

The Visual Analogue Scale. An Alternative to the Likert Type Response Scales. An Alternative to the Likert Type Response Scales

Adrian DUȘA*

University of Bucharest

Valeriu FRUNZARU

National School of Political and Administrative Sciences

Abstract: For about eight decades, research instruments in the social sciences have been orbiting around Likert's proposal for his famous response scale. Before him, and also after he managed to impose it, many researchers have tried to find a better solution. This, however, has proven difficult. While solving methodological problems for measuring concepts, by concentrating all the responses in only five categories brings major disadvantages as well: it has extremely low variation, it does not produce metric scores unless combined with similar items, and it cannot be used as such for advanced statistical analysis. In this article, we propose using a continuous response scale as a solution to each of these problems. In our opinion, the possible application of this solution has an extremely high potential to advance social science research methodology.

Keywords: *Research methodology, response scales, internal consistency, Likert Scale*

Scala de răspuns tip Likert. Între eficiență practică și limite teoretice

În colaborare cu Gardner Murphy, Rensis Likert a propus un instrument pentru măsurarea atitudinilor cu privire la relațiile internaționale, relațiile rasiale, conflictul economic, conflictul politic și religie. În efortul de a da o definiție 'atitudinilor', Likert subliniază faptul că acestea nu sunt

elemente inflexibile sau rigide ale personalității, ci sunt mai degrabă 'un anumit șir pe care răspunsurile se pot muta' (Likert, 1932, 8).

Doi ani mai devreme, o poziție similară a avut-o și Louis L. Thurstone: 'O atitudine este reprezentată ca un punct pe un continuum atitudinal. Există un număr infinit de atitudini ce pot fi reprezentate de-a lungul unei scale de atitudine. În practică noi nu facem

*email: corresponding author Adrian Dușa dusa.adrian@unibuc.ro. [Scala analog vizuală. O alternativă la scala de răspuns tip Likert]. Autorii sunt recunoscători D&D/TestCentral/OS România (<http://www.testcentral.ro>), pentru furnizarea cu generozitate a testului TOM – Testul de Orientare Motivațională, esențial pentru desfășurarea experimentului. Mulțumiri speciale, de asemenea, conf. C. Rughiniș și prof. D. Sandu care ne-au oferit observații extensive la redactarea textului.

diferențieri atât de fine. O atitudine, practic vorbind, este un anumit rang determinat sau o anumită vecinătate pe scală.’ (Chelcea, 2004, 355).

Scala de răspuns ierarhică, cu cele cinci categorii de răspuns în varianta ei clasică (1. dezacord puternic, 2. dezacord, 3. indecis, 4. acord, 5. acord puternic), reprezintă o astfel de soluție practică pentru măsurarea atitudinilor. Acest tip de scală de răspuns a existat și înainte de Likert, însă ca urmare a articolului pe care acesta l-a scris în 1932 scala respectivă a rămas în literatura de specialitate ca ‘scală tip Likert’ (în engl. *Likert-type scale*). Asocierea pentru cele cinci variante de răspuns a unor cifre de la 1 la 5 a dus la transformarea artificială a unei scale ordinale într-o scală de răspuns măsurată la nivel interval, în care distanțele dintre variantele de răspuns sunt tratate ca și când ar fi egale. De exemplu, distanța dintre ‘acord’ și ‘acord puternic’ ar fi la fel de mare ca distanța dintre ‘indecis’ și ‘dezacord’. Iar o persoană care a optat ‘acord’ (căreia îi corespunde cifra 4) ar avea un scor de două ori mai mare decât o persoană care a optat ‘dezacord’ (căreia îi corespunde cifra 2). În opinia noastră, această exagerare metodologică a fost și mai mult agravată de faptul că această scală ‘metrică’ este utilizată ca și când ar fi o variabilă discretă; cu alte cuvinte, respondentul sare brusc de la ‘dezacord’ la ‘acord’ (via ‘indecis’) ca și când între aceste categorii nu ar mai exista nici o altă variantă de răspuns.

În fapt, atitudinea este o dimensiune continuă. La fel cum este fericirea, la fel cum este mulțumirea, ori satisfacția. Nici un respondent nu poate să spună cu exactitate cât de mulțumit este într-un moment al vieții, așadar mulțumirea

nu poate fi încadrată între niște limite clar definite.

Să luăm spre exemplu iubirea: sunt frecvente, printre îndrăgostiți, declarațiile de genul ‘Te iubesc până la cer și înapoi!’. Cu alte cuvinte, iubirea este atât de mare încât respondenților le este greu să-i definească o limită maximă; practic, este indefinită. Din acest motiv, este extrem de dificil să restrângem o asemenea întindere în clasicele 5 cifre ale lui Likert, întrucât gradul maxim al scalei de răspuns este ‘foarte mult’. În realitate iubirea, sau mulțumirea, se întinde mult mai departe decât atributul ‘foarte’.

Referindu-se la atitudini, însuși Likert confirmă caracterul continuu al conceptelor:

... it seems justifiable for experimental purposes to assume that attitudes are distributed fairly normally, and to use this assumption for combining the different statements. (Likert, 1932, 22)

To avoid any space error or any tendency to a stereotype response it seems desirable to have the different statements so worded that about one-half of them have one end of the attitude continuum corresponding to the left or upper part of the reaction alternatives and the other half have the same end of the attitude continuum corresponding to the right or lower part of the reaction alternatives. (Likert, 1932, 46)

Dat fiind faptul că atitudinea este o dimensiune continuă (mai mult decât atât, chiar normal distribuită), în principiu cele cinci cifre ar trebui să fie poziționate la o distanță egală unele față de celelalte. În practică însă, respondenții aleg una dintre cele cinci

categoriilor de răspunsuri fără să perceapă distanța dintre acestea. Respondenții sunt pur și simplu fericiți sau nefericiți, ori foarte fericiți. În loc de atributul ‘foarte’ am putea să utilizăm fără probleme atributul ‘extrem de’, pentru a genera răspunsul ‘extrem de fericit’, iar respondenții nu ar sesiza nici un fel de problemă, atâta timp cât cele cinci răspunsuri sunt diferite între ele și oferă o anumită ordonare de la un pol negativ la un pol pozitiv. În consecință, cele cinci cifre nu pot avea semnificație de numere ci diferențiază pur și simplu cinci categorii ordonate. În loc de cifrele 1, 2, 3, 4, 5 am putea să utilizăm fără probleme literele a, b, c, d, e iar respondenții nu ar avea nici un fel de dificultate în a alege unul sau altul dintre răspunsuri.

Faptul că există o anumită tendință tacit acceptată în analiza datelor, de a calcula măsuri ale tendinței centrale (de exemplu, media) pe baza cifrelor de la 1 la 5 este, în opinia noastră, o exagerare metodologică generatoare de erori atât în rezultatele analizelor, cât și în interpretarea acestora. Din motive practice însă, în ciuda criticilor pe durata celor aproape opt decenii de la elaborarea scalei de răspuns, aceasta a fost unul dintre cele mai utilizate instrumente de către psihologi sau sociologi în măsurarea atitudinilor, opiniilor, satisfacției etc¹.

Măsurarea pe un continuum a atitudinilor a fost reprezentată grafic încă din anul 1920 la Scott and Co. Laboratory. Originalitatea ei consta în folosirea atât a unei linii, cât și a unor propoziții descriptive (patru sau cinci) care indicau variații ale personalității de la o extremă la alta (Freyd, 1923, 92). În instrucțiunile de folosire a scalei construite de Freyd, li se cerea

subiecților să bifeze un punct pe linie corespunzător situației lor, nu neapărat deasupra propozițiilor descriptive. Linia era împărțită apoi în 20 de intervale cărora le corespundea câte un număr, astfel fiecărei bifări care se plasa într-un interval îi corespundea un număr. Dacă subiectul bifa mai mult de o dată pe scală (posibil din cauza îndoielilor), atunci se făcea media scorurilor. Printre avantajele acestei scale se numără ușurința de a răspunde la întrebări, ușurința scorării și discriminarea fină a răspunsurilor (Freyd, 1923, 94). În ciuda acestor calități, scala grafică nu a devenit populară, așa cum se aștepta Freyd, din cauza efortului mare (mai ales dacă sunt sute sau chiar mii de subiecți) în obținerea scorurilor. Astfel, din motive practice, a fost folosită scala de tip Likert și nu o scală de răspuns continuă care ar fi corespuns faptului că atitudinile sunt ‘puncte pe un continuum atitudinal’.

Scala de răspuns tip Likert a cunoscut diverse variante în funcție de prezența sau absența expresiilor corespunzătoare punctelor de scală, respectiv de numărul de puncte de scală. Cu privire la primul aspect, dincolo de varianta clasică a scalei tip Likert, subiecților li se poate cere să-și exprime gradul de acord pentru o anumită afirmație alegând un număr pe o scală de la 1 la 5, unde 1 reprezintă ‘dezacord puternic’, iar 5 ‘acord puternic’. Acest tip de scală numerică prezintă avantajul că distanța dintre punctele scalei sunt egale și se pot construi scale cu un număr mai mare de categorii, fapt ce permite o mai fină discriminare a răspunsurilor. Dezavantajul este că odată cu creșterea numărului de puncte de scală, subiectul

este supus dificultății de a face diferența între două puncte de scală foarte apropiate (Kent, 2007, 136).

De exemplu, pe o scală de la 1 la 20, este greu de spus în ce măsură punctele de scală 16 și 17 reprezintă diferențe ale gradului de acord cu privire la o afirmație sau gradul de satisfacție față de un anumit produs sau serviciu. Iar pentru o scală cu peste cinci categorii de răspuns, este greu de găsit expresiile potrivite punctelor de scală.

Problema numărului de trepte pe scală a fost abordată în numeroase cercetări care au comparat fidelitatea măsurării în funcție de mărimea scalei și numărul de variante de răspuns. Cook et al. (2001, 698) atrag atenția asupra faptului că mai multe alternative de răspuns nu vor duce inevitabil la o variație mai mare, dar permit un mai mare scor numeric al acesteia. De exemplu, în situația teoretică în care avem șase subiecți care trebuie să aleagă dihotomic în cazul a patru itemi, gradul de eterogenitate ar fi maxim când trei subiecți ar răspunde 'Da' (codat cu cifra 1) la toți cei patru itemi, iar ceilalți trei ar răspunde 'Nu' (codat cu cifra 0) la toți cei patru itemi. În această situație, abaterea standard ar fi egală cu $2,2^2$. Dacă vom crește numărul variantelor de răspuns la trei (codate cu minim 0 și maxim 2), în situația de eterogenitate maximă abaterea standard ar fi egală cu 4,4 iar dacă numărul variantelor de răspuns ar fi patru (codate cu minim 0 și maxim 3), abaterea standard ar fi egală cu 6,6. Cook et al. concluzionează că mai multe puncte de scală permit teoretic o mai mare fidelitate a măsurării, însă doar pentru situațiile în care caracteristica măsurată are un grad ridicat de eterogenitate. La o concluzie

asemănătoare au ajuns și Komorita și Graham (1965), care indicau în plus că, dacă itemii măsoară o caracteristică omogenă, atunci nu există nici un fel de câștig în mărirea numărului de variante de răspuns. În aceste condiții, două alternative polare ar fi suficiente.

În cercetările realizate, experimentale sau simulate, au fost comparate scale de răspuns tip Likert cu un număr diferit de puncte de scală, luându-se în considerare coeficientul Cronbach-Alfa, alungirea și aplatizarea, sau după rescalare media și abaterea standard. Unele dintre rezultatele cercetărilor se contrazic. În urma unei cercetări cu privire la produsele alimentare, Bending a concluzionat că nu există diferențe în privința fidelității scalelor cu 3, 5, 7, 9 și 11 categorii (Krieg, 1999, 750). Komorita și Graham (1965) au folosit scale cu 2 respectiv 6 puncte, variind numărul de itemi și omogenitatea itemilor. Concluziile au fost că fidelitatea crește odată cu creșterea numărului de itemi, însă numai în cazul itemilor cu omogenitate scăzută sau ridicată, nu și pentru cei cu omogenitate medie. Iar creșterea numărului de variante de răspuns de la 2 la 6 duce la creșterea fidelității măsurării numai pentru itemii eterogeni, nu și pentru cei omogeni. Autorii compară creșterea preciziei măsurării folosind o scală cu mai multe puncte cu măsurarea înălțimii în inch, și nu în picioare sau yarzi.

Pentru a răspunde la întrebarea dacă există un număr optim de variante de răspuns, Matell și Jacoby (1971) au realizat o cercetare folosind o scală de răspuns tip Likert, variind numărul punctelor de scală de la 2 la 19. Concluzia celor doi autori a fost că fidelitatea test-retest, coeficientul

Cronbach-Alfa, validitatea concurrentă și predictivă nu au fost afectate de numărul de puncte de scală. Utilizarea unor scale cu mai multe puncte decât scalele dihotomice sau trihotomice nu crește rafinamentul măsurării.

Rezultate diferite au fost obținut de Finn (1972), într-o cercetare cu privire la complexitatea unor locuri de muncă. Finn a comparat scalele cu 3, 5, 7 și 9 puncte și a concluzionat că numărul optim al punctelor de scală este 7, luând în calcul fidelitatea și dorința de a maximiza varianța scorurilor obținute.

O cercetare simulată a fost realizată de Lissitz și Green (1975) care au comparat fidelitatea măsurării unor scale cu 2, 3, 5, 7, 9 și 14 puncte de scală, pentru trei niveluri de corelație între itemi. Rezultatele cercetării au arătat că fidelitatea crește odată cu creșterea numărului punctelor de scală, indiferent de nivelul de corelație dintre itemi.

Rezultate contradictorii au fost obținute de cercetători și în ultimii ani. În urma unei cercetări de piață de tip CATI (computer-assisted telephone interviews) Dowl (2007) a tras concluzia că nu există diferențe între scalele de 5, 7 și 10 puncte în ceea ce privește abaterea standard, skewness și kurtosis. Singura diferență este aceea că mediile scalelor de 5 și 7 puncte (între care nu există diferențe semnificative) sunt semnificativ mai mari decât media scalei de 10 puncte (după rescalare). Rezultate opuse au fost obținute de Bandalos și Enders (1996) în urma unei cercetări simulate. Aceștia afirmă că fidelitatea crește odată cu creșterea corelației dintre variabile și cu numărul punctelor de scală, cu un nivel maxim pentru scalele cu 5 și 7 puncte. Tot în urma unei cercetări simulate, Krieg

(1999) trage concluzia că erorile de măsurare variază în funcție de numărul punctelor de scală, numărul itemilor în scală, media și varianța cantității măsurate. În consecință, afirmă Krieg (1999, 764), ‘cea mai simplă cale de a evita erorile introduse de scalele de măsurare cu categorii distincte este să nu le mai folosim’.

Alternativa: scala de răspuns continuă

Alternativa este utilizarea unei scale grafice similară cu cea construită de Freyd în 1923. Apariția computerului poate ajuta la depășirea dezavantajelor practice pe care le prezenta scala grafică. Cook et al. (2001) au realizat o cercetare cu privire la gradul de satisfacție al utilizatorilor unui librării cu privire la calitatea serviciilor acestei instituții. Măsurarea satisfacției a fost realizată cu ajutorul a șapte dimensiuni și 41 de itemi aplicați în varianta de tip Likert și pe o scală continuă. În scala de tip Likert variantele de răspuns au fost numerice, de la 1 la 9, având design-ul de tip butoane radio. Scala continuă a fost construită în așa fel încât subiectul să aibă 100 de opțiuni. Pentru comparații aceasta a fost rescalată în două scale, una de 9 puncte, alta de 5 puncte. Cele două tipuri de scale au fost aplicate pe eșantioane diferite. Rezultatele cercetării au arătat că scala de tip Likert cu butoane a obținut un coeficient alfa mai mare în șase din șapte dimensiuni față de scala continuă, însă, per ansamblu, diferența a fost foarte mică (0,965 pentru scala de tip Likert comparativ cu 0,960 pentru scala continuă). Așa cum teoretic autorii s-au așteptat, coeficientul alfa

a fost mai mare pentru scala continuă (cu 100 de puncte) decât pentru scala de 9 puncte, care a fost mai mare decât pentru scala de 5 puncte. Prin urmare, comparativ cu scala continuă, este mai avantajos să folosești scala cu butoane doar pentru că subiecții necesită un timp de răspuns mai scăzut.

O comparație între scala continuă și cea de tip Likert a fost realizată și de Gerich (2007) care a aplicat scala controlului de sine atât pe hârtie, cât și pe computer (CASI – computer-assisted self-administrated interviews), în ambele variante discretă și continuă (în total patru versiuni). Scala continuă pe calculator a fost numită Scala Analog Vizuală (SAV). Pentru comparație, lungimea liniei scalei continue pe hârtie a fost de 66 de milimetri, iar în cazul computerului SAV a fost scalat cu 66 de puncte distincte (a căror valoare crește de la 0 la 66). Rezultatele cercetării au arătat că valoarea coeficientului Cronbach-Alfa este mai mare pentru scala SAV decât pentru cea de tip Likert, atât atunci când chestionarul a fost aplicat clasic pe hârtie, cât și în varianta CASI. De asemenea, au fost obținute rezultate mai bune în ceea ce privește fidelitatea măsurării prin administrarea CASI, decât pe hârtie în cazul scalei de tip Likert. Utilizarea unei scale SAV poate ridica probleme subiecților în ceea ce privește capacitatea lor de a se poziționa pe linia continuă, situație similară scalelor de tip Likert cu multe puncte scală (peste șapte).

Reips și Funke (2008) au realizat un experiment prin care au testat capacitatea subiecților de a poziționa cursorul pe SAV la anumite poziții date de cercetător în format procent (de exemplu, 75%) și în format raport

(de exemplu, $\frac{3}{4}$). O altă variabilă luată în calcul a fost lungimea liniei SAV, care a luat trei valori: 50 de pixeli, 200 de pixeli și 800 de pixeli. Au calculat apoi distanța dintre valoarea cerută și valoarea realizată, iar concluziile cercetătorilor au fost că subiecții au arătat o bună poziționare, cu valori mai mici pentru SAV de 50 de pixeli. Acest lucru, spun autorii, arată că măsurarea cu ajutorul SAV (indiferent de lungimea acesteia) este robustă la diferitele mărimi ale ecranului sau la rezoluții.

O posibilă explicație pentru rezultatele contradictorii este dată de Velicer și Stevenson (1978, 293) care atrag atenția că diferențele dintre rezultatele cercetării practice pot apărea din faptul că, în studii diferite au fost folosite instrumente de testare diferite și metodologii diferite. Un argument în favoarea a ceea ce au spus cei doi autori este faptul că, în comparație cu cercetările practice, în toate cercetările simulate numărul punctelor de scală a contat în măsurarea fidelității scalelor.

Experiment propriu

În ceea ce ne privește, am considerat necesară efectuarea unui experiment propriu, prin care să încercăm controlarea tuturor aspectelor contradictorii din studiile prezentate. Experimentul a fost efectuat cu ajutorul a 438 studenți ai secției de Sociologie de la Universitatea din București, care au răspuns consecutiv la două seturi de itemi identici, diferența fiind dată de scala de răspuns. Administrarea itemilor a fost făcută pe calculator, răspunsurile fiind indicate ori selectate cu ajutorul mouse-ului. Spre exemplu,

două scale de răspuns (una continuă și una cu cinci variante de răspuns) sunt prezentate în Figura 1.

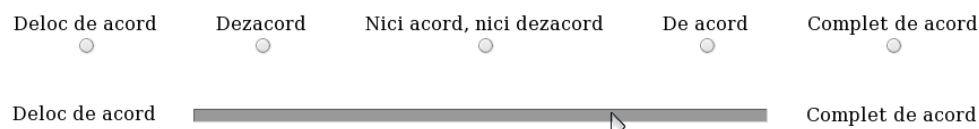


Figura 1. Două scale de răspuns, prima cu cinci trepte iar a doua continuă

Una dintre presupunerile noastre, pe care nici unul dintre studiile anterioare nu a luat-o în calcul, este aceea că diferențele dintre scorurile obținute la cele două tipuri de scale de răspuns pot să apară din construcția efectivă a scalelor (subliniem că prin scală înțelegem un instrument cu mai mulți itemi măsurați în același fel, pentru a face diferența față de scala de răspuns). Faptul că itemii aferenți unei dimensiuni nu merg în aceeași direcție poate explica diferența dintre scorurile de fidelitate, acestea fiind atribuite în mod eronat scalelor de răspuns (de unde și concluziile contradictorii dintre studii). Pentru a controla acest aspect, dat fiind faptul că este extrem de dificilă crearea unei scale cu o înaltă consistență internă (de obicei acestea sunt chiar patentate), am căutat o scală care să fie extrem de solidă metodologic recunoscută ca atare și validată în comunitatea academică.

Cele mai multe scale sunt construite în țări cu o mai bogată experiență metodologică, ajungând în România prin traducere. Acest fapt putea să se dovedească un alt factor de perturbare a răspunsurilor, întrucât prin traducere se pot introduce erori de măsurare. Chiar dacă în limba originală scala este solid validată, traducerea acesteia în limba română putea să introducă erori dificil de cuantificat, în consecință putea să scape condițiilor stricte ale controlului experimental.

Din acest motiv, am avut nevoie de o scală care să fie recunoscută atât în limba originală, dar să fie validată riguros și în varianta tradusă în limba română. Din fericire, astfel de instrumente există deja, cu precădere în domeniul testărilor psihologice. În România, una dintre companiile care administrează astfel de teste este D&D TestCentral, specializată în validarea de instrumente psihometrice consacrate la nivel internațional, care s-au impus în literatura de specialitate și sunt recunoscute de către mari asociații cum ar fi APA – American Psychology Association.

Un astfel de test este TOM – Testul de Orientare Motivațională (Borgogni, Petita și Barbaranelli, 2004) – utilizat în problematica motivației la locul de muncă, pe care o măsoară cu ajutorul a patru dimensiuni:

Orientarea spre Obiective (OO)

La locul de muncă, unele persoane tind să dea tot ce e mai bun dacă li se oferă oportunitatea de a interacționa cu sarcini de dificultate din ce în ce mai mare, dacă au ocazia să răspundă la provocări și încercări realiste și dacă primesc un feedback privind prestația realizată. Unii oameni sunt atrași de activități dificile și provocatoare și doresc să exceleze pentru simpla plăcere de a-și pune la încercare competențele și de a-și etala propriile posibilități.

Orientarea spre Inovație (OI)

Alte persoane au o performanță maximă dacă la locul de muncă pot să experimenteze lucruri noi, să exploreze situații mai puțin cunoscute și să lucreze la mai multe activități deodată, mai degrabă decât să se concentreze asupra unor activități repetitive. Acestor oameni le place să modifice proceduri și să militeze pentru schimbarea soluțiilor consacrate, să progreseze în mod constant și să gândească în mod creativ și divergent față de ceilalți.

Orientarea spre Conducere (OC)

Această dimensiune verifică ideea că unele persoane au o performanță optimă la locul de muncă dacă își pot asuma poziții care presupun influență și control, care le dau posibilitatea de a-i ghida pe ceilalți și de a trasa sarcini și responsabilități grupului de muncă. Aceste persoane sunt motivate să își impună alegerile și să fie în centrul

atenției, agreează să ia decizii în numele altora și doresc să îi determine pe ceilalți să le urmeze.

Orientarea spre Relaționare (OR)

În fine, mai sunt unele persoane care au o performanță maximă doar dacă pot lucra în cadrul unui grup și dacă climatul organizațional este pozitiv, bazat pe colegi care le sunt și prieteni și dacă pot primi din partea celorlalți susținere afectivă. Aceste persoane sunt motivate de colaborare, de posibilitatea de a evita orice tip de conflict și de solidaritatea cu colegii.

În total, au fost utilizați un număr de 70 de itemi (câte 15 pentru fiecare dimensiune), din care ultimii 10 aparțin unei a cincea dimensiuni numită Dezirabilitate Socială (pentru a testa dacă răspunsurile sunt sincere). Această dimensiune nu a fost utilizată în cadrul acestui articol, întrucât scopul nostru nu a fost de a valida testul de orientare motivațională, ci

Tabelul 1. Saturațiile itemilor pe cei patru factori rezultați din analiza factorială

Item	OO	Item	OI	Item	OL	Item	OR
1	0.86	16	0.84	31	0.83	46	0.86
2	0.86	17	0.84	32	0.84	47	0.85
3	0.87	18	0.84	33	0.84	48	0.86
4	0.84	19	0.86	34	0.84	49	0.84
5	0.86	20	0.84	35	0.86	50	0.86
6	0.87	21	0.85	36	0.82	51	0.86
7	0.86	22	0.85	37	0.86	52	0.86
8	0.87	23	0.85	38	0.84	53	0.86
9	0.86	24	0.83	39	0.87	54	0.87
10	0.85	25	0.84	40	0.82	55	0.86
11	0.87	26	0.84	41	0.87	56	0.87
12	0.87	27	0.85	42	0.86	57	0.86
13	0.86	28	0.85	43	0.88	58	0.87
14	0.87	29	0.83	44	0.88	59	0.86
15	0.86	30	0.83	45	0.84	60	0.84

Sursa: date preluate din studiul de validare al scalei TOM, cf. Pitariu, Vercelino & Iliescu (2009).

doar de a observa diferența dintre consistența internă a răspunsurilor pentru scale de răspuns diferite.

Tabelul 1 prezintă rezultatele analizei factoriale asupra itemilor care aparțin de cele 4 dimensiuni, mai exact modul cum itemii respectivi corelează cu factorii. Studiul de validare a acestei scale, în limba română, a fost efectuat pe un eșantion de aproximativ 2000 de respondenți, utilizându-se maximum likelihood ca metodă de extragere a factorilor și Promax cu normalizare de tip Kaiser ca metodă de rotire a acestora. Întrucât toate saturațiile sunt mai mari decât 0.8, se poate afirma fără îndoială că itemii respectivi aparțin dimensiunii corecte.

Respondenții au fost alocați în mod

aleator în cadrul a trei grupuri care au răspuns consecutiv la același set de itemi dar cu scale de răspuns diferite, pentru a testa diferențele dintre:

A. Scala de răspuns tip Likert cu 5 trepte versus SAV continuă.

B. Scala de răspuns tip Likert cu 7 trepte versus SAV continuă.

C. Scala de răspuns tip Likert cu 5 trepte versus aceeași scală cu 7 trepte.

Pentru a controla efectul ordinii scalei de răspuns asupra consistenței interne a răspunsurilor, cele două scale de răspuns au fost prezentate în mod alternativ. Spre exemplu în grupul A, unii studenți au răspuns prima oară la scala cu 5 trepte apoi la cea continuă, în timp ce alții au răspuns prima oară la scala continuă și apoi la cea cu 5 trepte.

Tabelul 2. Scorurile Cronbach-Alfa pentru toate dimensiunile și scalele de răspuns (R5 este SAV recodificată în 5 trepte iar R7 este SAV recodificată în 7 trepte)

Grup	N	Scală	Dimensiuni			
			OO	OI	OL	OR
A	170	5	0.82	0.81	0.92	0.87
		SAV	0.90	0.88	0.94	0.90
		R5	0.89	0.87	0.93	0.89
B	171	7	0.85	0.82	0.92	0.87
		SAV	0.90	0.86	0.93	0.90
		R7	0.89	0.86	0.92	0.90
C	97	5	0.84	0.80	0.90	0.86
		7	0.91	0.87	0.93	0.90

Sursa: Calcule proprii.

Tabelul 2 prezintă scorurile Cronbach-Alfa (C-A) pentru cele patru dimensiuni, în cadrul celor trei categorii de experimente A, B și C. Unul dintre aspectele cele mai importante care se pot observa din start este acela că, pentru scalele cu mai multe trepte, scorurile sunt mai mari. Spre exemplu, scorul C-A pentru prima dimensiune (OO) din cadrul grupului A este 0.90 pentru scala continuă, comparat cu 0.82

pentru scala de răspuns tip Likert cu 5 trepte. Unele diferențe sunt relativ mici (cum ar fi în cadrul grupului C pentru dimensiunea OL, adică 0.94 comparat cu 0.92), altele sunt de-a dreptul spectaculoase, în general însă pattern-ul se păstrează în mod sistematic. Așadar, unul dintre posibilele rezultate care merită a fi raportate în urma acestui experiment este faptul că scorul C-A crește pe măsura ce scala de răspuns își

mărește numărul de trepte, controlând toți ceilalți factori care pot influența răspunsurile respondenților.

Pentru a ne asigura de acest lucru, am dorit să verificăm dacă scorul C-A poate fi mărit în mod artificial pentru simplul motiv că variația scalei de răspuns cu mai multe trepte este mai mare. Dacă acest lucru ar fi adevărat, atunci comparația nu și-ar mai avea locul, întrucât am compara unități de măsură diferite.

Verificarea acestei posibilități se poate face prin recodificarea scalei continue într-o scală de răspuns cu același număr de trepte ca și cea cu care se compară. Procesul de recodificare este unul relativ simplu, prin împărțirea scalei continue pe quintile pentru scala cu 5 trepte (notată cu 'R5') și pe septile

în cazul scalei de răspuns cu 7 trepte (notată cu 'R7').

Rezultatele sunt de asemenea prezentate în Tabelul 2, confirmând rezultatul prezentat mai sus, anume că scorurile pe scala continuă sunt în mod sistematic mai mari decât scorurile pentru scalele cu mai puține trepte. Spre exemplu, în cazul aceleiași dimensiuni OO din cadrul grupului A, scorul C-A pentru scala continuă recodificată R5 este de 0.89, substanțial mai mare decât scorul C-A pentru scala de răspuns clasică.

O altă verificare posibilă a faptului că scorul C-A ar putea fi mărit în mod artificial în cazul scalei de răspuns cu mai multe trepte este utilizarea scorurilor C-A standardizate, pe care le prezentăm în Tabelul 3.

Tabelul 3. Scorurile Cronbach-Alfa standardizate pentru toate dimensiunile și scalele de răspuns

Grup	N	Scală	Dimensiuni			
			OO	OI	OL	OR
A	170	5	0.83	0.83	0.92	0.87
		SAV	0.90	0.89	0.94	0.90
		R5	0.89	0.87	0.93	0.89
B	171	7	0.87	0.83	0.92	0.87
		SAV	0.91	0.87	0.93	0.90
		R7	0.90	0.87	0.92	0.90
C	97	5	0.84	0.81	0.90	0.86
		7	0.91	0.88	0.93	0.90

Sursa: Calcule proprii.

În valori standardizate, scorurile C-A sunt de cele mai multe ori egale cu scorurile nestandardizate, ori cel mult mai mari cu 0,02. Se păstrează însă același tipar sistematic al scorurilor mai mari în cazul scalelor de răspuns cu mai multe trepte.

Faptul că pentru scalele recodificate, scorurile sunt (sistematic) mai mari

decât scorurile la scalele clasice cu același număr de trepte, relevă un alt aspect pe care îl considerăm important, anume că erorile de măsurare pentru scala continuă sunt mai mici, cu demonstrația că răspunsurile indivizilor au un grad mai înalt de consistență internă chiar și pentru scalele recodificate. Acest

fapt ne îndreptățește să avansăm următoarea ipoteză: *pe măsură ce se reduce numărul de trepte într-o scală de răspuns, scade precizia (consistența internă a) răspunsurilor*.

Într-o scală cu 5 trepte, faptul că subiectul alege 3 sau 4 contează foarte mult în economia scorului final și al consistenței răspunsurilor la itemi diferiți care aparțin de aceeași dimensiune. Noi credem că percepția intuitivă a subiecților cu privire la conceptul măsurat este reprezentată printr-o dimensiune continuă, de unde și dificultatea alegerii pe scala de răspuns cu 5 trepte, întrucât subiectul nu se poate poziționa cu exactitate să spunem pe cifra 3 sau pe cifra 4 ci, de cele mai multe ori, răspunsul se află între aceste cifre.

Concluzii

Deși extrem de solidă din punct de vedere metodologic, considerăm că neajunsurile scalei de răspuns tip Likert sunt mai mari decât beneficiile aduse de simplitatea utilizării acesteia. Aceste neajunsuri au fost identificate de multă vreme, însă până de curând a lipsit tehnologia necesară în utilizarea unei scale continue.

Scala Analog Vizuală reflectă mult mai bine natura continuă a conceptelor din științele sociale, care sunt măsurate în prezent într-un mod extrem de artificial prin utilizarea unor categorii distincte. Dintr-o dimensiune continuă este ușor de obținut o variabilă ordinală cum este scala de răspuns tip Likert, însă ceea ce se pierde nu mai poate fi recuperat decât cu extrem de mare dificultate.

Apariția computerelor și utilizarea

acestora în procesul de culegere a datelor (chiar în cadrul cercetărilor pe teren cu chestionare față în față), oferă în sfârșit tehnologia mult așteptată pentru măsurarea continuă a conceptelor. Așa cum am arătat prin experimentul nostru, fidelitatea măsurării continue nu este inferioară, ci superioară măsurării prin scala de răspuns tip Likert, fapt ce întărește motivația utilizării scalei continue.

Ca un argument final, apreciem că, științele sociale sunt extrem de văduvite prin imposibilitatea aplicării (de cele mai multe ori) a unor analize statistice avansate pentru simplul motiv că datele culese nu respectă asumțiile acestora (spre exemplul faptul că nu sunt metrice, dat fiind că (în aprecierea noastră) peste 90% din itemii chestionarelor sociologice sunt mășurați prin variabile categoriale).

Măsurarea continuă a conceptelor, cu ajutorul Scalei Analog Vizuale, deschide o poartă științelor sociale nu doar către analize statistice avansate, ci și către crearea mult mai facilă (am spune și mai corectă) a indicilor și indicatorilor sociali.

Concluziile prezentate în acest articol necesită o validare pe un eșantion mai larg, în special pentru variabilele educație și vârstă (dat fiind că în eșantionul de studenți acestea sunt constante), pentru a testa spre exemplu dacă persoanele cu o educație scăzută pot să răspundă pe un astfel de instrument.

Note

¹ Un efort considerabil a fost făcut pentru a adapta analize statistice la acest tip de date, prin dezvoltarea

corelației tetrachorice pentru datele bipolare (binare) și chiar a unui tip special de regresie ordinală.

² În articolul original, Cook et al. menționează o abatere standard egală cu 2, însă în opinia noastră cifrele

prezentate de aceștia sunt greșit calculate întrucât utilizează la numitor întregul volum al eșantionului n , în timp ce formula corectă a varianței și abaterii standard utilizează $n-1$.

Bibliografie

- Bandalos, D.L. and C.K. Enders (1996) The Effects of Nonnormality and Number of Response Categories on Reliability. *Applied Measurement Education*, 9(2):151-160.
- Borgogni L., Petitta L. and C. Barbaranelli (2004) Il Test di Orientamento Motivazionale (TOM) come strumento per la misurazione della motivazione al lavoro (The Test of Motivational Orientations as a tool to measure work motivation). *Bollettino di Psicologia Applicata*, 243:43-52.
- Chelcea, S. (2004) *Metodologia cercetării sociologice. Metode cantitative și calitative*. București: Editura Economică.
- Cook, C., Heath, F., Thomson, R.L. and B. Thomson (2001) Score Reliability in Web- or Internet-Based Surveys: Unnumbered Graphic Rating Scales Versus Likert-Type Scales. *Educational and Psychological Measurement*, 61(4):697-706.
- Dawes, J. (2008) Do Data Characteristics Change According to the Number of Scales Points Used? An Experiment using a 5-point, 7-point and 10-point scales. *International Journal of Market Research*, 50(1):66-77.
- Finn, R.H. (1972) Effects of some variations in rating scale characteristics on the means and reliabilities of rating. *Educational and Psychological Measurement*, 32(2):255-265.
- Freyd, M. (1923) *The Graphic Rating Scale*. *Journal of Educational Psychology*, 14(2):83-102.
- Kent, R. (2007) *Marketing Research. Approaches, Methods and Applications in Europe*. Londra: Thomson.
- Komorita, S.S. and W.K. Graham (1965) The number of scale points and the reliability of scales. *Educational and Psychological Measurement*, 25(4):987-995.
- Krieg Jr., E. (1999) Biases Induced by Coarse Measurement Scales. *Educational and Psychological Measurement*, 59(5):749-766.
- Gerich, J. (2007) Visual Analogue Scales for Mode Independent Measurement in self-Administered Questionnaires. *Behavior Research Methods*, 39(4):985-992.
- Likert, R. (1932) *A Technique for the Measurement of Attitudes*. New York: New York University Press.
- Lissitz, R.W. and S.B. Green (1975) Effect of the number of scale points on reliability: A Monte Carlo approach. *Journal of Applied Psychology*, 60(1):10-13.
- Matell, M.S. and J. Jacoby (1971) Is there an optimal number of alternatives for Likert scale items? Study I: Reliability and Validity. *Educational and Psychological Measurement*, 31(3):657-674.
- Pitariu, H., Vercelino, D. și D. Iliescu (2009) *Manual tehnic pentru Testul de*

Orientare Motivațională. Cluj-Napoca: Odiseea.

Reips, U.D. and F. Funke (2008) Interval-level measurement with visual analogue scales in Internet-based research: VAS Generator. *Behavior Research Methods*, 40(3):699-704.

Velicer, W.F. and J.F. Stevenson (1978) The Relation Between Item Format and the Structure of the Eysenck Personality Inventory. *Applied Psychological Measurement*, 2(2):293-304.

