот е сигурен во точноста на вистинскиот точен одговор, за него не постои никаков проблем. Меѓутоа, многу често, особено кога се работи за многу тежок ајтем, испитаникот е до извесен степен несигурен во точноста на одговорот. Со цел да му овозможи на чесниот испитаник да одговара на начин кој е најповолен за него во ситуација кога се соочува со несигурност, de Finetti предложил да се користи теоријата на одлуки преку изразување на субјективната веројатност на испитаниците. Под субјективна веројатност тој подразбира „степен на увереност на индивидуата во точноста на алтернативите, во моментот кога се соочува со проблемот на одговарање на определен ајтем“ (de Finetti, 1965, стр. 88).

Основната филозофија на природот на de Finetti се базира врз неколку прелиминарни претпоставки: (а) испитаниците мораат добро да го познаваат и разбираат и начинот на одговарање и методот на бодување; (б) испитаниците мораат да бидат искрено заинтересирани за постигнување висок скор; (в) испитаниците треба добро да ја разбираат врската помеѓу сопствените верувања и нумеричките веројатности во кои веру-

вањата треба да се преведуваат; (г) целокупното знаење и верување на определен испитаник за ајтемот и понудените алтернативи мора да се изразува и целосно да се претставува преку нумеричките веројатности кои тој им ги припишува на алтернативите; (д) скоровите треба да се пресметуваат на начин кој ќе ги одразува парцијалните знаења на испитаниците; и (ѝ) процената на оваа постапка треба да се однесува на тоа колку добро методот на бодување ја одразува увереноста на испитаникот и колку можните предности ги надминуваат недостатоците на усложеноста и продолжноста на бодувањето на одговорите.

De Finttiј прикажал и продискутирал различни шеми на бодување, обидувајќи се да ги прилагоди на однесувањето на испитаниците во тест-ситуацијата. Прво, тој го проанализирал таканаречениот „континуиран метод“ (continuous method) кој се смета за најмоќен затоа што овозможува точно определување на сите веројатности. Кога се применува континуираниот метод, од испитаникот се бара да ги определи и означи субјективните веројатности поврзани за секоја од алтернативите при што мора да ги открие своите вистински верувања затоа што фалсификувањата не му одат во прилог. Според наједноставниот приод од овој вид, тест-скорот „S“ се пресметува по формулата:

$$S = 2 p_h - \sum_{j=1}^r p_j^2$$

каде „ p_j “ е веројатноста припишана на j -тата алтернатива на ајтемот кој се состои од r алтернативи чиј точен одговор е обележан со h .

Предложени се и неколку варијанти на наведената формула: $S' = (S + 1)/2$ и $S'' = (S - 1)/2$ кои меѓусебно се разликуваат само по опсегот на корисната скала на мерење: за „S“ опсегот се движи од -1 до $+1$ за „S'“ од 0 до $+1$, а за „S''“ од -1 до 0 . Во сите три случаи, минималната вредност се постигнува кога целокупната веројатност е концентрирана на погрешната алтернатива, а максималната кога целат веројатност ($p=1$) ѝ е припишана на точната алтернатива.

Подоцна, de Finetti увидел дека е доста тешко да се дистрибуираат точни веројатности на алтернативите и затоа предложил други алтернативни методи. Наместо да бара од субјектите да ги означат точните субјективни веројатности за секоја од алтернативите, тој применувал нови методи со кои ги проценувал субјективните веројатности на субјектите користејќи најразлични шеми на одговори. Новите методи барале од испитаниците да ги рангираат сите или дел од алтернативите спо-

ред нивната точност, да означат извесен број алтернативи кои ги сметаат за прифатливи, или да ги прецртаат другите како неприфатливи итн.

Понатаму, de Finetti ги развил и таканаречените „методи на точно најмало растојание“ (strict least-distance methods) како поедноставна верзија на првобитно предложениот континуиран метод. Овие методи се базираат на обрасци на бодови кои се конзистентни (претставуваат веројатности чија сума не надминува единица). Кога се применуваат овие методи, од испитаниците се бара на алтернативите да ги распоредат единиците на веројатност дадени со обрасците на бодови. Најзабележителен образец на бодови е таканаречениот „систем од пет ѕвездички“ кој се користи за одговарање на ајтеми со пет понудени алтернативи. Дозволените веројатности изнесуваат 0, 0,2, 0,4, 0,6, 0,8 и 1, при што секоја ѕвездичка означува единица на веројатност која вреди 0,2 бодови. Кога испитаникот одговара на ајтемот, тој е задолжен да ги распореди сите пет ѕвездички на алтернативите кои ги смета за веројатно точни на тој начин што ги користи како пондери кои ја изразуваат јачината на неговата увереност во точноста на алтернативите. Потполна увереност во точноста на една алтернатива се означува со доделување на сите пет ѕвездички на таа алтернатива, а потполна несигурност се изразува со доделување на по една ѕвездичка на сите пет алтернативи. Скорот на секој ајтем (ајтем-скорот) може да се движи во границите од 0 до 25, зависно од искажаната увереност во точноста на избраните алтернативи (како што е прикажано во дадената табела).

На пример, ако испитаникот поставил три ѕвездички на една алтернатива и две ѕвездички на друга, ајтем-скорот се определува врз база на типот 3—2. Ако трите ѕвездички ѝ се доделени на точната алтернатива, ајтем-скорот изнесува 21. Ако пак на точната алтернатива ѝ се доделени само две ѕвездички, ајтем-скорот изнесува 16. Во случај пак сите ѕвездички да се доделени на неточната алтернатива, ајтем-скорот изнесува само 6.

тип на одговор	дистрибуција на ѕвездички	соодветни веројатности	ајтем—скор
5	+++++	1,0 0,0	25 0
4—1	++++ +	0,8 0,2 0,0	24 9 4
3—2	+++ ++	0,6 0,4 0,0	21 16 6

3—1—1	+++	0,6	22
	+	0,2	12
	+	0,2	12
		0,0	7
2—2—1	++	0,4	18
	++	0,4	18
	+	0,2	13
		0,0	8
2—1—1—1	++	0,4	19
	+	0,2	14
	+	0,2	14
	+	0,2	14
		0,0	9
1—1—1—1—1	+	0,2	15
	+	0,2	15
	+	0,2	15
	+	0,2	15
	+	0,2	15
		0,0	10

Значаен придонес на проблемот на тестирање на увере-носта дале и Shuford, Albert и Massangill. Нивната студија од 1966 год. имала за цел да го изводи и да го користи поголемиот дел од расположивите информации од објективните тест-ајтеми. По нивно мислење, овие информации се содржат во веројатностите кои го изразуваат степенот на увереноста на субјектите во точноста на различните можни одговори. За да може успешно да се мерат овие субјективни веројатности, тие сметале дека е неопходно да се развие систем на бодување кој би гарантирал дека испитаникот, независно од неговото ниво на знаење би можел максимално да го зголеми очекуваниот скор ако и само ако чесно го одразува степенот на увереност во точноста на избраните одговори. Постапките на тестирање кои ги применуваат таквите системи на бодување на испитаници кои добро ги разбираат се наречени „прифатливи постапки на мерење на веројатноста“ (admissible probability measurement procedures). Според авторите, сите претходно користени постапки на мерење не биле прифатливи во смисла на тоа што не ги исполнувале барањата на поставената дефиниција.

Shuford и соработниците вовеле и поим за систем на бодувања со особина на репродуцирање. Системот на бодување е репродуцирачки кога субјективните веројатности на испитаникот се идентични со веројатностите кои тој им ги припишува на своите одговори. Авторите ги извеле потребните и доволните услови за обезбедување на репродуктивност на тест-ајтемите со две понудени алтернативи. Понатаму, тие покажале дека

репродуктирачките системи на бодување се практично неисцрпни и демонстрирале соодветен метод на нивно конструирање.

Shuford и соработниците развиле различни функции на кои се базира пресметувањето на скоровите на испитаниците кога се применуваат репродуктирачките системи на бодување на ајтеми со повеќе од две понудени алтернативи (а само еден точен одговор). Прво, тие предложиле примена на квадратна и сферична функција:

$$\text{квадратна: } S = 2 r_k - \sum_{j=1}^n r_j^2$$

$$\text{сферична: } S = r_k / \left(\sum_{j=1}^n r_j^2 \right)^{1/2}$$

каде што скорот на испитаникот („S“) се изразува преку веројатностите припишани на j-тата алтернатива (r_j) и веројатностите припишани на точната алтернатива „гк“.

Пресметувањето на скорот на испитаникот со примена на квадратната и сферичната функција има еден голем недостаток: скорот зависи од дистрибуцијата на вредностите на веројатностите доделени на дистракторите, како што е прикажано во долунаведениот пример:

понудени алтернативи:	субјект А	субјект Б	субјект В
А	0,00	0,15	0,15
Б-точна	0,60	0,60	0,60
В	0,40	0,20	0,10
Г	0,00	0,05	0,05
Д	0,00	0,00	0,10
примена на:	— износ на скорот		
— квадратна ф.	1,040	1,135	1,155
— сферична ф.	1,500	2,353	2,828

Во примерот се дадени дистрибуции на веројатности припишани на понудените алтернативи кои се разликуваат само по вредностите на веројатностите доделени на дистракторите. Може да се види дека во услови кога скорот не зависи од веројатностите доделени на точниот одговор (тие се во примерот константни), неговиот износ варира под влијание на дистрибуцијата на веројатностите припишани на неточните алтернативи: скорот е повисок кога на повеќе дистрактори им

се припишуваат веројатности. Заради овој недостаток, сферичната и квадратната функција се препорачуваат како прифатливи за бодување на ајтеми со две понудени алтернативи од точно-неточно, или за бодовни системи кои ги ограничуваат можните износи на веројатностите кои им се припишуваат на понудените алтернативи.

Shuford и соработниците ја разгледале и можноста да се конструира репродуктивен систем на бодување во кој скорот на субјектот би зависел само од веројатноста која тој ѝ ја припишал на точната алтернатива, а не и од веројатностите припишани на дистракторите. Како резултат на оваа барања особина, тие развиле логаритамска функција која подоцна ја ограничиле за да ги избегнат можните бесконечни предности (кога испитаникот не припишувал никаква веројатност на точниот одговор, неговиот скор изнесувал минус бесконечно). Ограничената логаритамска функција го има следниот облик:

$$S = \begin{cases} 2 + 10g r_k / 2 & \text{за } 1 \geq r_k > 0,01 \\ 0 & \text{за } 0 \leq r_k \leq 0,01 \end{cases}$$

каде „ r_k “ е веројатноста доделена на точната алтернатива.

Подоцна (1968) Shuford и Massengill развиле сет од помагала за тестирање на верноста кој се состоел од посебни листи за одговори, табели за бодување и механичко помагало за бележење на субјективните веројатности. Овој сет бил конструиран за да го олесни бодувањето и одбележувањето на увереноста во одговорите на секоја од понудените алтернативи на еден објективен тест кој се применува за тестирање во наставни цели. Применувајќи го вака подготвениот сет од помагала во неколку наврати, тие дошле до заклучок дека скорите добиени со ваквиот начин на тестирање се повалидни и порелијабилни од скорите кои се резултат на конвенционалниот начин на тестирање, затоа што се елиминира погодувањето и се зголемува количеството на инфомации кои ги дава тестот.

По извршената ревизија на пробаблистичкиот начин на тестирање предложен од Shuford и Massengill, Ebel (1968) се осврнал критички на сложеноста на постапката, проценувајќи дека времето потребно за примена на ваквата постапка е двојно подолго од времето потребно за примена на вообичаените постапки на тестирање. Тој заклучил дека предложената постапка прибавува повеќе информации за секој ајтем, но поставил прашање дали количеството и вредноста на дополнителните информации се доволни да го оправдаат порастот на трошоците и зголемувањето на времето потребно за примена и бодување на тестот.

Rippey (1968) експериментирал со три шеми на пробабилистичко бодување на тестот: (а) логаритамска, (б) сферична (како што ги предложил Sbuford со соработниците) и (в) евклидова. За разлика од логаритамската и сферичната кои се применуваат само за бодување на ајтеми кои имаат еден точен одговор, евклидовата е погодна и за бодување на ајтеми кои бараат повеќе одговори со различни степени на точност, кои одговараат на одредена дистрибуција на преференција определена врз база на согласност на група експерти. Скорот добиен со примена на евклидовата функција се пресметува по формулата:

$$S = 1 - \left(\sum_{i=1}^n (r_i - k_i)^2 \right)^{1/2} / \sqrt{2}$$

каде „ r_i “ е веројатноста припишана на i -тата алтернатива. „ k_i “ е критериумската веројатност определена од експертите за i -тиот одговор.

Непостојаноста на добиените коефициенти на релијабилност, покажала дека релијабилноста на тестот автоматски не се зголемува со примена на пробабилистичките постапки на тестирање. Најдените коефициенти на корелација помеѓу скоровите добиени со сите применети постапки и скоровите добиени на есејски тестови, укажале дека традиционалните скорови повисоко корелираат од пробабилистичките, навестувајќи дека ајтемот бодуван по пробабилистичките постапки навистина содржи поинаков вид информации од ајтемот бодуван традиционално (што најверојатно се должи на фактот дека пробабилистички бодуваните ајтеми содржат информации поврзани со увереноста на ученикот во своето знаење).

Во една подоцнежна студија, истиот автор (Rippey, 1970) применил неколку делови од еден стандардизиран тест на пишување на различни групи средношколци. Скоровите ги пресметувал користејќи пет различни бодовни системи: (а) веројатности припишани на точниот одговор ($S = r_k$), (б) логаритамска функција, (в) сферична функција, (г) евклидова функција и (д) „изведен избор“ (inferred choice) кој одговара на конвенционално бодување (каде што субјектот добива бод само ако покажал најголема увереност во точната алтернатива). Споредувањето на добиените коефициенти на релијабилност (проценти преку Хојговата постапка на анализа на варијанса) покажало дека наједноставниот систем на бодување кој ги користи веројатностите припишани на точниот одговор има највисока релијабилност (од околу 0,69 во просек) во споредба со сите други системи (0,50 за логаритамската, 0,49 за сферич-

ната, 0,58 за евклидовата и 0,47 за изведениот избор). Најдените податоци навестуваат дека во отсуство на информации за системот на бодување кој ќе се користи, испитаниците ја припишуваат својта увереност во точноста на понудените алтернативи врз база на барањата на наједноставниот модел, поради што и рејабилноста опаѓа со усложнување на системот на бодување кој се користи за пресметување на скоровите. Авторот дошол за заклучок дека најпожелно би било сите тестови составени од ајтеми со по еден точен одговор да се бодуваат користејќи ги веројатностите припишани само на точниот одговор, затоа што релијабилноста на така добиените скорови е повисока, затоа што не се вклучени посебните трошоци ниту посебните компликации во бодувањето, како и затоа што идејата да се бара од испитаникот да ги изнесе степените на увереноста воточноста на одговорите (наместо да се бара од него да го искаже своето мислење за несигурни перференции) изгледа по-реалистична и почесна и за испитаникот и за испитувачот.

Michael (1968) го проучувала влијанието на различни методи на бодување на ајтеми од типот повеќечлен избор врз релијабилноста на тест-скоровите и врз стандардната грешка на мерењето. Методот на субјективна веројатност кој таа го користела барал од субјектите да распоредат 10 бодови меѓу понудените алтернативи за секој ајтем. Ајтемите биле бодувани врз база на пропорциите на бодови припишани на точната алтернатива и скоровите добиени на овој начин биле споредувани со двете конвенционални методи на бодување: бодувањето по формула и бодувањето само на точниот одговор.

Добиените резултати укажале на повисока релијабилност и пониски стандардни грешки на мерењето при примената на постапките на субјективна веројатност, додека обете конвенционални методи на бодување се покажале скоро еднакви во поглед на релијабилноста, но бодувањето само на точните одговори предизвикало помали грешки во мерењето од бодувањето по формула. Авторот нашол дека тестот кој се бодува конвенционално земајќи ги во предвид само точните одговори, би требало да биде 1,7 пати подолг за да се постигне релијабилност која се добива со бодување на тестот по методот на увереност. Со други зборови, таа заклучила дека продолжувањето на времето потребно за тестирање за нови 18 минути (колку што било потребно за постапките на тестирање на увереноста) е теориски еквивалентно на додавањето на нови 25 ајтеми на тестот кој се состои од 35 ајтеми, за да се добие тест од 60 ајтеми кој бодуван конвенционално ќе достигне исто ниво на релијабилност.

Boldt (1971) вовел нов метод на тестирање кој ги комбинира постапките на пондерирање на увереноста со постапките

кои ги користат субјективните веројатности. Според овој метод, наречен „тестирање на увереноста со избирање на еден“ (Pick-One Confidence Testing), испитаникот прво ја избира онаа алтернатива од повеќе понудени за која најмногу верува дека е точна. Откако ја забележува таа алтернатива, тој означува на скала од пет поени колку е уверен во нејзината точност. Иако испитаникот не одговара користејќи ги субјективните веројатности во вистинска смисла, упатството за овој вид тестирање вклучува субјективни веројатности за секое можно ниво на увереност. Кога се применува ваквиот начин на тестирање, бодувањето се базира на означените субјективни веројатности и зависи од бројот на понудените алтернативи. Кога е изразена увереноста во точноста на точниот одговор, ајтем-скорот се движи во границите меѓу 0 и 1: кога е изразено најниско ниво на увереност скорот е 0, а кога е изразено највисоко ниво на увереност скорот е 1. За определување на бодовите се користени квадратни функции, една за случајот кога е избрана точната алтернатива, друга за случајот кога е избрана неточната. Табелата за определување на бодовите за ајтеми со две, три, четири и пет понудени алтернативи и за различни нивоа на веројатност припишани на избраната алтернатива, го има следниот облик:

веројатност припишана на избраната алтернатива	А Л Т Е Р Н А Т И В И							
	2		3		4		5	
	Т	Н	Т	Н	Т	Н	Т	Н
0,2							0	0
0,25					0	0	0,12	-0,04
0,3					0,03	-0,05	0,23	-0,08
0,33			0	0	0,21	-0,09	0,31	-0,11
0,35			0,04	-0,01	0,25	-0,11	0,34	-0,13
0,4			0,18	-0,11	0,36	-0,17	0,44	-0,19
0,45			0,32	-0,21	0,46	-0,25	0,53	-0,25
0,5	0	0	0,44	-0,31	0,56	-0,33	0,61	-0,33
0,55	0,19	-0,21	0,54	-0,43	0,64	-0,43	0,68	-0,41
0,6	0,36	-0,44	0,64	-0,56	0,72	-0,53	0,75	-0,50
0,65	0,51	-0,69	0,72	-0,70	0,78	-0,64	0,81	-0,60
0,7	0,64	-0,96	0,80	-0,85	0,84	-0,76	0,86	-0,70
0,75	0,75	-1,25	0,86	-1,02	0,89	-0,89	0,90	-0,82
0,8	0,84	-1,56	0,91	-1,19	0,93	-1,03	0,94	-0,94
0,85	0,91	-1,89	0,95	-1,38	0,96	-1,17	0,96	-1,07
0,9	0,96	-2,24	0,98	-1,57	0,98	-1,33	0,98	-1,20
0,95	0,99	-2,61	0,99	-1,78	1,00	-1,49	1,00	-1,35
1,00	1,00	-3,00	1,00	-2,00	1,00	-1,67	1,00	-1,50

Т — ако избраната алтернатива е точна

Н — ако избраната алтернатива не е точна

Главниот недостаток на овој начин на бодување е во тоа што применетата постапка не ги зема предвид реакциите на испитаниците на алтернативите кои не се избрани, поради што се губат информации за нивото на увереност во точноста на дистракторите (може само да се заклучи дека степенот на увереност во точноста на неизбраните алтернативи е помала или еднаква на веројатноста означена за избраната алтернатива).

Во една подоцнежна студија (1974) Boldt ја прилагодил логаритамската функција да ги третира испуштените алтернативи на ист начин како и алтернативите избрани со случајно погодување. Авторот пошол од претпоставката дека скорот на ајтем чиј одговор испитаникот не го знае, би требало да има иста вредност и во случај кога испитаникот одговара на ајтемот заокружувајќи неточен одговор и во случај кога едноставно пропушта да одговори на ајтемот. Затоа, при примена на постапките на тестирање на увереноста, секогаш кога испитаникот означува дека неговата неувереност е $1/k$ (каде „ k “ е бројот на понудени алтернативи) што се случува во ситуација кога дава неточен одговор, неговиот скор треба да е еднаков (и да изнесува нула) на скорот кој се добива како резултат на neodговарање на ајтемот. Така, според Boldt, логаритамската функција треба да го има следниот облик:

$$S = A (\log k + \log x)$$

каде скорот „ S “ се изразува преку бројот на понудени алтернативи „ k “, преку бројот на точно одговорени алтернативи „ x “ и преку константата $A = 10/\log k$ (која ја поставува горната граница на скорот на 10).

Најголемиот број студии кои се бавеле со анализирање на постапките на тестирање на увереноста вршеле проценка на овие постапки користејќи го само порастот во коефициентот на релијабилноста како критериум. Иако е тоа често пати најдобриот критериум кој е на располагање, релијабилноста сама по себе има мала вредност за опишување на ефикасноста на еден тест. За соодветна проценка на ефикасноста на постапките на тестирање на увереноста треба да се користат и мерките на валидноста како дополнителен критериум. Со таа цел, Koehler (1971) ги споредил конвергентата и дискриминативната валидност на две техники на пробаблистичкото тестирање со валидноста на конвенционалната постапка на тестирање. Тој ги применил континуираната метода на бележење на увереноста развиена од de Finetti ($S = 2r_k - \sum_j r_j^2$) и репродуктивниот систем на бодување предложен од Shuford со соработниците (логаритамската функција).

Резултатите од студијата покажуваат дека примената на постапките на тестирање на увереноста, општо земено, не ја подобрува релијабилноста на традиционалните објективни тестови. Споредувањето на коефициентите на конвергентна валидност (коефициентите на корелација помеѓу постапките на тестирање на увереноста и конвенционалните постапки на тестирање) не покажало значајни разлики меѓу нив. Анализата на коефициентите на дискриминантната валидност не дала задоволувачки резултати: корелациите помеѓу мерките добиени со мерење на различните варијабли со еден метод се покажале повисоки од корелациите помеѓу мерките добиени со мерење на една варијабла со различни методи. Од студијата на Koehler може да се заклучи дека со примената на постапките на субјективна веројатност не се постигнува подобрување на кострукт-валидноста на тестот.

Hakstian и Kansup (1975) спровеле истражување со цел опопфатно да ја испитаат релијабилноста (изразена и како интерна конзистентност и како стабилност) и истовремено да прибават податоци за критериум-валидноста на тест-постапките. Во истражувањето учествувале голем број ученици поделени во три групи: на едната ѝ бил зададен текст со конвенционално упатство за решавање (конвенционално тестирање), другата била упатена да распоредува 10 поени кои изразуваат увереност во точноста на алтернативите на секој ајтем (пробабилистичко тестирање), а третата група одговарала на тој начин што ги прецртувала неточните алтернативи (елиминационо тестирање). Скоровите на конвенционалната група биле претставени со сумите на точните одговори. Бодувањето на одговорите на групата на која било применето пробабилистичко тестирање, било извршено на пет различни начини. Испитаниците од кои се барала примена на постапките на елиминационо тестирање добивале по +1 бод за секоја точно прецртана алтернатива и -4 бода за секоја неточно прецртана.

Добиените резултати пружаат доказ дека тестирањето на увереноста доведува до највисока релијабилност изразена и како интерна конзистентност и како стабилност. Меѓутоа, студијата не докажала дека скоровите добиени со елиминационо тестирање се порелијабилни од скоровите добиени на конвенционален начин.

Авторите ги процениле и критериум-валидностите на скоровите добиени со сите три постапки користејќи како критериуми: (а) четири мерки на школско знаење и (б) слични мерки на вербални и математички способности. Студијата не нашла значајни разлики меѓу добиените коефициенти на валидност. Ниту едната, ниту другата експериментална постапка

на тестирање не предизвикала постојан пораст во валидноста во споредба со конвенционалното тестирање, туку напротив, во неколку случаи било забележано опаѓање.

Kansur и Hakstian (1975) ги употребиле истите податоци со цел да спроведат пет различни постапки на пробабилистичко бодување во поглед на релијабилноста и валидноста. Одговорите на испитаниците биле бодувани на следните начини:

- (1) $Y_{ij} = c_{ij}$ (опсег: од 0 до 10)
- (2) $Y_{ij} = \log c_{ij}$ (опсег: од 0 до 1)
($Y_{ij} = 0$ кога $c_{ij} = 0$)
- (3) $Y_{ij} = 1 - \log (10 - c_{ij})$ (опсег: од 0 до 1)
($Y_{ij} = 1$ кога $c_{ij} = 10$)
- (4) $Y_{ij} = e_{ij}^2/10$ (опсег: од 0 до 1)
- (5) $Y_{ij} = z_{ij}$ (каде $+ 1,69 \geq z \geq - 1,69$)

каде „ Y_{ij} “ е ајтем-скорот за субјектот i на ајтемот j ; „ c_{ij} “ е бројот на бодови кој ја изразува увереноста на субјектот i и во точноста на избраната алтернатива на ајтемот j ; „ z_{ij} “ е нормализиран стандарден скор за c_{ij} за дистрибуција од 11 бодови.

Резултатите од анализата покажале дека (3) и (4) постапката биле неоправдано фаворизирани во поглед на интерната конзистентност и тест-ретест стабилноста затоа што очекувањата не се исполниле: ниедна разлика помеѓу добиените коефициенти на релијабилност не достигнала ниво на значајност 0,05 ниту за скоровите од тестот на вербални способности, ниту за скоровите од тестот на математичко резонирање. За скоровите од тестот на вербални способности биле најдени незначителни разлики во добиените коефициенти на валидност за сите пет постапки на бодување. За критериумот школско постигнување се добила мала, но значајна разлика помеѓу (3) и (4) постапка, главно затоа што корелацијата помеѓу така добиените скорови била многу висока (0,997). Не биле најдени значајни разлики помеѓу коефициентите кои ја изразувале корелацијата меѓу скоровите добиени со примена на слични тестови на вербални способности и скоровите добиени со петте постапки на тестирање на увереноста. За скоровите од тестот на математичко резонирање се добиле дури и помали разлики во коефициентите на валидност, при што како значајна се појавила само разликата помеѓу (2) и (5) функција кога се корелирани со двата критериума.

Тестирањето на увереноста и личните особини на испитаниците

Експериментирајќи со постапките на тестирањето на увереноста, извесен број студии покажале дека постигнувањата на испитаниците на така задаваните и бодувани тестови не ги одразуваат само знаењата на испитаниците по однос на варијаблата која е предмет на мерењето, туку и дека се определени од извесни особини на личноста на испитаниците кои инаку не доаѓаат до израз при традиционалното тестирање. Таквите фактори во личноста кои произлегуваат од општата самоувереност на испитаниците ја намалуваат вредност на добиените скорови и ја зголемуваат во мерењето на така применетите тестови како мерни инструменти.

Уште во 1936 година Wiley и Trimble спровеле истражување со цел да го утврдат уделот на факторите на личноста во постапките на тестирање на увереноста кои биле предложени од Nevner (1923). Тие задале четири различни тестови барајќи од испитаниците да означат (а) дали се сигурни во точноста на избраната алтернатива, (б) дали алтернатива ја избираат со неизвесност или (в) само погодуваат дека конкретната избрана алтернатива е точна. Иако на испитаниците им било кажано дека нивните одговори ќе бидат вреднувани според искажаниот степен на увереност во нивната точност, одговорите всушност не биле различно бодувани — скоровите биле пресметувани по вообичаената формула за ајтеми од тепот за точно-неточно: сумата на точните одговори минус сумата на неточните. Зборовите на најтеми чии одговори биле означени со „сигурно“, „неизвесно“ и „подоудвање“, биле користени како основа за пресметување на корелации помеѓу сите четири тестови. Аритметичките средини за добиените коефициенти на корелација изнесувале: 0,393 за скоровите пресметани како сума од точните минус сума од неточните одговори, 0,656 за скоровите пресметани како збир од ајтемите во чија точност испитаниците биле сигурни, 0,576 за скоровите пресметани како збир од ајтемите во чија точност испитаниците не биле сигурни (искажале изнесуваат) и 0,566 за скоровите пресметани како збир од ајтемите кои испитаниците ги избрале со погодување. Овие вредности покажуваат дека применетите тестови ја мереле и искажаната увереност на субјектите во точноста на нивните одговори поконзистентно (подоследно) отколку што го мереле знаењето од испитуваната област. Со други зборови, авторите заклучиле дека резултатите добиени од примената на постапките на тестирање на увереноста се под влијание на фактори на личноста кои можат да бидат мерени од објективен начин.

Поттикнувата од оваа студија Swineford (1938) спровела своја, со цел да најде мерка на оние особини на личноста кои го определуваат изборот на одговорите на испитаниците. Таа зададе тест кој се состоел од ајтеми од типот точно- неточно со истите упатства како Soderquist (1936) барајќи од испитаниците да ја означат својата увереност во точноста на избраните алтернативи преку вложување на 1, 2, 3 или 4 бодови. Како резултат на применетото бодување се добиле скорови кои се изразувале како разлика помеѓу пондерираните точни и пондерираните неточни одговори чии пондери се определиле врз основа на бодовите кои испитаниците ги вложиле во секој даден одговор (при што пондерите за неточните одговори биле двојно поголеми од износот на вложените бодови).

Врз основа на направената анализа на добиените резултати Swineford издвоила посебна особина на личноста која деги извршувале домашните обврски ќе бидат толку погрешно индекса ризикува доколку вложува бодови во овие ајтеми чии одговори не ги знае. Иако било невозможно да се определи кои од точните одговори биле резултат на погодување, се претпоставило дека сите неточни одговори биле давани со случајно погодување. Понатаму, се сметало дека е крајно неверојатно дека испитаниците кои биле редовни на настава и кои успешно ги извршувале домашните обврски ќе бидат толку погрешно информирани за да прават грешки комплетно уверени во точноста на дадениот одговор, па затоа се претпоставило дека испитаниците кои вложиле 4 бодови во погрешни одговори се труделе да го зголемат својот скор иако казните биле доста високи. Со други зборови, испитаниците ризикувале на оние ајтеми чии одговори не ги знаеле. Поаѓајќи од овие претпоставки, Swineford дефинирала таканаречен „скор на ризикување“ (gambling score) како процент од погрешителните одговори во кои се вложени 4 бодови:

$$\text{Скор на ризикување} = G = \frac{\text{број на погрешни одговори означени со „4“}}{\text{број на } 1/2 \text{ од бројот на погреш. одгов.} + \text{неодговорени ајтеми}} \times 100$$

Врз база на високата вредност на добиениот коефициент на релијабилност за G-скоровите (0,798) и безначајниот износ на коефициентот на колерација помеѓу G-скоровите и непондерираните скорови на испитаниците (0,084 \mp 0,05), авторот заклучил дека со формулата за пресметување на скорот на ризи-

кување се добива релијабилна мерка на тенденцијата за ризикување како особина на личноста која е независна од знаење то покажано на применетиот тест.

Во една подоцнежна студија (1941) Swineford применила четири теста за да ја измери тенденцијата за ризикување Студијата не само што го потврдила наодот дека G-скоровите се независни од тест-скоровите кои се добиваат на тестовите на знаење, туку довела и до нови сознанија. Врз основа на добиените резултати било заклучено дека момчињата многу почесто ризикуваат од девојчињата и тоа за 12% повеќе кога се применуваат вербатни тестови и за 21% повеќе кога се користат невербални тестови. Исто така, било најдено дека различните типови на тестовни материјали предизвикуваат значајно различни реакции: тенденцијата за ризикување била помала кога се применувале вербални тестови кои се составени од ајтеми поблиски за испитаниците, а многу поголема кога се користеле сосема нови тестови (порастот во G-скоровите изнесувал до 27% за момчињата и до 18% за девојчињата).

Проучувајќи го придонесот на психолошките фактори во тестирањето на увереноста, Hansen (1971) спровел истражување кое имало за цел да ја открие врската помеѓу определени особини на личноста и степенот до кој испитаниците исполнуваат увереност во точноста на своите одговори. Тој нашол дека испитаниците одговараат на ајтемите од типот повеќечлен избор со стабилна и карактеристична увереност која не може да биде припишана на нивното знаење. Оваа тенденција да се биде уверен или неуверен се покажало поврзана со скоровите на еден прашалник кој мери преференција за ризикување (Kogan and Wallace Choice Dilemmas Questionnaire), укажувајќи дека испитаниците со преференција за преземање ризик изразуваат поголема увереност во точноста на своите одговори, отколку што би се очекувало од индивидуи со нивно знаење.

Исто така, било најдено дека мерките добиени на една модификувана варијанта на F-скала (која мери авторитарност) корелираат значајно и негативно со увереноста на испитаниците, но дека скоровите добиени на два теста кои мереле анксиозност и ридигност не биле поврзани со тенденцијата да се биде уверен или неуверен.

Jacobs (1971) презел истражувања за да определи кои фактори на личноста се поврзани со изразувањето на увереност, како и да ги открие ефектите од примена на две нивоа казни врз неоправданото изразување на увереност. Тој задал тест од типот повеќечлен избор на машки и женски студенти да го решаваат по постапка на пондерирање на увереноста. Од студентите се барало да изберат која алтернатива ја сметаат за најточен одговор на ајтемот и да го изразат степенот на увереност во точ-

носта на извршениот избор на скала од три бода (погоднување, прилична увереност и потполна увереност) за што биле соодветно наградувани и казнувани. Еден бод им се доделувал за точен одговор означен како погоднување, два бода за изразување прилична увереност и три бода за означување потполна увереност. Ако избраната алтернатива била погрешна, студентите од групата на поблаго казнувани губеле по 0, 2 или 3 бода, додека студентите од групата на посторого казнувани губеле по (0, 4) или 6 бода зависно од изборната категорија на увереност. Мерката на изразување на увереноста (confidence expression) користена во оваа студија се базирала на формулата на Swineford (1968) и била дефинирана како:

$$\text{CONF} = \frac{\text{број на погрешни одговори за кои е изразена максимална увереност}}{\text{вкупен број на погрешни одговори}} \times 100$$

Двонасочната анализа на варијанса направена на скоровите добиени за машките и женските субјекти на кои се применети двете нивоа на казнување, не покажала значајни разлики. Коефициентите на релијабилност за конвенционално добиените тест-скорови, за пондерираните тест-скорови и за CONF-скоровите за поблаго казнување група изнесувале 0,89, 0,87 и 0,86. Соодветните коефициенти на релијабилност за посторого казнуваната група изнесувале 0,79, 0,39 и 0,68. Вака добиените резултати укажуваат на голема разлика во интерната конзистентност на пондерираните тест-скорови определени за двете групи субјекти. Падот на коефициентот на релијабилност од 0,87 на 0,39 покажува дека со зголемување на износот на бодовите кои се одземаат за неточно искажана увереност, се зголемува и непостојаноста на скоровите добиени со примена на постапката на пондерирање на увереноста. За да го разјасни овој наод, авторот ги пресметал и коефициентите на корелација помеѓу конвенционалните и пондерираните скорови за двете групи посебно и нашол дека корелацијата за поблаго казнуваната група изнесува 0,88, а за посторого казнуваната само 0,095. Врз основа на тоа, Jacobs заклучил дека забележаните промени во пресметаните коефициенти на релијабилност најверојатно се последица на ниска релијабилност на тест-скоровите која дозволува вмешување на независни и неконтролирани извори на варијанса кои предизвикуваат непостојаност и непоузданост на добиените тест-скорови.

Echternacht, Boldt и Sellman (1972) спровеле истражување обидувајќи се да ја откријат врската помеѓу некои особини на личноста и тест-скоровите добиени со примена на тестирање на

увереноста. Тие примениле батерија на инвентари на личност за мерење на особини како што се: догматизам, авторитатизам, анксиозност, ригидност, импулсивност и самоувереност. Тест-скоровите биле добивани на два теста на знаење кои се состоеле од ајтеми од типот повеќеочлен избор и кои биле применувани на искусни испитаници на два различни начина користејќи: (а) постапка на пондерирање на увереноста со избирање на еден точен одговор (каде од испитаникот се барало прво да ја избере точната алтернатива, а потоа да ја означи увереноста во нејзината точност на скала од пет бода) и (б) метод на дистрибуирање на 100 бода на предложените алтернативи на начин кој ќе го одрази степенот на увереност во точноста на секоја од алтернативите.

Авторите ги пресметале коефициентите на корелација помеѓу мерките добиени на инвентарите на личност и тест-скоровите пресметани по постапката на пондерирање на увереноста. Иако дел од коефициентите на корелација се покажале значајни, тие не останале постојани при повторувањето. Различните мерки на особините на личноста кои се покажале значајно поврзани со скоровите добиени на едниот тест на знаење, не биле значајно поврзани и со скоровите добиени со примена на истата постапка на бодување на одговорите дадени на другиот тест на знаење. Врз основа на тоа, Echternaht со соработниците заклучиле дека не може да се зборува за влијание на некои нови особини на личноста врз постапките на пондерирање на увереноста.

Заклучок

Од сето досега изнесено, може да се заклучи, дека постапките на тестирање на увереноста во потполност ѝ ги исполниле очекувањата во врска со подобрување на релијабилноста и валидноста на тест-скоровите.

Најдениот пораст во релијабилноста кој е резултат на примената на техниките на тестирање на увереноста, не е потполно оправдан, затоа што е проследен со продолжено време потребно за задавање и бодување на тестовите. Постапките на тестирање на увереноста главно бараат околу 25—35% подолго време од традиционалните постапки на тестирање, поради што ја доведуваат во прашање економичноста на тестирањето. Сосема е веројатно дека зголемувањето на релијабилноста кое му се припишува на тестирањето на увереноста може да се постигне и со традиционалното тестирање, ако на тестот кој се задава и бодува традиционално му се додадат нови ајтеми за чие решавање и оценување е потребно онолку време колку што изнесува продолженото време за примена на тестирањето на увереноста.

Валидноста на скоровите прсметани по постапките на тестирање на увереноста се доведува во прашање поради придонесот на индивидуалните разлики во личноста на испитаниците за зголемување на уделот на варијансата на грешката варијанса на тест-скоровите. Зголемувањето на варијансата на грешка за сметка на вистинската варијанса (варијансата на точноста на мерењето) потекнува од фактот што постојат определени особини на личноста на испитаниците кои неоправдано придонесуваат за зголемување или намалување на нивните постигнувања на тестовите задавани и бодувани по постапките на тестирање на увереноста, особини кои инаку не се опфатени со постигнувањето дефинирано со традиционалните формули на бодување.

Земајќи ги во предвид слабостите на постапките на тестирањето на увереноста, Koehler (1971) препорачал тестирањето на увереноста да се користи како наставно средство затоа што „може да му помогне на наставникот преку посматрање на означените уверености во точноста на одговорите избрани од страна на неговите ученици, полесно да ги открие јаките и слабите страни во нивното знаење од тестираната област“ (стр. 302).

Литература

- Archer, N. S. A COMPARISON OF THE CONVENTIONAL AND TWO MODIFIED PROCEDURES FOR RESPONDING TO MULTIPLE-CHOICE ITEMS WITH RESPECT TO TEST RELIABILITY, VALIDITY AND ITEM CHARACTERISTICS. Докторска дисертација, Syracuse University, 1962.
- Boldt, R. F. A SIMPLE CONFIDENCE TESTING FORMAT. Research Bulletin, Vol. 71—42, Princeton, N. J.: Educational Testing Service, 1971.
- Boldt, R. F. AN APPROXIMATELY REPRODUCING SCORING SCHEME THAT ALLIGNS RANDOM RESPONSES AND OMISSIONS. Educational and Psychological Measurement, 1974, 34, 57—61.
- De Finetti, B. METHODS FOR DISCRIMINATING LEVELS OF PARTIAL KNOWLEDGE CONCERNING A TEST ITEM. British Journal of Mathematical and Statistical Psychology, 1965, 18, 87—123.
- Dressel, P. L. i Schmid, J. SOME MODIFICATIONS OF THE MULTIPLE-CHOICE ITEMS. Educational and Psychological Measurement, 1953, 13, 574—595.
- Ebel, R. L. CONFIDENCE WEIGHTING AND TEST RELIABILITY. Journal of Educational Measurement, 1965, 2, 49—57. (a)
- Ebel, R. L. MEASURING EDUCATIONAL ACHIEVEMENT. Englewood Cliffs, N. J. Prentice-Hall, 1965 (b).
- Ebel, R. L. VALID CONFIDENCE TESTING — DEMONSTRATION KIT. Journal of Educational Measurement, 1968, 5, 353—354.
- Echternacht, G. J. THE USE OF CONFIDENCE TESTING IN OBJECTIVE TESTS. Review of Educational Research, 1972, 42, 2176—2236.
- Echternacht, G. J., Boldt, R. F. i Sellman, W. S. PERSONALITY INFLUENCES ON CONFIDENCE TEST SCORES. Journal of Educational Measurement, 1972, 9, 235—241.
- Jacobs, S. S. CORRELATES OF UNWARRANTED CONFIDENCE IN RESPONSES TO OBJECTIVE TEST ITEMS. Journal of Educational Measurement, 1971, 8, 15—20.
- Kansup, W. i Hakstian, A. R. A COMPARISON OF SEVERAL METHODS OF ASSESSING PARTIAL KNOWLEDGE IN MULTIPLE-CHOICE TESTS: I. SCORING PROCEDURES. Journal of Educational Measurement, 1975, 12, 219—229.
- Koehker, R. A. A COMPARISON OF THE VALIDITIES OF CONVENTIONAL CHOICE TESTING AND VARIOUS CONFIDENCE MARKING PROCEDURES. Journal of Educational Measurement, 1971, 8, 297—303.
- Michael, J. J. THE RELIABILITY OF A MULTIPLE-CHOICE EXAMINATION UNDER VARIOUS TEST-TAKING INSTRUCTIONS. Journal of Educational Measurement, 1968, 5, 307—314.
- Petroska, V. PROBABILITY SCORING: CONFIDENCE TESTING AND DIFFERENTIAL WEIGHTING, Teachers College, Columbia University 1982. (магистерска теза).
- Riphey, R. PROBABILISTIC TESTING, Journal of Educational Measurement 1968, 5, 211—215.
- Riphey, R. A COMPARISON OF FIVE DIFFERENT SCORING FUNCTIONS FOR CONFIDENCE TESTS. Journal of Educational Measurement, 1970, 7, 165—170.
- Shuford, E. H., Albert, A. i Massengill, H. E. ADMISSIBLE PROBABILITY MEASUREMENT PROCEDURES, Psychometrica, 1966, 31, 125—145.
- Shuford, E.H. i Massengill, H.E. CONFIDENCE TESTING TA THE OFFICER TRAINING SCHOOL, Lexington, Mass: Shuford-Massengill Corpor, 1968.

- Soderquist, H.O. A NEW METHOD OF WEIGHTING SCORES IN A TRUE-FALSE TEST. *Journal of Educational Research*, 1936, 30, 290—292.
- Swineford, F. MEASUREMENT OF A PERSONALITY TRAIT. *Journal of Educational Psychology*, 1938, 29, 295—300.
- Swineford, F. ANALYSIS OF A PERSONALITY TRAIT. *Journal of Educational Psychology*, 1941, 32, 438—444.
- Hakstian, A. R. i Kansup, W. A COMPARISON OF SEVERAL METHODS OF ASSESSING PARTIAL KNOWLEDGE IN A MULTIPLE-CHOICE TESTS: II. TESTING PROCEDURES. *Journal of Educational Measurement*, 1975, 12, 231—239.
- Hansen, R. THE INFLUENCE OF VARIABLES OTHER THAN KNOWLEDGE ON PROBABILISTIC TESTS. *Journal of Educational Measurement*, 1971, 8, 9—13.
- Hevner, K.A. A METHOD OF CORRECTING FOR GUESSING IN TRUE-FALSE TESTS AND EMPIRICAL EVIDENCE IN SUPPORT OF IT. *Journal of Social Psychology*, 1932, 3, 359—362.
- Hopkins, K. D., Hakstian, A.R. i Hopkins, B.R. VALIDITY AND RELIABILITY CONSEQUENCES OF CONFIDENCE WEIGHTING. *Educational and Psychological Measurement*, 1973, 33, 135—141.
- Coombs, C.H. ON THE USE OF OBJECTIVE EXAMINATION. *Educational and Psychological Measurement*, 1953, 13, 308—311.
- Coombs, C.H., Milholland, J.E. i Womer, F.B. THE ASSESSMENT OF PARTIAL KNOWLEDGE. *Educational and Psychological Measurement*, 1956, 16, 13—37.
- Wiley, L.N. i Trimble, O.C. THE ORDINARY OBJECTIVE TEST AS A POSSIBLE CRITERION OF CERTAIN PERSONALITY TRAITS — *School and Society*, 1936, 43, 446—448.

Viолета PETROSKA — BEŠKA

CONFIDENCE TESTING WITH MULTIPLE-CHOICE TESTS

S u m m a r y

This study is a theoretical survey of the confidence testing procedures which emerged from the tendency to improve conventional testing techniques through assessing partial knowledge and reducing the effects of guessing in multiple-choice tests.

Confidence testing is presented as a special method of responding to multiple-choice test-items and a special method of scoring the obtained responses. It is said to be a method of testing where examinees indicate their confidence in the correctness of the response alternatives of test-items on the basis of what the item-scores are determined.

All available techniques of confidence weighting, probabilistic testing and elimination testing as three different approaches to confidence testing are described separately, compared with the conventional testing techniques, and evaluated in the light of test reliability and validity. In addition to that, the study illustrates the effects of personality factors on confidence testing procedures as found in the research done so far.