

Statisztikai Szemle

Közzététel: 2024. február 12.

A tanulmány címe:

A mérés és a méréselmélet néhány alapkérdése

Szerző:

DUSEK TAMÁS

a Széchenyi István Egyetem egyetemi tanára, a Statisztikai Szemle főszerkesztője

E-mail: dusekt@sze.hu, Tamas.Dusek@ksh.hu

DOI: <https://doi.org/10.20311/stat2024.02.hu0107>

Az alábbi feltételek érvényesek minden, a Központi Statisztikai Hivatal (a továbbiakban: KSH) Statisztikai Szemle c. folyóiratában (a továbbiakban: Folyóirat) megjelenő tanulmányra. Felhasználó a tanulmány vagy annak részei felhasználásával egyidejűleg tudomásul veszi a jelen dokumentumban foglalt felhasználási feltételeket, és azokat magára nézve kötelezőnek fogadja el. Tudomásul veszi, hogy a jelen feltételek megszegéséből eredő valamennyi kárért felelősséggel tartozik.

1. A jogszabályi tartalom kivételével a tanulmányok a szerzői jogról szóló 1999. évi LXXVI. törvény (Szt.) szerint szerzői műnek minősülnek. A szerzői jog jogosultja a KSH.
2. A KSH földrajzi és időbeli korlátozás nélküli, nem kizárólagos, nem áthatható, térítésmentes felhasználási jogot biztosít a Felhasználó részére a tanulmány vonatkozásában.
3. A felhasználási jog keretében a Felhasználó jogosult a tanulmány:
 - a) oktatási és kutatási célú felhasználására (nyilvánosságra hozatalára és továbbítására a 4. pontban foglalt kivétellel) a Folyóirat és a szerző(k) feltüntetésével;
 - b) tartalmáról összefoglaló készítésére az írott és az elektronikus médiában a Folyóirat és a szerző(k) feltüntetésével;
 - c) részletének idézésére – az átvevő mű jellege és célja által indokolt terjedelemben és az eredetihez híven – a forrás, valamint az ott megjelölt szerző(k) megnevezésével.
4. A Felhasználó nem jogosult a tanulmány továbbértékesítésére, haszonszerzési célú felhasználására. Ez a korlátozás nem érinti a tanulmány felhasználásával előállított, de az Szt. szerint önálló szerzői műnek minősülő mű ilyen célú felhasználását.
5. A tanulmány átdolgozása, újra publikálása tilos.
6. A 3. a)–c) pontban foglaltak alapján a Folyóiratot és a szerző(ke)t az alábbiak szerint kell feltüntetni:
„*Forrás: Statisztikai Szemle c. folyóirat 102. évfolyam 2. számában megjelent, **Dusek Tamás** által írt, **A mérés és a méréselmélet néhány alapkérdése** című tanulmány (link csatolása)*”
7. A Folyóiratban megjelenő tanulmányok kutatói véleményeket tükröznek, amelyek nem feltétlenül esnek egybe a KSH vagy a szerzők által képviselt intézmények hivatalos álláspontjával.

Dusek Tamás

A mérés és a méréselmélet néhány alapkérdése

Some basic questions of measurement and measurement theory

Dusek Tamás, a Széchenyi István Egyetem egyetemi tanára, a Statisztikai Szemle főszerkesztője
E-mail: dusekt@sze.hu, Tamas.Dusek@ksh.hu

A statisztikai elemzésnek meglehetősen különleges a méréshez való viszonya, mivel a mérés és a méréselmélet nemcsak az elemzés alapjául szolgáló adatokat biztosítja, hanem jelentős bizonytalanságok forrása is lehet az elméleti, módszertani és alkalmazott kutatások során. Ezért különösen fontos a méréssel kapcsolatos kérdések árnyalt, részletes bemutatása, a különböző megközelítésekkel és a mögöttük álló érvekkel együtt. A tanulmány a mérés hétköznapi és természettudományos jelentésének ismertetését követően a mérés eltérő felfogásait vizsgálja, majd a társadalomstatistikában a 20. század közepére bekövetkezett drasztikus jelentésváltozására tér rá. Foglalkozik a mérési skálák elméletével, a skálák kritikájával, bővítési javaslataival, végül a mérési skálák keretén belül nem elhelyezhető Likert-skála típusú állítások egyes sajátosságait tárgyalja. Az írás egyik konklúziója, hogy a mérési skálák Stevens-féle rendszere nem tekinthető az adattípusok adekvát leírásának.

Kulcsszavak: mérés, mérési skálák, adattípusok, Stevens, Likert-skála

Statistical analysis has a special relationship to measurement. For it, measurement and measurement theory not only provide the data on which analysis is based, but are also a source of important uncertainties in theoretical, methodological and applied research. A detailed presentation of measurement issues is therefore justified, together with the different approaches and the arguments behind them. After a careful discussion of the everyday and scientific meanings of measurement, the paper deals with the different meanings and conceptions of measurement, and then turn to the drastic change in meaning in social statistics in the mid-20th century. Then the theory of measurement scales, their criticism and proposals for expansion will be discussed, and finally some of the specific aspects of Likert-type statements that cannot be placed within the framework of measurement scales will be examined. One of the conclusions of the paper is that the Stevens system of measurement scales cannot be regarded as an adequate description of data types.

Keywords: measurement, measurement scales, data type, Stevens, Likert scale

A mérés a tudományban és a hétköznapi életben is rendkívül sok kontextusban – a bolti súlyméréstől a vérnyomásmérésen át a kvantumfizikáig – előforduló, sokféle definícióval rendelkező fogalom. A méréselmélet kifejezést is számos tudományterület alkalmazza, eltérő értelemben. Ez a mérés mindenhol jelen lévő voltával és a valóságról szerzett alapvető információk nyérésében betöltött szerepével magyarázható. A mérésnek mint tevékenységnek az eredménye a tudománynak és a racionális hétköznapi döntéseknek is az egyik legfontosabb, objektív alapját biztosítja. Ez már önmagában is olyan helyzet, amely indokoltá teszi a méréssel kapcsolatos, tudományterületeken túlnyúló értelmezési, fogalmi, gyakorlati és a hátterben megbúvó ismeretelméleti kérdések tisztázó jellegű áttekintését.

Ezen túlmenően a statisztikai elemzés különleges, kitüntetett viszonyban van a méréssel. A mérés és a méréselmélet nemcsak az elemzés alapjául szolgáló adatokat biztosítja, hanem jelentős bizonytalanságok forrása is lehet az elméleti, módszertani és alkalmazott kutatások során. Az egymásnak ellentmondó vagy indokolatlan módszertani ajánlások befolyásolhatják a gyakorlati statisztikai munka bármely fázisát az adatfelvételtől az alkalmazott módszereken át az eredmények értékeléséig és a belőlük levont következtetésekig. Ezt tovább súlyosbítja a szakterületenként eltérő és szakterületeken belül sem mindig egységes terminológia, aminek jellemző tünete az egyes elemzéseket érő, ilyen és hasonló jellegű kritikai megjegyzések használata: „a munka szóhasználata bírálható”, „az adatok dichotomizálása érzékeny információvesztéssel eredményezett”, „az adatok dichotomizálása jelentősen növelte volna az eredmények áttekinthetőségét”, „a változó ordnális mérési szintű, ezért nem volt indokolt az alkalmazott módszer használata”, „a változó intervallum mérési szintű, ezért lehetőség lett volna megfelelőbb módszer alkalmazására”. Ilyen típusú, olykor egymásnak is ellentmondó bírálatok ugyanazon adatok elemzésével kapcsolatban is érkehetnek, attól függően, hogy a kritika megfogalmazója az adatfelvételre és a mérésre vonatkozóan melyik tankönyvet vagy ajánlást ismeri vagy fogadja el. Ha pedig a következtető statisztikai eljárásokra gondolunk, akkor tovább bővül a mérésre vonatkozó indokolatlan előírásokon alapuló bírálatok lehetősége (de az ilyen eljárások hiánya is kritika forrása lehet), például ilyen formában: „a normális eloszlás tesztelésének hiányában a módszer alkalmazása nem indokolt”.

A mérés kitüntetett szerepét az is mutatja, hogy nemcsak a statisztikai és a szaktudományos folyóiratok, tanulmányok, tanulmánykötetek és monográfiák foglalkoznak méréselméleti kérdésekkel, hanem önálló folyóiratai is vannak a témának.

Ezek jellemzően nem interdiszciplinárisak, hanem valamely tudományágra épülnek. Ha megnézzük a természettudomány és a technológia méréssel kapcsolatos kérdésköreinek folyóiratait (*Measurement, Measurement Techniques, Measurement and Control, Metrology, International Journal of Metrology and Quality Engineering, Measurement Techniques*), azok témáit, cikkeit, valamint a pszichológia, a viselkedéstudomány mérésével foglalkozó orgánumokat (*Applied Psychological Measurement, Educational and Psychological Measurement, Educational Measurement: Issues and Practice, Psychometrika, Measurement: Interdisciplinary Research and Perspectives, Measurement in Physical Education and Exercise Science*), akkor láthatjuk, hogy a közöttük lévő tartalmi átfedés minimális. Ami közös bennük, az a mérés fogalma, de a mérés jelentése, koncepcionális értelmezése, ismeretelméleti háttere már eltér, és legfeljebb nagyon általános szinten lehet olykor erőltetett analógiákat találni például a talajnedvesség mérése és az intelligencia mérése között.

Mindez tehát kellőképpen indokolja a méréssel kapcsolatos kérdések árnyaltabb, részletesebb bemutatását, a különböző megközelítésekkel és a mögöttük lévő indokokkal együtt. Az 1. táblázat összefoglalja, hogy mivel foglalkozik a tanulmány (az általánosabb kérdésekkel), és mivel nem, illetve azt is jelzi, hogy néhány kérdés csak érintőlegesen kerül majd szóba. Először a mérés hétköznapi és természettudományos jelentését árnyalom, majd a mérés jelentéstartalmának eltolódásával foglalkozom, végül a társadalomstatisztikában a 20. század közepére bekövetkezett drasztikus jelentésváltozására térek rá. Ez egyben előkészíti a mérési skálák elméletét, a skálák kritikáját és bővítési javaslatait. Ennek kapcsán bírálom azt a megközelítést, amely következtető statisztikai eljárásokat alkalmaz egyedi történeti események leírásakor, indokolatlanul vizsgálva a tesztek alkalmazásának olyan feltételeit, amelyek véletlen folyamatok esetén lennének csak érdekesek. Az adatok nem tudják, honnan származnak (*Lord, 1953*), de az adatokat elemzésre felhasználó kutatóknak tudniuk kellene. Ezután a társadalomstatisztikában kitüntetett szerepe miatt külön tárgyalom a Likert-skála típusú állítások (a fogalmat a tárgyalás során majd pontosítom) statisztikai problémáit, különös tekintettel arra a folyamatos vitára, amely a mérési skálák Stevens-féle kategorizálása óta egyfolytában napirenden van, hogy vajon mennyiségi, sorrendi, vagy esetleg kategorikus változók-e, illetve hogyan lehet ezeket kezelni, melyek a megengedett műveletek és statisztikai módszerek. A kérdés rendre felmerül, akár elméleti, módszertani tanulmányok témájaként, akár alkalmazott tanulmányok kapcsán, amikor az egyik táborba tartozó kutató azt kifogásolja, hogy miért kezelik a Likert-skála eredetű adatokat mennyiségi ismérvként, a másik pedig pont ennek az ellenkezőjét bírálja, vagyis miért nem kezelik azokat mennyiségiként. Emellett természetesen elképzelhetők egyéb megközelítések is, amelyek nem ennyire dogmatikusan kezelik a

kérdést. A probléma tulajdonképpen ezzel a vitával kapcsolatban az, hogy választási kényszert szeretne előírni egy olyan mérésiskála-rendszeren belül, amelybe annak hiányos volta miatt a Likert-skála be sem illeszthető.

1. táblázat

A méréssel kapcsolatos néhány fontosabb témakör
Some important issues on measurement

Kérdéskör	A tárgyalás mértéke
Általános, metatudományos, metakutatási méréselmélet	Részletes
Mérési skálák	Részletes
Méréstörténet (általános és résztudományokon belüli is)	Részleges
Különböző tudományágak és speciális területek méréselmélete	Részleges
Metrológia	Érintőleges
Adattranszformációk	Érintőleges
Kompozit mutatók	Nem tárgyalta
Mérési hiba (pontosság, megbízhatóság, hibaforrások)	Nem tárgyalta
Mérési eredmények közlése, kommunikációja	Nem tárgyalta
Digitalizáció, Big Data	Nem tárgyalta
A mérés társadalmi, gazdasági, kulturális, egészségügyi, szervezeteken belüli, hétköznapi szerepe, hatásai	Nem tárgyalta

Forrás: saját szerkesztés.

A tárgyalta témák történetére jellemző, hogy az elmúlt évtizedek során sem értek nyugvópontra, hanem ugyanazok a vitatémák és érvek bukkannak fel rendre, újra és újra. Ez a helyzet egyrészt ellentmond a kumulatíván növekvő és tökéletesedő tudomány eszményének, másrészt indokolja az azokra a régebbi tanulmányokra való hivatkozást, amelyek olyan koncepcionális kérdésekkel foglalkoznak, amelyek nem lezártak, ráadásul sokszor a kérdést később tárgyaló tanulmányoknál elmélyültebben elemezznek. A statisztikai elemzés környezetének változása (a digitalizáció, az adatforradalom, az adatokhoz való hozzáférés könnyebbége, a statisztikai szoftverek használata) a koncepcionális alapokat nem változtatta meg, de újabb kérdéseket vetett fel, összetettebbé tette a ténylegesen létező adattípusokat és növelte a többváltozós módszerek arányát. Ezen túlmenően a statisztikai szoftverek megerősíthettek olyan kétséges eljárásokat és értelmezéseket, amelyeket a készítőik beléjük építettek. Közülük a leglátványosabbak a következtető statisztika automatikusan szolgáltatott azon eredményei, amelyek eredetileg véletlen mintavételből származó adatok esetén értelmezhetők, és nem az önkényes és kényelmi mintákra, az önkényes elemzési egységekre, a teljes populációra vagy a módszertantól függő aggregált, súlyozott és transzformált adatokra.

A tárgyalás természetesen nem lehet mentes a korlátoktól és a hiányosságoktól. A téma hatalmas szakirodalmának csak egy szűk szeletére reflektálok. Magyar nyelven fontos előzmény Kehl (2011) tanulmánya, Fatalin (2017) munkája, valamint számos statisztikatankönyv mérési skálákkal foglalkozó része. Igyekeztem a különböző álláspontokat nemcsak bemutatni, hanem ütköztetni, értékelni is. Tisztában vagyok vele, hogy szűkebb szakterületek sajátosan használhatnak egy-egy fogalmat, illetve számos fogalomnak egymással párhuzamosan többféle meghatározása is létezik. Ezekre többször is utalok, de terminológiai megjegyzésekkel, pontosításokkal, kitérőkkel nem kívántam egyfolytában megtörni a gondolatmenetet, így előfordulhat, hogy ugyanaz a méréselméleti kifejezés a szövegkörnyezettől függően mást jelent. Ez a sokszínűség, a területenként és időben változó jelentés hozzátartozik a témához. Magára a mérésre többféle meghatározást ismertetek majd. Nem gondolom, hogy általános, mindenhol érvényes definíció adható, azt viszont igen, hogy lehetnek rossz definíciók, aminek a miértjét is megindokolom az adott helyen. Előljáróban pontosítanék: általában mérés alatt egyetlen megfigyelési egység vagy jelenség valamilyen jellemzője, változója értékének megállapítását értem. A mérés ennél szélesebb értelmezéseit alkalmazzák azonban például a kompozit mutatók képzésénél, a mutatószámok számításánál, amelyek több mérés eredményét használják fel. Ezekre néhol utalni fogok. További korlátot jelent, hogy nem foglalkozom az adatforradalom társadalmi és egyéb következményeit tárgyaló munkákkal és a mérés hétköznapi életben és adminisztratív szervezetekben betöltött szerepével, amelyek *A metrikus társadalom*, *A mérőszámok zsarnoksága*, *a Bizalom a számokban* és *a Bad Data. Miért mérünk rossz dolgokat, és miért hagyjuk ki gyakran a fontos mérőszámokat?* című könyvek témái.¹ Ezek nem a mérést bírálják, hanem inkább annak valamilyen szempontból inadekvát alkalmazásaival foglalkoznak, ami más szempontból rossz gyakorlat, mintha semmit se mérnének és csak az intuíciónak lehetne hagyatkozni a döntéseknél, értékeléseknél.

¹ Az említett könyvek eredeti címei: Steffen Mau: *The metric society. On the quantification of the social*; Jerry Z. Muller: *The tyranny of metrics*; Theodore M. Porter: *Trust in numbers. The pursuit of objectivity in science and public life*; Peter Schryvers: *Bad Data. Why we measure the wrong things and often miss the metrics that matter*. (A részletes bibliográfiai adatokat lásd az Irodalomban.)

1. A mérés hétköznapi, klasszikus, hagyományos és általánosabb természettudományos definíciói

Hétköznapi szóhasználatban a mérés olyan művelet, tevékenység, amellyel elvileg folytonos tulajdonságok nagyságát határozzák meg, a mérési pontosság és a kerekítések miatt gyakorlatilag diszkrét skálán. Az elvileg diszkrét nagyságokra már nem a mérést, hanem a számlálás, megszámlálás kifejezést használja a köznyelv.²

Mérésre utaló régészeti leletek már az írott történelem előtti időkből is fennmaradtak.³ A súly, a távolság és az idő mérése, a környezetben való tájékozódás, a navigáció a világgal való újfajta anyagi kapcsolat kialakítását tették lehetővé, amelynek egyszerre volt gyakorlati és koncepcionális, fogalmi hatása. A mérés egy megismerési tevékenység, amely a tapasztalati világ leírását, megértését, előrejelzését, racionalizálását segíti (*Renfrew–Morley, 2010; Morley, 2010*). A mérés valamilyen formájának korai megjelenése magyarázza, hogy klasszikus meghatározása, amely a mai hétköznapi szóhasználatban is tetten érhető, az ókorig vezethető vissza. Ez alapján a mérés egy olyan művelet, amely egy önkényesen választott egységgel hasonlítja össze a mérendő mennyiséget. A megmért mennyiség egy számból (mérőszámból) és egy mértékegységből áll. A mérőszám megmutatja, hogy a mértékegységet hányszor lehet a mérendő mennyiségbe belefoglalni. Ez a meghatározás szemléletesen működik az egyszerű mérőeszközöket igénylő, az ember méretéhez hasonló hosszúságok és súlyok mérésénél, aminek az ókorban is gyakorlati jelentősége volt. A hosszúság szubatomi, mikroszkopikus és csillagászati szintjén, illetve mozgó testek mérésénél azonban már egyéb mérési eljárásokra van szükség.

A mérésre adott korai definíció megtalálható Euklidész *Elemek* című művében, annak is az arányokról szóló, ötödik könyvében: „1. Egy kisebb mennyiség a nagyobbak része, ha osztja a nagyobbat. 2. Egy nagyobb mennyiség egy kisebbnek többszöröse, ha a kisebb osztja. 3. Az arány két egynemű mennyiség nagyságbeli viszonya.” (*Euklidész, 1983, 150. o.*). Ez a meghatározás csak szűk körben, a geometriai nagyságokra alkalmazható. A mennyiségek egyneműsége is korlátozza az arány jelentését. Euklidész arányokról szóló tanítása ugyanakkor jelentős befolyást gyakorolt az olyan reneszánsz művészekre, mint Leonardo da Vinci vagy Albrecht Dürer. Dürer Euklidész munkáját a mérésről írt tankönyvében és az arányelméleti értekezésében is felhasználta, amelyek kicsit később már Galileire és Newtonra voltak nagy hatással (*Michell, 2004, 30–32. o.*).

²Angolul a számlálás *counting*. A *measurement* (mérés) kiterjesztett jelentésébe beletartozik a *counting* is, ami a két kategorikusan különböző tevékenység egybemosása révén ugyanolyan megtévesztő, mint a magyar nyelvhasználatban.

³Ezekkel foglalkozik *A mérés archeológiája (The Archaeology of Measurement)* című kötet több tanulmánya (*Morley–Renfrew, 2010*).

Galilei (1564–1642) az ókori és a középkori spekulatív természetkutatással való szakítás egyik kiemelkedő alakjaként, a tapasztalati megfigyeléseken és méréseken alapuló természetkutatás úttörőjeként, szakítva az arisztotelészi hagyománnyal, azt írta: „Számold meg, ami megszámlálható, mérd meg, ami mérhető, ami nem mérhető, tedd mérhetővé” (*Ferris, 2004, 101. o.*). Ezt a számlálást és a mérést még külön fogalmakként kezelő gondolatot a modern menedzsment is előszeretettel idézi. Az ilyen rövid, frappáns, részletesen nem kifejtett megfogalmazások általános problémája azonban, hogy többféleképpen értelmezhetők, eltérő szinteken általánosíthatók, ami aztán a bírálatok és viszontbírálatok egész láncolatát indíthatja el. Ferris szerint naiv az az értelmezés, miszerint Galilei mindent számszerűsíteni akart. Az akkori (1500-as évek végi, 1600-as évek eleji), spekulatív jellegű természetfilozófiai környezetben próbálta meggyőzni kortársait az empirikus alapú, mérésekkel alátámasztott kutatás fontosságáról. Ez fokozatosan sikerült is, munkásságával jelentősen hozzájárult a természettudományos világkép kialakításához, és látványos felfedezésekkel és eredményekkel elméleti és technológiai haladást ért el. Galilei olyan kísérleteket végzett, amelyek számokkal leírható eredményeket hoztak. A számok között pedig a matematika segítségével tudott kapcsolatokat feltárni. Arra azonban többen is rámutattak, hogy önmagában az empiria, a tapasztalati kísérletezés nem vezetett volna el a mechanika matematikai összefüggéseinek a feltáráshoz. Ezekhez gondolatkísérletekre is szükség volt, ahogyan azt *Koyré (1968)* nagyon részletesen be is mutatta.

Sok természettudós értékelt a későbbiekben Galileihez hasonlóan a mérés tudományos felfedezésekben játszott szerepét. James Clerk Maxwell (1831–1879), akinek munkássága az egységes nemzetközi mértékegységrendszer kidolgozására és a fizika világképére is jelentősen hatott, így fogalmazott: „a tudomány fejlődésével a mennyiség területe mindenütt behatolt a minőség területére, mindaddig, amíg a tudományos vizsgálat folyamata úgy tűnt, hogy egyszerűen a mennyiségek mérése és regisztrálása lett, az így kapott számok matematikai elemzésével kombinálva” (*Maxwell, 1870, 419. o.*). Gyakran idézték Lord Kelvin (1824–1907, születési nevén William Thomson) következő, a mérés fontosságát hangsúlyozó előírását is: „Ha meg tudjuk mérni és számokban ki tudjuk fejezni azt, amiről beszélünk, akkor valamit tudunk róla, de ha nem tudjuk megmérni, nem tudjuk számokban kifejezni, akkor tudásunk szegényes, nem kielégítő”⁴ (idézi *Yule–Kendall,*

⁴ Az ökonometria és a matematikai közgazdaságtan elterjesztésében jelentős szerepet játszó Cowles Bizottság eredeti mottója, Kelvin mondásától ihletve, 1932-től „A tudomány: mérés” volt, ami 1952-ben „Elmélet és mérés”-re változott (*Christ, 1994*). A chicagói egyetem társadalomtudományi épületére Kelvin mondásának ezt a rövidített változatát írták: „Ha nem tudod mérni, a tudásod csekély és nem kielégítő” (*Kuhn, 1961, 161. o.*). Kuhn tanulmányában ismerteti Frank Knight értékelését erről a mondásról, amelynek a társadalomtudósok számára a gyakorlati jelentése szerinte a következő: „Ha nem tudod mérni, mérd meg bárhogy!” (vagy nem szó szerint fordítva: „Ha nem tudod megmérni, akkor is mérd meg!”) (*Kuhn, 1961, 164. o.*).

1964, 15. o.). Ezt, a Galilei előbb említett mondásához hasonlóan szintén többféleképpen értelmezhető állítást Yule és Kendall így értékeli: „Ezt a kifejezést sokszor idézték olyan helyesléssel, amelyet egyáltalán nem érdemel meg – például ennek alapján nem lenne említésre méltó Darwin és Pasteur munkássága sem, hogy Kelvin kortársai közül csak kettőt említsünk” (Yule–Kendall, 1964, 15. o.).

Az egységes nemzetközi mértékegységrendszer létrejöttében fontos szerepet játszott a már említett Maxwellnek a mérésről szóló meghatározása. Eszerint a mérés két összetevőből áll: egy ismert, a kifejezendő mennyiséggel azonos jellegű mennyiség nevéből (mértékegység), valamint abból, hogy hányszor kell a mértékegységet venni a kívánt mennyiség előállításához. Ez a fogalmi előrelépés segített a fizikai mérések egységesített nemzetközi mértékegységrendszerének a kidolgozásában és fokozatos továbbfejlesztésében (De Boer, 1995). A Nemzetközi Mértékegységrendszer (*Système International D'Unités – SI*) jelenleg hét fizikai alapmennyiséget fogad el: hosszúság, tömeg, idő (1954-ig ez volt a három alapmennyiség), hőmérséklet, anyagmennyiség, fényerősség és áramerősség. Ezek közül a hőmérsékletet Arisztotelésztől kezdve sokáig minőségi, és nem pedig mennyiségi kategóriának tartották úgy, hogy a hideg és a meleg két külön tulajdonság, amelyek egymástól elkülönítve azonosíthatók. Ez a felfogás uralkodott a középkor végéig, mígnem Galilei elkészítette az első hőmérőt⁵ (Barnett, 1956). A hőmérséklet abban különbözik a hosszúságtól, a tömegtől és az időtől, hogy nincs kiterjedése térben vagy időben. A hőmérséklet mérése kapcsán ezért már nem lenne kielégítő a mérés fenti klasszikus definíciója (amely ráadásul negatív előjelű mérési eredményt nem is enged meg), ahogyan a fényerősségre és az áramerősségre sem, amelyek műszeres mérést igényelnek. Így a fizikai mennyiségek mérésére vonatkozó általánosabb meghatározás szerint a mérés olyan műveletek összessége, amelynek célja egy mennyiség értékének a meghatározása. A hét fizikai alapmennyiség meghatározása alapvető méréssel, míg az alapmennyiségek közötti viszonyokon alapuló származtatott egységek – amelyekből száznál is több van – meghatározása származtatott méréssel történik. Az anyagmennyiség annyiban tér el a többi alapmennyiségtől, hogy elvileg diszkrét, megszámlálható nagyságot jelöl, és nem folytonost. Az alapmennyiségek kiegészülnek a szöggel és a térszöggel.

Az SI-hez és a műszaki, technológiai, laboratóriumi mérésekhez kötődik a metrológia, más néven méréstudomány, amely a fizikai mennyiségek mérésével foglalkozik. A metrológián belül a mérés fogalmára számos meghatározás született. A metrológia alapvető és általános fogalmainak nemzetközi szótára első kiadása (1984) szerint a mérés olyan műveletek összessége, amelyek célja egy mennyiség értékének meghatározása. A szótár harmadik, 2008-as kiadása a mérést már úgy

⁵ Az első hőmérőt Galilei vagy Galilei kortársa, ismerőse, munkatársa, Santorio Santorio készítette, aki bevezette az orvoslásba a kísérleti módszereket és a mérésen alapuló eljárásokat. Arról is vita van, hogy az első pulzuszámológót Santorio vagy Galilei készítette-e (Bigotti–Barry, 2022; Hay, 2020).

definiálja, mint egy vagy több, egy mennyiséghez észszerűen hozzárendelhető mennyiségérték kísérleti úton történő meghatározásának folyamatát (Mari, 2013).

A metrológia a Nemzetközi Súly- és Mértékügyi Hivatal (*International Bureau of Weights and Measures*) meghatározása szerint a mérés tudománya, amely magában foglalja mind a kísérleti, mind az elméleti vizsgálatokat a tudomány és a technológia bármely területén, a bizonytalanság bármely szintjén (Brown, 2021). Itt a tudomány inkább csak a természettudományt jelenti, a meghatározást kiegészítő szöveg mérésre vonatkozó példái mind fizikai és kémiai mérések. A következő tankönyvi meghatározásban szerepel is ez a korlátozás: „A metrológia a mérések tanulmányozásával foglalkozik. Rendkívül fontos, hogy különböző típusú paramétereket vagy fizikai változókat mérjünk, és mindegyiket egy meghatározott mértékegységgel számszerűsítsük. A mérés tehát egy fizikai változóhoz pontos és precíz értéket rendel. A fizikai változó ezáltal átalakul egy mért változóvá. Az értelmes mérésekhez közös mérési szabványokra van szükség, amelyeket kötelező alkalmazni. A közös mérési módszerek a nemzetközi specifikációs szabványok kidolgozásán alapulnak” (Raghavendra–Krishnamurthy, 2013, 3. o.). Végül egy tankönyvi leírás, amely a metrológia alkalmazott mivoltát emeli ki: „A metrológia a mérések tudománya. A mértékegységek és szabványaik megállapításával, reprodukálásával, megőrzésével és átadásával foglalkozik. A metrológia a különböző mérőműszerek átfogó tanulmányozását végzi a precizitás (*precision*), a pontosság (*accuracy*), a lehetséges hibaforrások és a hibák kiküszöbölésére szolgáló módszerek meghatározására a pontosság (*accuracy*) javítása érdekében” (Patil, 2018, 1. o.). (A meghatározásokban szereplő precizitás és pontosság fogalmak tartalma is érdekes lenne.)

Sok hasonló, egymástól hangsúlyjaiban némileg eltérő definíciót lehetne még idézni, de a műszaki, laboratóriumi mérés valamilyen módon általában szerepel bennük. A metrológia művelői jellemzően olyan laboratóriumokban és szabványokkal foglalkozó hivatalokban dolgoznak, amelyek felelősek a mérőműszerek kalibrációjáért, a mérési bizonytalanság feltárásáért, a mértékegységek közötti átváltásokért. A szakemberek között sok az olyan specialista, aki csak egy-egy szűkebb terület mérésével találkozik. A metrológia tárgyát áttekintő tanulmányok, a metrológia kézikönyvei is nagyrészt műszaki jellegű kérdésekkel foglalkoznak, azon belül is főleg a technológiai változásokkal, a digitalizációval, a mérőeszközök és berendezések, rendszerek kalibrálásával, és csak kisebb részben a mérés koncepcionális problémáival. Ezeknek a műszaki, technológiai, minőségbiztosítási, minőség-ellenőrzési kérdéseknek az alkalmazásai jócskán túlnyúlnak az iparon és a laboratóriumokon, mivel a társadalmi-gazdasági élet minden területe érintett (élelmiszer-biztonság, kommunikáció, közlekedés, környezet, egészségügy, oktatás stb.). A metrológia története pedig nagyrészt a mérési szabványok és technológiák története, kiegészítve a mérés társadalmi, jogi, gazdasági és üzleti kör-

nyezetével. Ezeken az alkalmazási területeken túl azonban a metrológiának a méréselmélettel foglalkozó általánosabb, elméleti, ismeretelméleti, koncepcionális, fogalmi részei is vannak.

2. A mérés modern jelentése (általános igényű meghatározások)

Ha a mérést csak a természettudományhoz kötnénk, és általánosítanánk Kelvinnek azt a mondását, amelyet eredetileg az elektromos mértékegységek kapcsán fogalmazott meg, akkor az – ahogyan *Yule és Kendall (1964)* korábban bemutatott értékelése erre rá is mutatott – a tudomány jelentős részének leértékelését jelentené, köztük olyan területekét, amelyek rendkívül komplex jelenségek magyarázatával foglalkoznak, és tapasztalatok és megfigyelések sokasága áll mögöttük. Teljes mértékben elfogadható, hogy a hideg, langyos, meleg meghatározások nem elégségesek, nem egzakttak a hőmérséklet jellemzésére. A szabadon eső testek gyorsulásának érzékszervi megfigyelése is hiányos, a gyorsulás mértékének kifejezéséhez mérésre van szükség. Az ilyen és ehhez hasonló esetekkel kapcsolatban Kelvin és Maxwell jogos igényt fogalmaz meg, de a tudomány nemcsak mechanisztikus természeti jelenségekkel és fizikai jelenségek mérésén alapuló következtetésekkel foglalkozik.

A helyzetet azonban ennél bonyolultabbá tette a méréselmélet az ismeretelmélet, azon belül főleg a természettudomány ismeretelmélete felől megközelítő egyes szerzők azon törekvése, hogy általánosabb meghatározást adjanak a mérésnek, amely bármely területen alkalmazható. Már Kelvin idejében megkezdődött – még a természettudományt szem előtt tartó kutatók részéről – az a tendencia, amely a mérésnek egyre tágabb és tágabb értelmezést kívánt adni. Ez a jelentésbővítés egy darabig természetes volt, ahogyan az egyszerű földméréstől az egyre bonyolultabb kalibrációkat és segédelméleteket igénylő műszeres mérésekig és a társadalmi, pszichológiai jelenségek méréséig eljutott a tudomány.

A mérés tapasztalati tudományoktól elválasztott, absztrakt, logikai, ismeretelméleti tárgyalása a rendkívül sokoldalú orvos és fizikus, Hermann von Helmholtz (1821–1894) számolásról és mérésről írt 1887-es tanulmányával kezdődött. A modern méréselmélet⁶ vízvázlatját Helmholtzhoz vagy kicsit későbbre, Otto Hölder (1859–1937) 1901-es munkájához teszik a méréstörténettel foglalkozók. Hölder

⁶ A klasszikus és a modern jelző használata történelmi idősakra utal, így megvan az a hátránya, hogy nem tartalmi sajátosságok alapján utal a jelentésre. A többféle modern felfogás azonban jobban különbözik a klasszikustól, mint egymástól, ezért itt ezt a jelzőt használok, és később írok majd a modern felfogáson belüli különbségekről.

matematikus volt, Kronecker és Weierstrass tanítványa, írásának a címe pedig: *A mennyiség axiómái és a mérés elmélete*.⁷ Angol fordítása első részének bevezető tanulmányában *Michell (1996, 237. o.)* azt írta, hogy a mai matematikai tudás alapján Hölder munkája nem tartalmaz újdonságot, így történeti érdekesség. A fordítás második részét megelőző bevezetőben ezt a véleményét már némiképp korrigálta (*Michell, 1997a, 346. o.*). Későbbi szerzőkkel összehasonlítva Höldert, *Michell (1996)* azt írja, hogy Bertrand Russell (a matematika és a logika oldaláról közelítve a kérdést), Norman Campbell (a fizika oldaláról) és Ernest Nagel nézetei (ismeretelméleti megközelítéssel, a természettudományt szem előtt tartva) Hölderhez képest naiv és matematikailag kiforratlanok voltak. A három említett szerző mérésre vonatkozó meghatározása a 2. táblázatban olvasható. A meghatározásokban közös, hogy a mérés eredménye számmal van kifejezve. A mérés módjáról, részleteiről nem írnak, a mérés tárgya különbözőképpen van megfogalmazva, ahogyan a mérés tárgya és a mérés eredménye közötti viszony is (kölsönös megfelelés, rendelés, megfeleltetés). A mérés tárgya: nagyságok, minőségek, entitások – a fordítás során más szavak is szóba kerülhetnek, de az eredeti angol megfogalmazások is jelentősen eltérő szavakat használnak: *magnitudes, qualities, entities*. Campbell a részletes kifejtésben ugyanakkor lényegében visszatér a hagyományos meghatározáshoz, amikor arról ír, hogy a mérésnek egyetlen önkényes eleme van, a mértékegység meghatározása (*Campbell, 1920, 290–294. o.*). Tehát Campbell rövid meghatározása előrevetíti a mérés jelentésének a megváltozását, viszont egész fogalmi kerete a mérés hagyományos értelmezéséhez áll közel.

2. táblázat

A mérés absztrakt meghatározásai
Abstract definitions of measurement

Meghatározás	Forrás
„A mennyiségek mérése a legáltalánosabb értelemben, minden olyan módszer, amellyel egy egyedi és kölcsönös megfelelést állapítanak meg az összes vagy egyes nagyságok és az összes vagy néhány szám között, amelyek lehetnek egész, racionális vagy valós számok.”	Russel, 1903, 176. o.
„A mérés az a folyamat, amelynek során számokat rendelünk a minőségek jelölésére: a mérés célja az, hogy a matematikai analízis hatalmas fegyverét a tudomány tárgyára alkalmazhassuk.”	Campbell, 1920, 267. o.
„A mérés olyan entitások számokkal való megfeleltetése, amelyek nem számok.”	Nagel, 1931, 313. o.

⁷ Németül: *Die Axiome der Quantität und die Lehre vom Mass*. Angol fordításának címe: *The axioms of quantity and the theory of measurement*. Az angol fordításnak is meg kellett küzdenie azzal a problémával, hogy a Hölder publikációjának megjelenése óta eltelt időben megváltozott (új jelentéssel egészült ki) a német *Mass*, az angol *measure* és *measurement* jelentése is. Szó szerint a német *Mass* az angol *measure*-nek, a magyar mértéknek felelne meg. (Erről lásd még a fordítók megjegyzését, *Michell–Ernst, 1996, 250. o.*)

Míg a 2. táblázatban olvasható három meghatározásban a mérés eredménye még számmal van kifejezve, addig Herman Weyl matematikus, elméleti fizikus 1926-ban már megengedi a nem számmal, hanem bármilyen szimbólummal való kifejezést: „Úgy tűnik azonban, hogy minden mérés egyetlen meghatározó jellemzője a szimbolikus ábrázolás; még a számok sem az egyetlen használható szimbólumok. A mérés lehetővé teszi, hogy a dolgok (a feltételezett mérési alaphoz képest) fogalmilag, szimbólumok segítségével jelenjenek meg” (Weyl, 1949, 144. o.). (Az 1949-es angol kiadás az 1926-ban megjelent német kötet fordítása.) A mérés elkezdődött jelentésváltozása, amelynek összefoglalása a 3. táblázatban látható, jelentős problémákat okozott azzal, hogy az új definíciók egy része absztraktságukban megválaszolatlanul hagyja azt az alapvető kérdést, hogy a minőségek számokkal kifejezése mérés-e (bár Campbell a minőség kifejezést használta, ő a fizikára gondolva írta a meghatározást, míg Russel és Nagel általában a tudományra gondolt). A mennyiségek és a minőségek megkülönböztetése lételméleti kérdés, ami Arisztotelésznél úgy jelenik meg, hogy valami vagy mennyiség, vagy minőség. Ha a mérés tárgya felől közelítünk a méréshez, akkor vannak nem mérhető tulajdonságok, azaz minőségek. Ha a mérés eredménye, vagyis egy szám felől közelítünk, akkor minden mérhető lesz. Ez az egyik legfontosabb különbség a mérés hagyományos és modern felfogása, megközelítése között.

3. táblázat

A mérés hagyományos és modern felfogása
Traditional and modern concepts of measurement

Hagyományos felfogás	Modern felfogás
A mért objektumra, tulajdonságra összpontosít.	A mérés eredményére összpontosít.
A mennyiség a fontos.	A szimbólum a fontos, ami többnyire egy szám.
A mennyiségi tulajdonság létezése feltétele a mérésnek.	A mérés eredménye egy szám (vagy szimbólum), bármilyen tulajdonság kifejezhető számokkal (szimbólumokkal).

Forrás: saját szerkesztés.

Ezek az absztrakt, a tapasztalati jellegű vizsgálatoktól elszakított méréselméleti írások már igen korán oda vezettek, hogy az egyértelmű nyelvhasználat egyre inkább veszélybe került. A tartalmi kérdések tárgyalása helyett a formális és szemantikai álviták lehetőségét teremtették meg. A szemantikai vitára kiváló példa a következő eszmecsere. Az asztrofizikus Arthur Eddington 1928-as munkájában, *A fizikai világ természetében* (*Nature of the physical world*) Kelvin előírásához nagyon hasonló fogalmazott meg a mérésről és a tudományosságról, azt állítva, hogy a tudomány és a nem tudomány közötti határvonal a metrikus és a nem metrikus közötti határvonal. Ezt a kijelentést Herbert Dingle azzal támadta, hogy az

evolúció elmélete tagadhatatlanul tudományos, még sincs köze a mérésnek a minőségi változások leírásához, magyarázatához. Erre a megjegyzésre Vibert Douglas úgy reagált, hogy valójában az evolúció is mérhető különbségekkel foglalkozik, az élővilág egyedeinek a méretére, tömegére és azok időbeli változására vonatkozóan (*Douglas, 1947*).

A mérés most bemutatott, 20. század elejei absztrakt felfogásain alapul a Stevens-féle mérési skálák elmélete, amellyel hamarosan részletesen is foglalkozom, valamint a reprezentációs méréselmélet, amely a mérést két kapcsolódó struktúra, egy empirikus és egy szimbolikus (ami lehet numerikus) struktúra közötti leképezésként jellemzi. Ennek a megközelítésnek a paradigmaticus alkotása *A mérés alapjai* című, sokszerzős, háromkötetes, 1971-ben, 1989-ben és 1990-ben megjelent munka. Az első kötet bevezetőjében a szerzők arról írnak, hogy a közgazdasági, matematikai, filozófiai, fizikai, pszichológiai, statisztikai szakirodalomban olyan axiómarendszerek és tételek találhatók, amelyek azt hivatottak megmagyarázni, hogy a tárgyak és az események tulajdonságai miért reprezentálhatók számszerűen. Ezek az eredmények alkotják a mérés matematikai alapjait. A nem fizikai tudományokban a mérés mindig problémás volt, ezért nem véletlen, hogy négy viselkedéskutató próbálja meg összefoglalóan tárgyalni a méréssel kapcsolatos kérdéseket (*Krantz et al. 1971, XVII–XVIII. o.*). A közel 1500 oldalnyi írás nem könnyű olvasmány, mivel a szerzők halmazelméleti jelölésekkel fejezik ki magukat még a hétköznapi szóhasználatnál nagyon egyszerűen leírható koncepciók tárgyalásakor is. A kötetek sikeresek voltak abban az értelemben, hogy a filozófián belül dominánssá tették a reprezentációs elméletet az 1990-es évekre (*Michell, 1993*). Abban az értelemben azonban bírálható ez a megközelítés, hogy a mérést a reprezentációra szűkíti le, anélkül, hogy meghatározná valaminek a tényleges mérési folyamatát, és figyelmen kívül hagyja a tényleges mérés során felvetődő olyan alapvető kérdéseket, mint a mérési hiba és a megbízható mérőműszerek és eljárások készítése, kialakítása. A reprezentációs elmélet sajátosságainak ismertetésekor *Swistak (1990)* rámutat arra, hogy legszembetűnőbb vonásként nincsenek megkötések arra vonatkozóan, hogy mit nevezhetünk mennyiségnek, vagyis a mennyiség fogalma nem játszik korlátozó szerepet abban, hogy mit lehet (és mit nem lehet) mérni.

A mérésnek természetesen számos további megközelítése is létezik, de az eddig említett három, és a hamarosan részletezendő operacionális felfogás, amelyek lényegét a 4. táblázatban foglaltam össze, és amelyek közül a reprezentációs megközelítés áll a legtávolabb a mérés gyakorlatától, tekinthetők a legelterjedtebbeknek.⁸ Az általános hagyományos felfogást lehet realistának is hívni, amely szerint

⁸ A reprezentációs és az operacionális felfogásról magyarul áttekintést ad *Kehl (2011)* tanulmánya. Az egyéb felfogásokat *Swistak (1990)* és *Mari–Wilson–Mau (2023)* jól bemutatják, valamint *Michell* realista nézőpontú művei (egy részükre hivatkozom) szintén ajánlhatók.

minden objektum minden tulajdonságának eredendően van egy értéke, amit néha valódi értéknek neveznek. A mérés célja ennek az értéknek a felfedezése.

4. táblázat

A mérés általánosabb felfogásainak összefoglalása
Summary of the general concepts of measurement

Felfogás	A mérés meghatározása
Klasszikus	Olyan művelet, amely egy önkényesen választott egységgel hasonlítja össze a mérendő mennyiséget.
Hagyományos (általánosított)	Olyan műveletek összessége, amelynek célja egy méréstől függetlenül is definiált mennyiség értékének a meghatározása.
Modern – operacionális	Az a folyamat, eljárás, amelynek során számokat rendelünk a dolgokhoz és azok tulajdonságaihoz.
Modern – reprezentációs	Két kapcsolódó struktúra, egy empirikus és egy szimbolikus struktúra közötti leképezés.

Forrás: saját szerkesztés.

Az operacionális és a reprezentációs felfogás közötti különbséget kisebbnek érzem, mint a hagyományos és a modern felfogás közötti különbséget, mivel önmagában az a tény, hogy a mérés eredményének számnak kell-e lennie, vagy nem szám is lehet, a modern felfogások esetében nem annyira fontos kérdés, mert a nem számok könnyen helyettesíthetők számmal. Számokat fel lehet használni (1) címkeként, azonosító jelként, illetve (2) rangsorolásra, valamint (3) mennyiségi kapcsolatok, arányok kifejezésére. A mérés meghatározásának változásából fakadó problémát az okozza, hogy hagyományos értelemben csak a számok harmadik felhasználása mérés, míg modern értelemben mindhárom felhasználás mérés, csak más mérési szintű.

3. Mérés a pszichológiában (nem fizikai mennyiségek mérése)

Galilei mérésre, számszerűsítésre vonatkozó javaslatát a társadalom és az ember tanulmányozásában is alkalmazni kezdték, ami a mérés koncepciójának a további bővülésével járt. Ez egy nagyon lassú folyamat volt, és a 19. század közepéig a mérés és a kvantifikáció szórványos alkalmazását jelentette, nagyrészt standardizált, összehasonlítható mérési eredmények nélkül. A kérdést tudományáganként,

kutatási területenként sokan feldolgozták. Utaltam már Galilei kortársára, Santorióra (1561–1636), aki számos élettani jellemző mérésére készített mérőeszközt, de egységes skálákat nem tudott kidolgozni (Hollerbach, 2023). A mérés korai alkalmazásaival találkozunk a 17. századi Angliában politikai aritmetika néven, valamint a valószínűségszámítás gazdasági és demográfiai alkalmazásaival, majd statisztika elnevezéssel a német egyetemeken a 18. században. A gazdaságtudomány a 18–19. században még szorosan együttműködött a tapasztalati megfigyelésekkel, kortárs adatokkal és statisztikákkal, majd a 19. század végétől a két terület (a mérés/gazdaságtudomány/empíria és az elmélet) egyre jobban szétvált, aminek meglettek a negatív következményei. A mérés a közgazdaságtanon belül nagyon eltérő területeken jelenik meg, így egészen mások a szubjektív és a csereérték mérésével, a racionális gazdasági kalkulációval, a gazdasági mutatók és tevékenységek különböző aggregátumainak mérésével kapcsolatos kérdések.⁹ A gazdasági megfigyelések és mérések sajátosságait, a természettudományos méréstől való koncepcionális különbségeit legrészletesebben, legszisztematikusabban Oskar Morgenstern *A gazdasági megfigyelések pontosságáról (On the accuracy of economic observations)* című könyve tárgyalta, még 1963-ban, amelynek koncepcionális része ma ugyanúgy aktuális, mint a megírásakor. Szerinte az adatpontosság, amit nagyon szélesen értelmez, a közgazdászok részéről kevésbé vizsgált, mint a fizikában, ahol már rutinfeladattá vált (Morgenstern, 1963, 7. o.).

A fizikai mérésektől távol eső terület a pszichológia is, ahol a mérés szintén nagyon fontos szerepet játszik a gyakorlatban. A pszichológiai méréselmélet rendkívül kidolgozott, de az ezzel kapcsolatos bizonytalanságok miatt a pszichológiát az „inexact tudományok királynőjének”¹⁰ is nevezték már. Bizonyos dilemmák, mint a pszichológiai konstrukció közvetlen mérésének nehézségei, hasonló módon merülnek fel a szociológiában és a közgazdaságtanban is, az itt alkalmazott módszerek egy része (mint a kérdőívek és a később külön tárgyalandó Likert-skála) a társadalomtudományok számára is fontos. Emellett a fogyasztói magatartás elméleti és gyakorlati vizsgálatában a közgazdaságtan és a pszichológia is illetékes.¹¹ Az alábbiakban a társadalomstatisztikára gyakorolt erős hatás és szoros kapcsolat miatt a pszichológiai mérés koncepcionális háttérét vázoló fel, nagyon szelektíven, a történeti és egyéb részletekre nem is térek ki.

⁹ A gazdaságtudományi mérések történetét és egyes koncepcionális kéréseit dolgozza fel Hüttl Antónia két tanulmánya (Hüttl, 2003a; Hüttl, 2003b).

¹⁰ Comrey (1951) írta azt, hogy a természettudományokkal összehasonlítva valamelyest kedvezőtlen a hírneve a pszichológiai mérésnek („queen of the inexact sciences”).

¹¹ A hasznosság mérése kapcsán sok szempontból hasonló elméleti, koncepcionális viták zajlottak a gazdaságtudományban, mint a pszichológiában az érzelmek mérése kapcsán. Ezt jól rekonstruálja Ivan Moscati a 2019-ben megjelent könyvében (Moscati, 2019). A hasznosság mérésének és az érzelmek mérésének gyakorlati oldala eközben teljesen eltér egymástól.

A korai kísérleti pszichológiában a mérés az érzékenység (fény, hangerő, hangmagasság) és a mozgások megfigyelésével jelent meg (Pléh, 1992). Ezekhez még pszichofizikai jellegű mérések társultak, amelyekben ötvöződnek a fizikai, fiziológiai és pszichológiai jelenségek. A mérés nehézségeit itt az okozza, hogy bár a fizikai inger nagysága számszerűen mérhető, az arra adott érzetválasz mérésénél akadnak nehézségek. A pszichofizika megalapítójának tartott Gustav Fechner (1801–1887) a méréssel kapcsolatban Galileihez hasonlóan fogalmazott 1860-ban *A pszichofizika elemei* című könyve előszavában: „A pszichofizikának mint egzakt tudománynak a fizikához hasonlóan a tapasztalatra és a méréssel nyert empirikus tényeknek a matematikai összekapcsolására kell épülnie, vagy ha ilyen mérés még nem áll rendelkezésre, azt meg kell keresni” (idézi: Carterette–Friedman, 1974, XV. o.). Ahogy Carterette és Friedman (1974) fogalmaz, Fechner kísérletei óta a pszichofizika empirikus bázisa, mérései, eszközei rendkívüli mértékben bővültek az érzékelés, észlelés és megismerés minden területén. Ami még mindig kétséges, az a mérés elméleti alapja.

A pszichometria, amely tesztek¹² és kérdőíveket használ a lelki jelenségek kutatásában, nem sokkal a pszichofizikát követően jelent meg a 19. század harmadik harmadában. A 19. század egyik neves polihisztorának, Francis Galtonnak a kedvenc mottója Galilei mondásához hasonló volt: „Amikor csak tudsz, számolj!” Galton a mérés átfogó szemléletét vallotta, és úgy vélte, hogy társadalmi fejlődés csak akkor lehetséges, ha az emberi tulajdonságokat kvantitatív módon lehet tanulmányozni és nyomon követni (Briggs, 2022). Galton a tesztek és a kérdőíves felméréseket használó pszichometria – amely a pszichofizika mellett szintén kvantitatív jellegű pszichológiai irányzat – egyik legkorábbi alkalmazója volt. A pszichometria kifejezést elsőként használta és definiálta egy 1879-es, pszichometriai kísérletekről szóló írásában, ami miatt a pszichometria atyjának is nevezik (Briggs, 2022).

1918-ban Edward Thorndike, a tanuláslélektan egyik korai művelője kutatómódszertani hitvallását a természettudományok módszerével megegyezően fogalmazta meg, a következőképpen: „Ami egyáltalán létezik, az valamilyen mennyiségben létezik. Az alapos megismeréshez a mennyiségének és a minőségének ismerete is hozzátartozik. A nevelés az emberben bekövetkező változásokkal foglalkozik; a változás két állapot közötti különbség; e feltételek mindegyikét csak az általa létrehozott termékek – készített dolgok, kimondott szavak, végrehajtott cselekedetek és hasonlók – alapján ismerjük. E termékek bármelyikének mérése azt

¹² A teszt kifejezés is nagyon lényeges jelentésbővülésen esett át és sokkal szélesebb a jelentése, mint a kérdőívnek. A magyarba az angolból került át, oda pedig a franciából. A *testu* eredeti latin jelentése agyagedény volt. Az angol *test* szót, ami főnév és ige is lehet, magyarul többnyire tesztnek vagy próbának, néha vizsgálatnak fordítják, de sok más fordítási lehetőség is van.

jelenti, hogy valamilyen módon meghatározzuk a mennyiségét, hogy a hozzáértő személyek jobban tudják, mekkora az, mint mérés nélkül tudnák. Egy terméket jól mérni azt jelenti, hogy a mennyiségét úgy határozzuk meg, hogy a szakértők bizonyos pontossággal tudják, mekkora, és hogy ezt a tudást kényelmesen rögzíteni és felhasználni lehessen. Ez azoknak az általános hitvallása, akik az elmúlt évtizedben az oktatási termékek mérésének kiterjesztésével és javításával foglalkoztak” (*Thorndike, 1918, 16. o.*).

A pszichológia azonban mégsem fizikai objektumokkal foglalkozik, hanem lelki, érzelmi jelenségekkel, attitűdökkel, véleményekkel és hasonló koncepciókkal, amelyek mérése a fizikai mérésekétől eltérő kérdéseket vet fel. Meg lehet számolni az elszívott cigaretták számát, de a dohányzással kapcsolatos attitűdöt nem. Már a pszichológiai tesztek készítésének legkorábbi szakaszában felvetődött a mértékegység és az egyes mérések közötti viszony (összehasonlíthatóság) problémája.¹³ A fizikai mérésekhez hozzátartozik valamilyen mértékegység, amelynek nagysága önkényes, jellege viszont nem, mert azt a mért jellemző szükségszerűen meghatározza. Egy pszichológiai teszthez viszont egy olyan pontszám tartozik, amelyet csak maga a teszt határoz meg, mértékegység nem társul hozzá. Ezeknek a kérdéseknek és a pszichológiai mérések logikai alapjainak a megtárgyalására a Brit Tudományfejlesztési Társaság (*British Association for Advancement of Science*) 1932-ben létrehozott egy fizikusokból és pszichológusokból álló bizottságot. Vezetője Allan Ferguson fizikus volt, egyik befolyásos tagja pedig a korábban már említett Norman Campbell, szintén fizikus. A bizottság üléseit a Cambridge-i Egyetemen rendezték. Nyolc év ülésezés, levelezés, vita és több köztes publikáció után a végső jelentés bevezetője rögzíti, hogy nincs az a mennyiségű vita, ami lehetővé tenné a bizottság 19 tagja számára, hogy egyöntetű véleményt nyilvánítsanak a vitás kérdésekről. A fizikus résztvevők elutasították a pszichológiai mérés logikai alapjait, a tesztpontszámok összeadhatóságának és a mérések összekapcsolhatóságának a hiányára hivatkozva (*Borsboom, 2005*).

A nézeteltérést annak kifejezéseként is lehet értelmezni, hogy a pszichológiai mérések különböznek a fizikai mérésektől. A kérdőív alapú mérésekkel kapcsolatos tulajdonságok leírásának olyan szókincse és módszertana alakult ki, ami merőben eltér a fizikai mérések tulajdonságainak jellemzésétől: megbízhatóság (re-

¹³ Maga Thorndike is kritikusabban fogalmazott néhány évvel az előző idézete után szerzőtársaival közösen írt könyve első oldalán, az intelligenciatesztek akkori helyzetét áttekintve. A problémák között említette, hogy nem tudjuk, mit mérnek a tesztek, indokolt-e a tesztpontszámokat összeadni, kivonni, szorozni, osztani, arányokat képezni belőlük, és mi a jelentősége az eredményeknek az intellektusra vonatkozóan. Tehát bizonytalan a tartalom, önkényes a mértékegység és bizonytalan a jelentőség (*Thorndike et al., 1926, 1. o.*). Az intelligencia mérésének történetét, módszertani és fogalmi problémáit több magyar könyv is feldolgozta (*Kun-Szegeci, 1978; Horváth, 1991*).

liabilitás, reprodukálhatóság, megismételhetőség, stabilitás, precizitás, konzisztencia), érvényesség (validitás), tárgyilagosság (objektivitás), válaszkészség, értelmezhetőség, valamint ezek különféle jelzőkkel ellátva (pl. interkulturális érvényesség, strukturális érvényesség, tartalmi érvényesség, konstruktvaliditás, kritériumvaliditás, egyidejű és prediktív validitás). A felsorolt jellemzőknek magyarul és angolul is sokféle meghatározásuk és értékelési kritériumuk, jelzőszámuk ismert (magyarul ezekkel kapcsolatban *Horváth, 1997* ad jó összefoglalást).

A mentális konstrukciók mérésével kapcsolatos problémák áthidalására a pszichometria egyes művelői az operacionalizmushoz fordultak segítségért, amelynek megközelítése szerint – a rövidség kedvéért kiemelve és egyszerűsítve a lényegét – egy fogalmat vagy koncepciót a mérése, megfigyelése révén definiálunk és ismerünk meg. Az operacionalizmus kifejezés a fizikában jelent meg, Percy Bridgman 1927-es munkájával, *A modern fizika logikájával (The Logic of Modern Physics)*, de maga a koncepció (vagy ahhoz nagyon hasonló felfogás) már sokkal korábban, más neveken is felbukkant. Így szoros kapcsolatot mutat Galilei kísérleteivel, számszerűsítésre vonatkozó javaslatával, az empirista hagyományával, Charles Sanders Peirce pragmatizmusával és Pierre Duhem instrumentalizmusával, a korabeli tudományfilozófiai mozgalmak közül pedig közel állt a bécsi kör logikai pozitivizmusához. De míg Bridgman fizikusként a fizika ismeretelméletével foglalkozott, addig a bécsi kör¹⁴ tagjai általánosabb, a tudomány egésze számára alkalmazandó előírásokat igyekeztek megfogalmazni.

A pszichometrián belül az operacionalizmus már magának a kifejezésnek a megjelenése előtt jelen volt és gyakorlattá vált. Edwin Boring az intelligencia mérése kapcsán 1923-ban írta le azt, amit későbbi elnevezéssel élve az operacionalizmus nézőpontjának, összefoglalásának lehet tekinteni: „az intelligencia az, amit az intelligenciatesztek mérnek”.¹⁵ Erre sokan hivatkoztak azóta, bírálva úgy is, hogy „ma már csak ironikus szellemeskedésnek tekinthetjük” (*Kun–Szegedi, 1978, 11. o.*), de ennek ellenére tökéletesen kifejezi az operacionalizmus gyakorlatiasságát. Az operacionalizmus megközelítésének általánosabb megfogalmazása úgy hangzik, hogy „X tulajdonság” az, amit „T teszt” mér. Ezt hiba lenne az önkényesség engedélyezéseként értelmezni. A tesztek összeállítását a korábbi definíciók és a szakmai konszenzus alapján kell elvégezni (*Vessonen, 2019*). Ez a problémakör felmerül az olyan, a gazdaságelméletben és a szociológiában szerepet játszó koncepciók mérése során is, amelyek közvetlenül nem megfigyelhetők, mérhetők (például az egyénekre és csoportokra is értelmezett műveltség, tudás, emberi tőke, vállalkozói hajlam, bizalom, amelyek operacionalizálása sokszor egészen furcsa és vitatható változók révén történik).

¹⁴ A bécsi körhöz tartozó Nagel mérésdefinícióját a mérés modern jelentését tárgyalva ismertettem.

¹⁵ „Intelligence is what the tests test.” Megjelent 1923-ban, az *Intelligence as the tests test it* című írásban (*Boring, 1923*).

1931-ben Burrhus Frederick Skinner doktori értekezésében hivatkozott Bridgman operacionalizmusára (Rogers, 1989). Boring, valamint tanítványa, Stanley Smith Stevens (Bridgman, Skinner, Boring és Stevens is a Harvard Egyetem munkatársai voltak) sokat tett az operacionalizmus mint fogalom és ideológia pszichológián belüli terjesztéséért. Stevens 1935-ös, *A pszichológia operacionális alapja* című munkájában Bridgmanra hivatkozva próbálja az operacionalizmus révén megoldani, túlhaladni azokat a vitákat, amelyek a pszichológiai koncepciók tartalma, értelme, világossága körül zajlottak. A fogalmak jelentését bizonyos műveletek (operációk) révén pontosíthatjuk. Stevens nem részletezi a műveletek mibenlétét, és más szerzők is sokszor hallgatnak erről, vagy valamilyen szinonimát használnak.¹⁶ Egy másik, közel egyidejű tanulmányában az operacionalista módszereket a hangtani jellemzők tanulmányozásával kapcsolatban mutatja be, ami legfőbb alkalmazott vizsgálódási területe volt (Stevens, 1935b). Az operacionalista módszerek a gyakorlat alapján valamilyen megfigyelhető jellemző, a pszichofizikai kísérletek, a tesztekkel folytatott mérések és egy közvetlenül nem mérhető tulajdonságnak valamilyen mérhető tulajdonsággal történő helyettesítései lehetnek. Nem arról van szó, hogy minden egyes fogalom használatakor egyfolytában műveleteket kell végezni azok meghatározására, hanem arról, hogy amikor tudatos erőfeszítést tesznek egy fontos pszichológiai fogalom meghatározására, pontosítására, akkor kell szükségszerűen megfelelő műveletekhez folyamodni (Stevens, 1935a; Leahey, 1980; Feest, 2005).

4. A mérési skálák és a mérés Stevens-féle definíciója

Az eddig bemutatott előzmények megalapozták a mérésnek az elmúlt évtizedek statisztikatankönyveiben, -kézikönyveiben található leggyakoribb meghatározását. Ezek akár magyar, akár más nyelvűek, a statisztikai fogalmak bevezető tárgyalása kapcsán a mérésnek nem egy hétköznapi vagy természettudományos eredetű meghatározását ismertetik, hanem a következő definícióhoz hasonlót: a mérés számok meghatározott szabályok szerinti hozzárendelése a megfigyelési egységekhez. A konkrét szövegezés Stevens 1946-os definíciójának (amely Stevens ki-

¹⁶ Az operacionalizmus mibenlétét definiálhatják példákkal, az elvégzendő műveletek megnevezésével, a megfigyelések és műveletek megnevezésével.

fejezésével élve Campbell definíciójának a parafrázisa)¹⁷ valamilyen változata. Ezek közelebb állnak a reprezentációs elmélethez, mint az operacionálizmushoz (Stevens megközelítését mindkét felfogáshoz sorolták), a mérés hagyományos értelemezésétől pedig mindenképpen nagyon távoliak. Például: „Legáltalánosabb értelemben a mérés azt jelenti, hogy dolgokhoz szimbólumokat rendelünk. Így minden rendezés mérésnek fogható fel” (*Ferge–Schnell, 1975, 316. o.*). „A mérés során bizonyos hozzárendelési szabályok alapján szimbólumokat, számokat rendelünk dolgokhoz, tulajdonságokhoz. Ezek a hozzárendelési szabályok határozzák meg a mérési skálákat” (*Kérékgyártó–Mundruczó, 1991, 31. o.*). „A mérés számok meghatározott szabályok szerinti hozzárendelése jelenségekhez (dolgokhoz, tárgyakhoz, eseményekhez), illetve ezek bizonyos tulajdonságaihoz” (*Korpás, 1996, 11. o.*). „Bizonyos szabályok betartása mellett bármely, eredetileg nem mennyiségi ismérv lehetséges változatai számértékké alakíthatók, »kódolhatók«. Ilyen alapon a sokasági egységek bármely tulajdonságának észlelése és szám formájában történő rögzítése az egységek számokkal való jellemzésének, azaz mérésnek tekinthető” (*Hunyadi–Vita, 2008, 24. o.*). „Mérésnek nevezzük a változó numerikus értékeinek a változóhoz történő hozzárendelését” (*Benyovszki, 2013, 12. o.*).

Ehhez a definícióhoz társul Stevens négy mérési skálájának az ismertetése:

1. *Nominális mérési szint.* Más terminológia szerint névleges mérési szint. Megfelelne a minőségi változóknak, kategorikus változóknak, ha a megfigyelési egységek kódolását nem sorolnák ide. Stevens alapján a kódolás akkor mérés, ha szabályok szerint történik, és nem véletlenül. A szabály azonban csak annyi, amely által a számok megfigyelési egységekhez való hozzárendelése nem véletlen lesz, miszerint nem szabad azonos számot különböző osztályokhoz és különböző számot azonos osztályokhoz rendelni. Az osztályok lehetnek egyeleműek is, mint például a labdarúgók azonosítása egy számmal, amely a mezükön is olvasható (ez Stevens gyakori példája, sőt a legelső példája a mérésre az 1946-os tanulmányban, és szerepel az 1951-es, 1959-es, 1968-as és 1975-ös tanulmányaiban is; a legutolsó tanulmány már a halála után jelent meg). Stevens nyomán a tankönyvek rendszerint a kódolást is mérésnek fogják fel, olyan jellemző példákkal, mint az autók rendszámablaja, az irányítószám vagy a bankszámlaszám. Egyes tan-

¹⁷ Stevens meghatározása így hangzik angolul: „we may say that measurement, in the broadest sense, is defined as the assignment of numerals to objects or events according to rules” (*Stevens, 1946, 677. o.*). Ugyanezt a definíciót Stevens később sokszor megismételte írásaiban, így az 1951-es kísérleti pszichológiai kézikönyvében, amely rendkívül nagy hatásúvá vált. *Michell (1997b)* 44 könyv (elsősorban pszichometriai, kisebb részben általános és társadalomstatistikai tankönyv, kézikönyv) mérésdefinícióját vizsgálta meg. Ezekből 39 adott Stevens meghatározásához hasonló szerkezetűt, eltérő fogalmakkal. Például „számok” helyett szerepelhetnek a meghatározásban szimbólumok, numerikus értékek, pontszámok. Hagyományos mérésdefiníciót viszont egy könyv sem adott.

- könyvek példáiban csak kategorikus változók (minőségi ismérvek) szerepelnek. Más tankönyvekben csak kódok, azt állítva vagy sugallva, hogy a nominális mérési szint csak az elemek azonosítására szolgál.
2. *Ordinális mérési szint.* Más terminológia szerint sorrendi skála. A megfigyelési egységek között értelmezhető a kisebb-nagyobb reláció is.
 3. *Intervallumskála.* Nullapontja önkényes, ezért az ilyen skálán mért értékek arányai nem értelmezettek, csak a különbségek közötti arányok.
 4. *Arányskála.* Természetes nullapontja van.

A további, magyarul ritkábban előforduló elnevezéseket az 5. táblázatban foglaltam össze. Ezek mindegyike nyomtatásban megjelent munkákból származik, a zérópont kivételével, amit én fordítottam (angolul *zero-point*). Az arányskálára alkalmazott, racionális skála megnevezés egy rossz fordítás eredménye (*Benesch, 1994, 47. o.*). A táblázatban szerepeltetem az abszolút skálát is, amellyel Stevens 1946-ban, mint az arányskála egyik esetével röviden foglalkozott, és amellyel egyes tankönyvek még kiegészítik a négy skálát. Ez a diszkrét egységek megszámlálásához használatos (Stevens példái a tojások, pennyk és almák). Az abszolút skála azonban a kategorikus változókhoz is kötődik, az egyes kategóriákba tartozó elemek darabszámához.

5. táblázat

A mérési skálákra (mérési szintekre) használt elnevezések és viszonyuk a mérés hagyományos felfogásához

Terms used for measurement scales (levels) and their relation to the traditional concept of measurement

Mérési skála	Egyéb elnevezések	A hagyományos méréshez való viszony
Nominális	Névleges, nevesítő	Kódolás, kategorizálás, klasszifikáció, osztályba rendezés, osztályozás
Ordinális	Sorrendi, rendező, rendezéssel, rangsoroló	Sorba rendezés
Intervallum	Értékköz, különbségi, távköz	Hagyományos mérés
Arány	Hányados, viszony, proporcionális, racionális, zérópont	Hagyományos és klasszikus mérés
Abszolút	Kardinális	Számlálás

Forrás: saját szerkesztés.

Nominális és ordinális szinten a mérés eredményét nemcsak számok, hanem szimbólumok, elnevezések is jelölhetik (például az eper, autó, katicabogár, mint a gyerekek óvodai jelei; a platina, arany, gyémánt, prémium, mint a termékek kategóriái stb.). Az 5. táblázat hagyományos méréshez való viszony oszlopát azzal egészítettem ki, hogy a nominális és az ordinális szint hagyományosan nem mérés,

de a hagyományos mérésnek lehet az a célja, hogy a megfigyelési egységeket sorrendbe vagy nagyság szerinti kategóriákba, osztályokba rendezze. Azt viszont nem mondhatnánk a hagyományos mérési fogalom szerint, hogy méréssel állapítjuk meg egy ember nemét, iskolai végzettségét, hajának színét.

Stevens (1957) az inger nagysága és az érzés nagysága közötti kapcsolatot vizsgálva a négy mérési skála mellé (vagy az abszolút skálával együtt öt mellé) javasolta a logaritmikus skála bevezetését. Ezt azzal indokolta, hogy az inger és az érzés között nem lineáris a viszony, hanem az inger nagyságának valamilyen hatványa áll kapcsolatban az érzés nagyságával. A logaritmikus skála elnevezéséhez *Stevens (1957)* jónak tartotta volna az arányskála terminust, de az már foglalt és elterjedt volt. Ennek kapcsán írta, hogy az arányskála nevezhető lenne zérópontskálának is (*Stevens, 1957, 177. o.*). Stevens logaritmikus skáláról írt javaslata két okból is bírálható. Egyrészt, logaritmikus skálákat nem a mérés műveletére használnak – azzal analóg módon, ahogyan intervallumskálát szükségszerűen kell használni a földrajzi méréseknél (önkéntes nullapont, önkényes irány) vagy az időszámításnál (önkéntes nullapont, az irány nem önkényes) –, hanem a megmért eredmények transzformált közlésére néhány olyan esetben, amikor több nagyságrendnyi különbség jellemző az adatokra. Gyakran emlegetett példák erre a földrengések erősségének megadása a Richter-skálán, a savasság mértékének közlése a pH-értékkel, a hangerő megadása decibelben vagy a csillagok fényességéé magnitúdóban. A logaritmikus skálák előnye az, hogy nem kell egymástól nagyon eltérő nagyságrendű számokat használni. Ez azonban elég kicsi előny, hiszen végső soron az eredeti nagyságrend az, amiben helyesen lehet értelmezni az adatokat, ezt pedig legközvetlenebbül az eredeti nagyságrendben közölt adatok biztosítják. A skála hátránya, hogy a laikusok számára a skála kis különbségei megtévesztőek lehetnek, lebecsülhetik a különbségek jelentőségét: tízes alapú logaritmus 4 és 6 között a különbség nem másfélszeres, hanem százszoros, és 4 és 4,3 közötti különbség sem nüansznyszerű, hanem csaknem duplája a 4,3 a 4-nek. A logaritmikus skála tehát létezik, használatos, de nem a mérésre, hanem az eredmények közlésére. Másrészt, amit Stevens a mérések kapcsán végzett, az nem mérési skála létrehozása, mért eredmények átskálázása logaritmizálás segítségével, hanem függvényillesztés a tapasztalati adatokra, amely azzal az eredménnyel járt, hogy a lineáris regressziónál a hatványkitevős regresszió pontosabban írja le a megfigyeléseket. Ráadásul pszichológiai törvénynek nevezi azokat az egyedi megfigyelésekből kapott függvényillesztéseket, amelyek valójában az egyének egymástól eltérő érzékeléseit becsülik, közel sem determinisztikusan. A hallás- és látásérzékelés vizsgálatában például ugyanazon ember jobb és bal füle, szeme is eltérő eredményeket adhat.

A mérési skálákat mérési szintnek is nevezik. A mérési szint annyiból jobb elnevezés a Stevens által eredetileg használt mérési skálánál, hogy az utóbbi kifejezésben nemcsak a mérés, hanem a skála is egy merőben új és indokolatlan jelentéssel lett ellátva. A mérés kontextusában eredetileg a skála egy mértékegységet vagy egy mennyiséget jelentett, tehát a mérés hagyományos jelentéséhez kötődött. Ebben az új megfogalmazásban pedig nagyjából a típus, forma, fajta jelentést vette fel, a mérési skála kifejezés pedig az adattípus kifejezésnek felel meg. További terminológiai sajátosság, hogy Stevens a „mérési skálák elmélete” címet adta leírásának, amely inkább egy adattípus leírás, egy klasszifikáció, és nem elméletet fogalmaz meg, még akkor sem, ha a skálák matematikai tulajdonságainak leírása kiegészül az egyes mérési szinteken megengedett statisztikai eljárások, módszerek leírásával, illetve az alkalmazásukra vonatkozó előírásokkal. A Stevenstől eredő „mérési skálák elmélete” kifejezés máig megmaradt és széles körben használatos, amihez a saját szóhasználatomnak is igazodnia kell, de ennek ellenére a szóösszetétel mindhárom eleme vitatható.

A mérési skálákhoz (szintekhez) Stevens (1951) részben meghatározta, részben előírta, hogy az adott mérési szintű számokkal milyen műveletek értelmezhetők és végezhetők el. Ez később nagyon sok vita forrása lett, amely során nem mindig különböztették meg, hogy az egyes előírások közül melyikre gondolnak, mindegyikre, vagy csak egy részükre. A 6. táblázatban látható lehetőségek közül a legkisebb jelentősége a megengedett skálatranszformációknak van, hiszen ez ritka és alkalmazásakor eléggé egyértelmű. Ez inkább értelmezési szempontból lehet érdekes a kompozit mutatóknál, és minden olyan esetben, ahol transzformált adatokkal történnek a számítások, de formális értelemben ez kevésbé problémás terület.

6. táblázat

Az egyes mérési skálákon (szinteken) elvégezhető matematikai és statisztikai műveletek
Mathematical and statistical operations that can be performed on measurement scales (levels)

	Az eljárások, műveletek köre	Példák a megengedett és nem megengedett műveletekre
1.	Alapvető műveletek	Egyenlőség, különbség, arány
2.	Megengedett skálatranszformációk	Lineáris, logaritmikus
3.	Egyszerű leíró statisztikai eljárások	Módusz, medián, számtani átlag, szórás
4.	Összetett statisztikai eljárások	Asszociáció, varianciaanalízis, korreláció és a sokváltozós eljárások
5.	Következtető statisztikai eljárások	Paraméteres és nem paraméteres próbák

Forrás: saját szerkesztés.

A mérési skálákhoz (szintekhez) Stevens olyan értelmezést adott, amely szerint a szintek között nemcsak különbség van az elvégezhető matematikai műveletekben és eljárásokban, hanem ez a különbség egyben minőségi sorrendet, más szóval fejlettséget, információkülönbséget is jelent. Stevens szerint a mérés során törekedni kell a magasabb mérési szint elérésére, bár ennek kivitelezése nem mindig lehetséges. A legtöbb információt nyújtó, legfejlettebb skála az arányskála, a nominális pedig a legfejletlenebb. Ez az értelmezés annyira elterjedt, hogy valamilyen formában az összes, a mérés definíciója kapcsán hivatkozott tankönyvben szerepel. Azonban „a legtöbb információt nyújtó skála az arányskála” állítás legfeljebb azonos jelenségek eltérő mérési skálán történő méréseinél értelmezhető, de problémásnak tartom az olyan típusú összehasonlításoknál, mint mondjuk az emberek hajszíne (nominális skála) és az emberek testsúlya (amit jellemzően arányskálán adnak meg).

A hőmérséklet koncepciójának és mérésének fejlődése jól mutatja, hogy mikor lehet érvényes a nagyobb információtartalom kitétel. Az arisztotelészi koncepció minőségi volt, két kategóriával. A 2. században élt orvos és filozófus, Galénosz a hideg és meleg kategóriát is négy fokozatra osztotta fel, amelyet az orvos érzékelése alapján tudott megkülönböztetni. Ez a sorrendi felfogás azonban nem terjedt el annyira, mint az arisztotelészi (*Barnett, 1956*). A hőmérséklet mennyiségi koncepciójának felfedezésére volt szükség az intervallumszintű mérést lehetővé tevő hőmérők elkészítéséhez, ami a 16. század végére következett be. A 18. századig több mint 30 hőmérsékleti skálát javasoltak, más és más kitüntetett pontokkal és a pontok közötti intervallumokkal, amelyek közül hármat használtak a leginkább, a Fahrenheitről, a Réaumurról és a Celsiusról elnevezett skálát. Celsius skálája eredetileg fordított volt, 0 a víz forráspontjaként, 100 a fagyáspontjaként volt meghatározva (*Hay, 2020*). Végül a termodinamikai alapú hőmérsékletfelfogással Kelvin 1848-ban megalkotta a később róla elnevezett Kelvin-skálát, ami arányskálaként a fejlődés csúcspontja, ennek ellenére a mai napig nem szorította ki a Celsius- és a Fahrenheit-skálákat se a mindennapokból, se a tudományos mérésekből. Bizonyos esetekben pedig a hőmérséklet dichotóm formájú közlése is, bár információvesztéssel jár, de elég lehet, ha van valamilyen jelentős küszöbérték. Ilyen eset például, amikor a víz halmazállapotára vagyunk kíváncsiak, vagy a klinikai kutatásokban a nem lázas vagy lázas kategorizálás is elég informatív lehet.

A mérési skálák és a változók típusaira használt egyes elnevezések közötti kapcsolatot a 7. táblázatban foglaltam össze. Ezek olyan elnevezések, amelyekkel a magyar szakirodalomban találkozhatunk. Ha maguk az elnevezések esetleg bírálhatók is az egyes szavakra ráerakodott sokféle jelentés miatt, a nagyobb problémát az jelenti, ha nem következetesen használják őket.

7. táblázat

A mérési skálák (szintek) és a változók elnevezései közötti kapcsolat
Relationship between measurement scales (levels) and variable names

Mérési skála (szint)	Elnevezések a változókra
Nominális (és ordinális)	Minőségi, kvalitatív, kategorikus
Intervallum és arány (és ordinális)	Mennyiségi, kvantitatív, metrikus, numerikus, folytonos

Forrás: saját szerkesztés.

A Stevens-féle mérési skálák, valamint a minőségi és mennyiségi kategorizálás négy fontos jellemzőben különbözik egymástól:

1. A nominális szint a minőségi ismérveken kívül az elnevezéseket és a kódolásra, azonosításra használt változókat is tartalmazza. Az elnevezés és az azonosító között az a különbség, hogy míg az előbbi tartozhat több megfigyelési egységhez is (pl. az emberek személynevei), az utóbbi minden megfigyelési egységnél különböző. A minőségi ismérvek, kategorikus változók közé többnyire nem szokták besorolni az egyedi azonosítókat, de az is elképzelhető, hogy valamikor valaki már besorolta őket közéjük. Ugyanakkor, ha az egyedi azonosítókat valamilyen sorrend alapján képezik (például egyes szervezetekben a belépés időpontja alapján adhatnak egyéni személyi azonosítókat), akkor az ilyen azonosítók egyszerre számok és egyben sorrendi változók is, amelyek már értelmesen elemezhetők matematikai műveletekkel. De ebben az esetben a kódolás a megfigyelési egység valamilyen tulajdonságához kötődik, és nem önkényes.
2. A mérési skálák megkülönböztetik az ordinális változókat, a minőségi/mennyiségi dichotómia nem. Ugyanakkor az ordinális változók egy része minőségi változó jellegű (ezek a rangsorolt osztályok és fokozatok), más része mennyiségi változó jellegű (nagyság szerint sorrendbe rendezett egységek, holtverseny nélkül). A problémát az okozza, hogy a minőségi kategóriák (típusok, csoportok) nagyon eltérők, néhány fontosabb lehetőséget megemlítve:
 - nem rangsorolható minőségek (pl. emberek neme, anyanyelve),
 - nem rangsorolható csoportosított minőségek (pl. emberek anyanyelve nyelvcsaládok szerint),
 - gyengén rangsorolható minőségek (pl. emberek végzettsége, ahol a sokféle lehetséges kategóriát nem minden esetben lehet egyértelműen sorrendbe állítani),
 - elképzelt minőségek (pl. foglalkozások presztízse, autómárkák presztízse),
 - jól rangsorolható minőségek (pl. a gyémántok minősége),

- minőségi besorolások (pl. szállodák minősítése, építőanyagok minőségi osztályai, hitelminősítések, energiatanúsítványok),
 - több változóból képzett csoportok (ezek más adatokból származtatott mutatók),
 - egyértelműen rangsorolható, szabályozáson alapuló hierarchiák (pl. rendfokozatok),
 - mennyiségeket tartalmazó minőségek (pl. a ruhák méretszámozása).
3. A mérési skálák annyiban részletesebb a mennyiségi ismérvek leírásában, hogy megkülönbözteti az intervallum- és az arányskálát. Ez fontos előrelépés, mert az önkényes nullaponttal rendelkező skáláknál az arányok számítása hibaforrás. Ezt leszámítva az alkalmazható elemzési módszerekben nincs nagy különbség az intervallum- és az arányskálátípusú adatok között.
4. A mennyiségi ismérv megnevezés annyiban tágabb az intervallum- és arányskálaszintű mérésnél, hogy mennyiségi ismérvek lehetnek
- megoszlási és intenzitási viszonyszámok,
 - aggregátumok,
 - kompozit mutatók,
 - több változóból képzett indexek,
 - különböző módszerekkel képzett pontszámok,
 - együtthatók és egyéb mutatók.

Mivel ezeket más számokból képezték, és maguk is számok, a gyakorlat többnyire ezeket is intervallum- vagy arányskálájú adatokként kezeli. Ez azonban nem korrekt, mert a mérési szintek egy változóra értelmezhetők, nem pedig több változóból különféle transzformációk segítségével képzett összetett adatokra, amelyekre nem érvényesek a Stevens által leírt matematikai skálatulajdonságok.

A változók egy sajátos kategóriáját alkotják a két ismérvváltozattal rendelkező változók, amelyeket hívnak dichotómnak, binárisnak, kétértékűnek és alternatív ismérveknek is. Ezek lehetnek természetes módon adódó kettőségek és mesterségesen képzettek. Elemzésükre – mérési szintjüktől és minőségi vagy mennyiségi jellegüktől függetlenül – léteznek egységesen használható módszerek. A kettőnél több, de kevés ismérvváltozattal rendelkező diszkrét mennyiségi ismérvek esetén pedig haszonnal alkalmazhatók a kategorikus változóknál használt eljárások.

5. A mérési skálák kritikái és néhány egyéb tipizálás

Az eddig tárgyaltakból már látszik, hogy a mérési skálák Stevens-féle leírásának vannak érdemei a koncepciók tisztázásában, ahogyan problémái is. Nem célom annak elkülönítése, hogy mit mondott Stevens, és mit mondtak a rá egyetértően hivatkozó szerzők. Nem foglalkozom részletesen azzal sem, mennyire volt koherens a mondandója egy-egy tanulmányon belül és az alatt a harminc év alatt, amíg a méréselméletről publikált. Vannak arra utaló jelek, hogy ő maga nem kezelte dogmatikusan a saját rendszerét, több helyen is írt arról, hogy nem kell szigorúan venni az előírásait, amelyet pragmatikus szempontok felülírhatnak. 1946-os tanulmányában a sorrendi statisztikákkal kapcsolatban, amelyeknek az átlagolását egyébként tiltotta, megengedő volt: „A legszigorúbban véve szokásos mutatókat, beleértve az átlagokat és a szórásokat nem szabadna használni ezekkel a skálákkal, mivel ezek a statisztikák az adatok relatív rangsorolásánál több ismeretet feltételeznek. Másrészt, erre az »illegális« statisztikai számításra lehet hivatkozni egyfajta pragmatikus értelemben: Számos esetben gyümölcsöző eredményekhez vezet. Bár ennek az eljárásnak a betiltása valószínűleg nem szolgálna jó célt, helyénvaló rámutatni arra, hogy az ordinális skálán számított átlagok és szórások annyiban hibásak, amennyiben a skála egymást követő intervallumai egyenlőtlen nagyságúak. Ha csak az adatok sorrendjét ismerjük, óvatosan kell eljárunk a statisztikáinkkal, és különösen a belőlük levont következtetésekkel” (Stevens, 1946, 679. o.). Ugyanitt a nominális skála mérésnek nevezésével kapcsolatban megelőlegezte azt a kritikát, hogy sokan lesznek, akik azt fogják állítani, hogy abszurd dolog a számjegyek minőségi ismérvváltozatokhoz való hozzárendelésének folyamatához a mérés fogalmából fakadó méltóságot tulajdonítani (Stevens, 1946, 679. o.).

Az elkövetkező (és az eddigi) kritikák tehát egy szigorú, kivételeket nem ismerő rendszer egészének jellemző tételeire vonatkoznak, amelyeket Stevens számos tanulmányában és ennek nyomán mások fogalmaztak meg. Ezek röviden úgy foglalhatók össze, hogy

- kibővítve a mérés definícióját, koncepcionálisan nagyon eltérő műveleteket írt le egyetlen fogalommal,
- a mérési szintek merev kategóriáit hozta létre, amelyeket helyénvalóbb lett volna adattípusoknak neveznie,
- a kategóriákhoz rendelte a megengedhető műveleteket és statisztikai eljárásokat,
- teljes körűnek szánta, de a tényleges adatok egy része kategorizálhatatlan, más része nem egyértelműen kategorizálható.

8. táblázat

**A Stevens-féle mérési skálákkal kapcsolatos fontosabb problémák
(átfedést tartalmaznak)**
Main problems of Stevens' measurement scales (they contain overlaps)

	Megnevezés	Példa, rövid leírás, megjegyzés
1.	Terminológiai zavarkeltés	Hozzájárul a mérés jelentésének túlzott kibővítéséhez. Maga a definíció nem egyértelmű. A skála kifejezést az eredeti jelentéstől eltérően használja.
2.	A formális rész és a gyakorlat nem egyeztethető össze.	Az operacionalizálás során sokszor adódik, hogy nem lehet egyértelműen eldönteni, milyen „skálájú” az adat.
3.	A gyakorlati mérés mechanisztikus és leegyszerűsített képét sugallja.	A mérési eljárás megtervezése előfeltételezi az elméleti és konceptuális háttér tisztázását, megértését.
4.	A következtető statisztikai módszerek alkalmazásával kapcsolatos zavarhoz vezet.	Alapvetőbb az a kérdés, hogy véletlen-e a minta. Ha igen, akkor lehet érdekes a következtető statisztika.
5.	Bizonyos adattípusok nem vagy nem egyértelműen helyezhetők el.	Ilyen adattípusok: értékbeli aggregátumok, tesztpontszámok, sportágak világranglistapontszámai, pontozásos sportágak pontszámai, több változóból képzett kompozit indikátorok és mutatók, Likert-skála.
6.	A nominális szint heterogén.	A nominális szint jelenthet kódolást és minőségi ismérvet is. Nem indokolt a két kategória egybemosása és nem indokolt a kódolás mérésnek nevezése. Minőségi ismérvekre vonatkozóan indokoltabb az osztályozás kifejezés használata, mint a mérésé.
7.	A mérési szintek hiányosak, nem fedik le a lehetséges egydimenziós eseteket.	Bizonyos (igaz, ritkábban előforduló) adattípusok kívül maradnak a kategóriáin (mérési szög, ciklikus adatok).
8.	A magasabb mérési szint nem mindig előnyösebb.	A hőmérséklet mérése sokkal elterjedtebb Celsius- és Fahrenheit-skálán, mint Kelvin-skálán.
9.	Korlátozza az elvégezhető műveleteket.	A gyakorlatban kategorikus és sorrendi változók mennyiségi ismérvvé átalakítva hatékonyan bekapcsolhatók többváltozós mennyiségi módszerekkel végzett elemzésekbe.
10.	A Likert-skála nem helyezhető el rajta.	Ezt külön alfejezetben részletezem.

Forrás: saját szerkesztés.

Ezek a problémák kezelhetők lennének a merevség oldásával (ahogyan már utaltam rá, magának Stevensnek is vannak erre utaló megjegyzései), az indokolatlan szabályok ajánlásként való kezelésével és további kategóriák létrehozásával. A felsorolt problémákon túl számos további kritikus kérdést vetettek fel ezzel az előírásokat (és nem pusztán leírásokat) tartalmazó rendszerrel kapcsolatban, már rögtön a megfogalmazásakor is, csakúgy, mint az azóta eltelt évtizedekben. A 8. táblázatban látható bővebb összefoglalóban a kategorizálás nehézsége miatt lehetnek átfedések. A megfogalmazott kritikákkal azért érdemes foglalkozni, mert kialakult egy olyan statisztikai gyakorlat, rutin, amely vagy nem, vagy nem kellő mértékben veszi figyelembe azokat. Néhány kérdést már érintettem az előzőekben, most a mérési skálák eltérő aspektusaival foglalkozó, de átfogóbb bírálat ismertetése révén térek ki egyes további részletekre és indokokra.

Velleman és Wilkinson (1993) elsősorban az addigi kritikákat rendszerezte és tekintette át, nem vetettek fel új szempontokat. Szerintük a Stevens rendszerével kapcsolatos bírálatok három pontra összpontosulnak:

1. Az alkalmazható statisztikai módszerek korlátozása.
2. A taxonómia túl szigorú ahhoz, hogy a való világ adataira alkalmazható lehessen.
3. Indokolatlanul korlátozza a paraméteres becslési eljárások alkalmazását, nem paraméteres próbákat írva elő.

A 2. pontot már tárgyaltam. Mivel az 1. és a 3. pont is statisztikai módszerek korlátozásával kapcsolatos, érdemes itt egy fontos következménnyel járó különbséget tenni, ami a 6. táblázatban látható, és amit Velleman és Wilkinson nem ír le expliciten, ahogyan az általuk hivatkozott szerzők többsége sem. Léteznek egyrészt olyan leíró statisztikai eljárások, módszerek, amelyek Stevens előírásai ellenére hatékonyan, problémamentesen használhatók az elemzésben (néhány példát később említek). Másrészt léteznek következtető statisztikai eljárások, amelyek alkalmazásának csak az egyik feltétele az adatok mérési szintje és eloszlása. Ez a feltétel sokszor nem egzaktul vizsgálható, maga is valószínűségi jellegű, és függhet a sokaság elemszámától is. A következtető statisztika használatának hátterében van azonban egy gyakran figyelmen kívül hagyott szempont, nevezetesen az adatok keletkezési körülményei. Ha azok valamilyen véletlen (vagy véletlenhez közeli) mechanizmussal keletkeznek, akkor a valószínűségszámításon alapuló következtető statisztikai módszerek alkalmazhatók. Véletlen mechanizmus kétféleképpen lehetséges. Vagy magának a vizsgálat tárgyának olyanok a sajátosságai, ami miatt a vizsgált jellemzőnek az értéke véletlen lesz, vagy a vizsgált jellemző értéke nem véletlen, de a vizsgált mintaelemek kiválasztási mechanizmusába mesterségesen be van építve egy véletlen mechanizmus. A társadalmi, gazdasági adatok többségénél ezek a feltételek nem teljesülnek. Itt többnyire egy adott helyre és időre vonatkozó szituációt, történeti helyzetet lehet leírni az adatok segítségével,

felhasználva referenciának más hasonló, de nem azonos helyzetben nyert eredményeket. Tehát nem megismételhető megfigyelések, csak egymáshoz hasonló megfigyelések zajlanak, amelyek egyedi helyzeteket mérnek fel és értékelnek az adatok segítségével.¹⁸

A használható módszerekkel kapcsolatos diskurzus egy része egy olyan mellékvágány, amelynek nem lenne nagy gyakorlati jelentősége a következtető statisztika megfelelő, indokolt használata esetén. Frederic Lord (aki pszichológus és pszichometriával foglalkozó statisztikus volt) Stevensszel szemben megfogalmazott korai kritikája érdekes példa, mert egyszerre van érvényes üzenete és problematikus oldala is. *Lord (1953)* rövid írásában egy szatirikusan előadott szemléltető példán keresztül fogalmazta meg az elvégezhető statisztikai műveletek korlátozására vonatkozó bírálatát, amely röviden összefoglalva így szólt. Egy pszichometriával foglalkozó professzornak a nyugdíjazását követően, gondos munkájának elismeréseként az egyetem futbalszám-koncessziót adott, valamint nagy mennyiségű szövetszámot és egy automatát, hogy eladhassa azokat. Összesen 100 trillió két számjegyű szövetszámmal kellett kezdenie, amelyek gyakorisági eloszlását feljegyezte. A számok sorrendje és kiosztásának módja véletlen volt. Először a másodéveseknek értékesített a mezekből, majd 1600 elsőéves hallgatónak. Az elsőéves hallgatók panaszkodtak, hogy a kapott számok túl alacsonyok voltak, ami miatt a másodévesek kinevetik őket. Az elsőévesek mindannyian azt támogatták, hogy a másodéveseket egyenként kizavarják az ágyukból, és a folyóba dobják őket. A professzor megrémült ettől, és kérte az elsős hallgatókat, várjanak addig, amíg konzultál egy statisztikussal, mert talán csak véletlenül kaptak alacsony számokat. A megkérdezett statisztikus paraméteres próbával összehasonlította a sokasági átlagot és a mintaátlagot, miközben Stevens által tiltott műveletek sokaságát végezte el a professzor nagy megrökönyödésére. A professzor ugyanis úgy tudta Stevenstől, és korábban azt tanította a diákjainak, hogy a mezszámok csak nominális mérési szintű azonosítók, így nem lehetne velük matematikai műveleteket végezni, összeadni, osztani, négyzetre emelni, gyököt vonni. Azt még az olyan ordinális szintű változókkal sem szabadna, mint a tesztpontszámok. A statisztikus megnyugtatta, hogy mindezt szabad, és be is bizonyította, hogy a számolásoknak van értelme. A műveletek elvégzése miatt aggódó professzornak („De hát nem lehet futbalszámokat szorozni – jajgatott a professzor. – Azok még csak nem is sorszámok, mint a tesztpontszámok”) ezzel a híressé vált mondattal válaszolt a statisztikus: „Mivel a számok nem emlékeznek arra, hogy honnan jöttek, mindig ugyanúgy viselkednek” (*Lord, 1953, 751. o.*).

¹⁸ További részleteket tárgyaltam a területi statisztikai adatok jellege (*Dusek, 2006*) és a valószínűségszámítás közgazdasági alkalmazása kapcsán (*Dusek, 2013*). A kérdés kapcsolódik a szignifikanciavizsgálatok helytelen gyakorlatához is (*Bárdits–Németh–Terplán, 2016*).

Az adott példának az elrendezése annyira speciális volt, hogy a megszámok értelmezhetők voltak nominális és intervallumskálán is. Így a statisztikusnak igaza volt, indokoltan végezte el a műveleteket, amellyel a megszámok kiadásának véletlen voltát megfelelően tesztelte. Ha a példázat általános érvényét nézzük, akkor azonban árnyaltan kell fogalmazni, mert mindkét félnek (Stevens és Lord) részben igaza lehet, viszont teljesen általánosítva az állításokat, egyiküknek sem. Lord példája megmutatta, hogy az adatok, mérési eredmények körülményeinek ismeretében lehet dönteni az értelmesen elvégezhető, tartalmi jelentéssel bíró műveletekről. Ezek hiányában fennáll a lehetősége, hogy a számokkal végzett műveletek pusztán matematikán belül értelmezhető manipulációk lesznek, amelyek csak a számok tulajdonságait mutatják meg, és az eredmények nem köthetők a valósághoz. Lord elemzése a megszámokról szólt, és nem a diákokról. Lord mondása csak azzal a kiegészítéssel fogadható el, hogy bár a számok nem emlékeznek arra, honnan jöttek, azt a vizsgálat végzőjének tudnia kell, hiszen csak így tudja őket nemcsak számtanilag értelmezni, hanem az elemzett helyzetre, a tapasztalati világra vonatkozóan is. Lord 1953-as munkájára sokan hivatkoztak, használták fel támogatásul, elsősorban a sorrendi szint intervallumszintű kezeléséhez és paraméteres tesztek végzéséhez, valamint bírálták is. A támogatók közül Anderson úgy parafrázálta Lordot, hogy „Lord rámutat, a statisztikai teszt aligha lehet tisztában azoknak a számoknak az empirikus jelentésével, amelyekkel foglalkozik. Következésképpen egy statisztikai következtetés érvényessége nem függhet a használt mérőskála típusától” (*Anderson, 1961, 309. o.*). A napjainkban is tartó vita a méréselméleti (Stevens nézeteit követő) tábor és a számelemző statisztikai tábor között feloldható egy harmadik megközelítéssel, a mérések és az elemzések pragmatista felfogásával.

Velleman és Wilkinson (1993) rámutat arra, hogy Stevens kategóriái nem az adatok rögzített és változtathatatlan tulajdonságai, hanem ugyanaz az adat az értelmezésétől függően többféle skálájú is lehet. Egyik példájukban egy fogadásra a terembe érkező embereknek olyan tombolaszelvényeket osztanak ki, amelyek egyes sorszámokkal kezdődnek, és a belépés sorrendjével eggyel növekednek. Amikor a nyertes számot, a 126-ost kihúzták és kihirdették, az egyik résztvevő összehasonlította a szelvényével, hogy megtudja, nyert-e. Ekkor nominális skálájúként kezelte az adatot. Ezután körülnézett a teremben, és azt mondta, hogy nem tűnik úgy, hogy 126 személy lenne a teremben. Ez az értelmezés arányskálájú volt. Egy 56-os sorszámú résztvevő megállapította, hogy túl korán érkezett a terembe ahhoz, hogy nyerjen, ez az értelmezés ordinális skálájú volt. Ha a résztvevő ismerné a terembe érkezés gyakoriságát, akkor meg tudná becsülni azt az időt, ami a nyereshez szükséges lett volna. Ebben az esetben az adat értelmezése intervallumskálájú lett volna. A szerzők másik példája az autók hengerszáma, ami értelmezhető kategorikus változóként (nominális skálájúként, például megvizsgálva a hengerek

száma szerint az autók átlagos fogyasztását), sorrendiként, intervallum- és arányskálájúként (például megvizsgálva a hengerek átlagos számának időbeli változását) is (*Velleman–Wilkinson, 1993, 69. o.*). Az utóbbi példa a kicsi lehetséges darabszámú diszkrét mennyiségi változókra általánosítható, amelyeket sorrendbe helyezett kategóriákként is lehet kezelni, nemcsak arányskálájú adatokként.

Torgerson (1958) skálázásról írt nagyon alapos művében bírálja Stevens mérés meghatározásának több vonását. Stevens definíciójában nem szerepel a tulajdonság, csak az objektum. Stevens számára, ha a számjegyeket szabályok (bármilyen szabályok) szerint rendeljük a tárgyakhoz, akkor mérésről van szó, és a tárgy az, amit mérnek, és nem a tárgy valamely tulajdonsága. *Torgerson* példájában a mérés kifejezés használható botok halmokba rendezésének jelölésére aszerint, hogy tölgy-, szil- vagy fenyőfán nőttek-e, mindaddig, amíg a halmok megnevezésére számjegyeket használnak, nem pedig szavakat. E nézet szerint megmértünk egy botot nominális szinten, amikor megállapítjuk, hogy az a bizonyos bot egy „kettes”. Így e megközelítés számára az osztályozás, sőt az egyes példányok megnevezése is egyfajta méréssé válik. Stevensnek ezt az értelmezését *Torgerson* elutasítja, a mérés kifejezést a hagyományosabb nézet szerint alkalmazza, miszerint a mérés a tárgyak tulajdonságaira vonatkozik, és nem magukra a tárgyakra. Így egy bot *Torgerson* fogalomhasználatában nem mérhető, bár a hossza, súlya, átmérője és keménysége igen (*Torgerson, 1958, 14. o.*). *Torgerson* mérési szintjei annyiban bővebbek Stevensénél, hogy az ordinális szintnél megkülönbözteti a természetes nullaponttal nem rendelkező és rendelkező skálákat, és annyiban szűkebb, hogy a nominális mérési szint nem mérés, hanem klasszifikáció.

William Rozeboom pszichológus dogmatikusnak és merev kategorizálásnak tartva bírálja a Stevens-féle skálák szigorú rendszerét. Szerinte hibás az a kérdésfeltevési sorrend, amely a skálák típusából szeretné a megengedhető műveleteket meghatározni. Az a kérdés ezt megelőzi, amely a vizsgált jelenség, jellemző sajátosságai alapján előzetesen felméri, hogy az milyen módon mérhető. Az IQ mérésével kapcsolatos gyakori példával jól be lehet mutatni a különbséget. A 150-es IQ-ról nem mondható el, hogy kétszer akkora, mint a 75-ös IQ. A mérési skálák alapján mondhatnánk azt, hogy azért nem, mert az intelligencia nem arányskálán mért jellemző.¹⁹ De ennél pontosabb az a válasz, hogy azért nem, mert magának az intelligenciának nincsen olyan tulajdonsága, amelyet egy természetes nullapontból induló skálával lehetne megjeleníteni (*Rozeboom, 1966, 196. o.*). Rozeboom elismeri a mérés jelentésének kibővítését és Stevens meghatározásának a népszerűségét, de szerinte az a meghatározás nem jó, amely a kategóriák, minőségek számokkal kifejezését mérésnek tekinti (*Rozeboom, 1966, 223. o.*).

¹⁹ Bár volt kísérlet az abszolút nulla intelligencia kezdetű arányskála megalkotására *Thurstone (1928)* részéről.

Nicholas Cox geográfus szerint Stevens sémája, Stevens terminológiáját alkalmazva, egy négyponos ordinális skála, amely nem jeleníti meg a mérés sokszínűségét. A változók jelentős osztályai maradnak ki belőle, a kompozit változók, az irányítusú adatok. Nem különbözteti meg a diszkrét és folytonos, a korlátos és a nem korlátos változókat (Cox, 1996, 480. o.). Egy versenyben elért helyezés sorrendi skála Stevens szerint, amely sokféleképpen transzformálható anélkül, hogy a belőle számított statisztikai mutatók máshogyan változnának, mint a transzformáció. Cox szerint azonban a helyezések hasonlítanak a megszámlálással kapott arányskálátípusú adatokra, a helyezések pedig számos olyan fontos információt tartalmaznak, amely az adattranszformáció révén elvész. A negyedik és a hetedik helyezett között például két pozíció található, az első helyezett előtt viszont egy sem. Egy állandó hozzáadásával vagy állandóval való szorzással ezek az információk eltűnnek, ezen túlmenően az első helyezett sem venné jó néven, ha átneveznék kilencediknek (Cox, 1996, 481. o.). A nominális skálákról Stevens azt állítja, hogy csak önkényesen címkézhetők fel számokkal, amelyekkel nem megengedettek (vagy nem értelmesek) a műveletek. A bináris változókra adott 0-ás és 1-es kódokkal való számtani műveletek azonban sokszorosan bizonyítottan értelmesek, jól értelmezhetők és hasznosak (Cox, 1996, 481. o.).

Zumbo és Kroc (2019) frusztrálónak tartja, hogy sok kvantitatív társadalomkutató még mindig úgy hagyatkozik Stevens mérési skáláira a módszertani kritikáik megfogalmazásakor, mintha ez egy matematikailag koherens módja lenne a mérés osztályozásának. A szerzők szerint a matematikai és a statisztikai szakirodalomban az elmúlt harminc évben csaknem eltűnt ez a megközelítés, mivel nem osztályozza megfelelően a tényleges méréseket. Zumbo korábban több helyen is foglalkozott az elemzési eljárásokra vonatkozó, szerinte indokolatlan korlátozásokkal (Zimmerman–Zumbo, 1993; Zumbo–Zimmerman, 1993). Miközben Stevens korlátozta a nominális és az ordinális változókkal megengedhető műveleteket, valójában ezek a változók sokszor használhatók Stevens által csak intervallum- és arányszinten engedélyezett eljárásokban. A szerzők a regressziószámítást és a bináris változókkal végzett műveleteket hozzák fel az egyik típusú példának. Másik példájuk értelmezési. Ha egy, mindössze egy kérdésből álló tesztnél a jó választ 1-essel kódolják, a rossz választ 0-val, akkor az eredmény kétféleképpen értelmezhető. Egyrészt, mint a jóság nominális skálán mért mutatója. Másrészt, pontszámként értelmezve, intervallumskálát kapunk. Vagyis Zumbo és Kroc szerint a szemantikai döntéstől függően lehet ugyanaz az információtartalom vagy nominális, vagy intervallumskálájú, ami matematikai és statisztikai értelmezés alapján értelmetlen.

A tipizálás hiányos volta és a nem egyértelműen kategorizálható adattípusok problémája a javasolt mérési szintek bővítésével, a meglévő szintek átdolgozásával orvosolható. Sajnos Coombs (1952) a nominális és az ordinális szintet sokkal részletesebben felosztó rendszere közel sem lett annyira ismert, mint Stevensé.

Mosteller és Tukey (1977) a regressziószámítás kapcsán adták meg az adattípusok részletesebb tipológiáját:

1. megnevezések (nem mérés),
2. fokozatok,
3. sorrend,
4. törtszámok (százalékok),
5. egyenlegek (különbségek),
6. darabszámok,
7. mennyiségek.

Ezt a gyakorlatias kategorizálást csak részben lehet megfeleltetni Stevens skáláinak. Megkülönbözteti a fokozatokat és a sorrendeket, ami még a Stevensi ordinális szint indokolt kettébontása, de megjelennek benne olyan adattípusok (a törtszámok és egyenlegek) is, amelyek fontosak, azonban a mérési skálák rendszerében mégsem helyezhetők el.

Chrisman (1998) bemutatja Stevens hatását a kvantitatív geográfiára, majd úgy fogalmazza újra a mérési szinteket, azokból tízet meghatározva, hogy abban sajátosan területi adatok is elhelyezhetők legyenek (9. táblázat). A rendszerét azonban így sem tekinti teljesnek. Bár Chrisman lényegesen bővíti Stevens rendszerét, anynyiban megmarad a keretei között, hogy mérési szintekről, skálákról és azok hierarchiájáról beszél, miközben megfelelőbb lenne hierarchia nélküli adattípusokként felfogni a rendszert. Sajnos az egyes szintek ismertetése nem túl részletes, így maradnak nyitott kérdések öt szintnél (a 9. táblázat 2., 3., 5., 6. és 8. sorszámú adattípusainál) is. Az extenzív arányt nem részletezi, csak táblázatban adja meg. Amit a logaritmikus skáláról ír, annak egy része tévedés. A földrengések erősségének kifejezése Richter-skálán logaritmikus, de másik példája, a Mercati-skála lineáris. A földrajzban jelentős szerepet játszó sűrűségmutatók (pl. népsűrűség), vagy magyar terminológiával élve, intenzitási viszonyszámok szintén nem logaritmikus skálájúak.

A ciklikus arány nem illeszthető bele a négy hagyományos szintbe. Bár van egy nullapontja, ami önkényes irányú, 360 fokra osztott körnél az 1 fok és a 359 fok közötti különbség 2 fok lesz, a nullafoktól való távolságuk 1 fok. Ilyen ciklikusság az időbeli adatoknál is fellép, az éven belüli szezonoknál és a napszakoknál. Chrisman szerint az arányskála nem a legmagasabb mérési szint, mert csak a kezdőpontja természetesen rögzített a nullánál, a mértékegysége önkényes. Szerinte a legmagasabb mérési szintnél a skála egységnyi, kezdő- és végpontja is rögzített. Ilyen mérések a megoszlási viszonyszámok és a valószínűségek, amelyek nem szerepelnek Stevens rendszerében. Ezt Chrisman abszolút skálának nevezi, míg a korábbi terminológia az abszolút skálát a darabszámokra alkalmazta. A valószínűségek és a megoszlási viszonyszámokat kifejező százalékok egy kategóriába sorolása azzal a veszéllyel jár, hogy megoszlási viszonyszámokat valószínűségeként

értelmeznek és kezelnek, annak ellenére, hogy a valószínűségi értelmezéshez véletlen mechanizmusra lenne szükség.

9. táblázat

Chrisman (1998) mérési szintjei
Chrisman's (1998) measurement levels

	Mérési szint	Megjegyzés
1.	Nominális	A kódok nem szerepelnek benne, a kategóriák határai elmosódhatnak.
2.	Fokozatok	Megfelel Stevens sorrendi szintje egy részének.
3.	Ordinális	Megfelel Stevens sorrendi szintje egy részének.
4.	Intervallum	Megfelel Stevens intervallumszintjének.
5.	Log-intervallum	Logaritmikus skála.
6.	Extenzív arány	Összeadhatóság érvényesül.
7.	Ciklikus arány	Mértékegység és a ciklus hossza.
8.	Származtatott arány	Egynél több mennyiség kombinációja.
9.	Darabszámok	
10.	Abszolút	Valószínűség, százalék.

Forrás: Chrisman (1998, 236. o.).

6. A Likert-skála és a mérési skálák

Rensis Likert 1932-ben javasolt egy később róla elnevezett technikát az attitűdök mérésére, amit nem sokkal később tovább részletezett egy következő írásában.²⁰ A módszer egyszerűsége, könnyen kezelhetősége és értelmezhetősége miatt széles körben elterjedt az attitűdök mérésén és a pszichológián kívül a szociológiában, a közvéleménykutatásban, az oktatásban, az egészségügyi felmérésekben, a marketingkutatásban, a fogyasztói elégedettség felmérése során és sok más területen. Rése a statisztikai alapismereteknek és az elemzési gyakorlatnak. Gyakorlati használatán túl rendszeresen és gyakran jelennek meg róla módszertani tanulmányok és

²⁰ Likert tanulmányainak a címe: *Az attitűdök mérésére szolgáló technika, illetve Egyszerű és megbízható módszer a Thurstone-skálák pontozására*. A második címből látszik, hogy egy, az attitűd mérésére használt korábbi nehézkes eljárás helyett vagy mellett egy sokkal egyszerűbb eljárást javasolt, amely egyszerűen kezelhető voltának köszönhető gyors elterjedését.

áttekintő írások. Ennek ellenére néhány félreértés vagy hiányosság is jellemző a tárgyalására, ezért egy rövid, nem részletekbe menő ismertetést adok róla és a vele kapcsolatban folyamatosan napirenden lévő fontosabb módszertani kérdésekről.

10. táblázat

**A Likert-típusú állításokra adott gyakori válaszlehetőségek,
eltérő számú válaszlehetőséggel**

*Common response options to Likert-type statements,
with different numbers of response options*

Válaszok száma	Válaszok megnevezése								
	1	2	3	4	5	6	7	8	9
2	Nem értékek egyet	Nem							
3	Nem értékek egyet	Közömbös számomra	Egyet-érték						
4	Nem értékek egyet	Inkább nem értékek egyet	Inkább egyet-érték	Egyet-érték					
5	Nem értékek egyet	Inkább nem értékek egyet	Közömbös számomra	Inkább egyet-érték	Egyet-érték				
6	Egyáltalán nem érték egyet	Nem érték egyet	Inkább nem érték egyet	Inkább egyet-érték	Egyet-érték	Teljesen egyet-érték			
7	Egyáltalán nem érték egyet	Nem érték egyet	Inkább nem érték egyet	Közömbös számomra	Inkább egyet-érték	Egyet-érték	Teljesen egyet-érték		
8	Egyáltalán nem érték egyet	Nem érték egyet	Inkább nem érték egyet	Nagyon kicsit nem értékek egyet	Nagyon kicsit egyet-érték	Inkább egyet-érték	Egyet-érték	Teljesen egyet-érték	
9	Egyáltalán nem érték egyet	Nem érték egyet	Inkább nem értékek egyet	Nagyon kicsit nem értékek egyet	Közömbös számomra	Nagyon kicsit egyet-érték	Inkább egyet-érték	Egyet-érték	Teljesen egyet-érték

Forrás: saját szerkesztés.

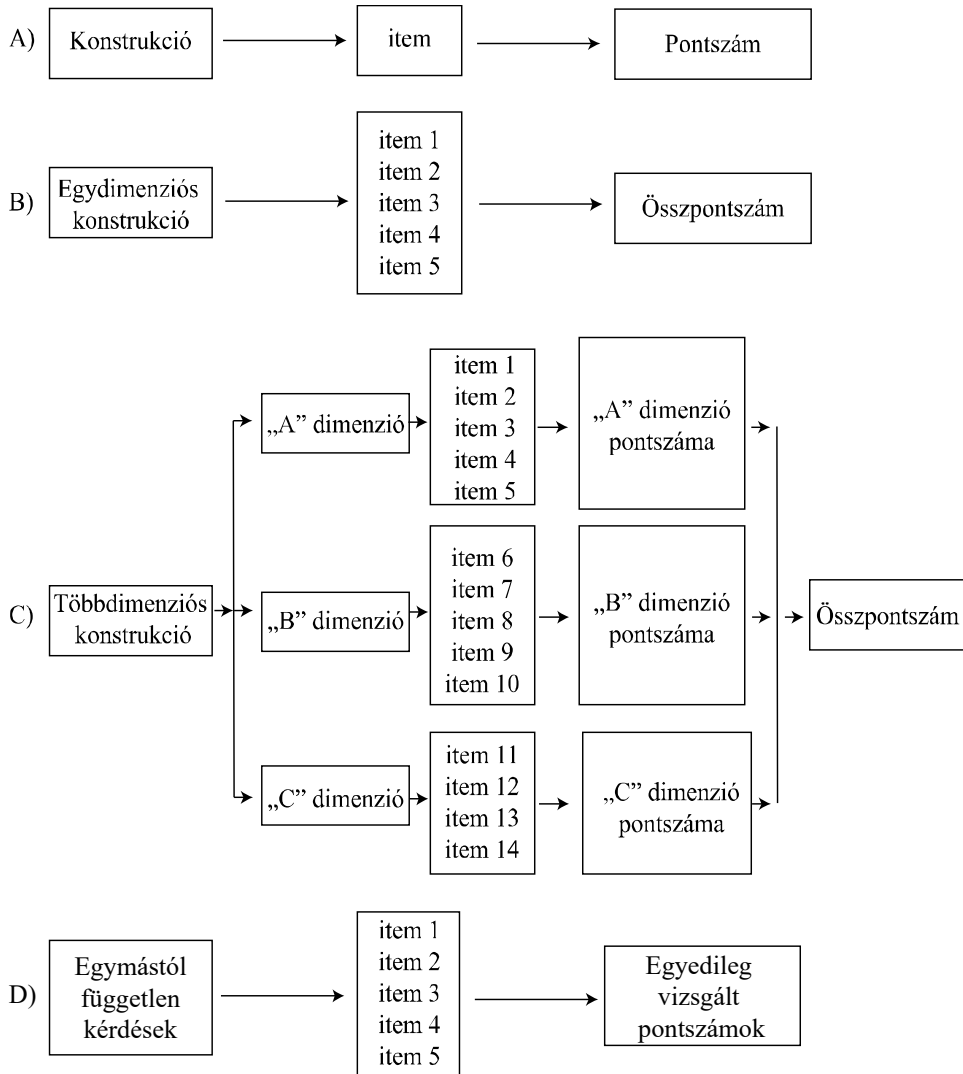
A Likert-skála eredetileg összefüggő állítások (itemek) sorozatát jelentette. Az állítások mindegyike a mérendő attitűdre vonatkozott. A felmérésben részt vevőket arra kérik, hogy jelezzék, mennyire értenek egyet (a teljes mértékben egyetértéstől a teljes mértékben nem egyetértésig) az adott állításokkal. A válaszlehetőségek számának megfelelően a válaszokat pontszámokká alakítják, például a következő módon: egyáltalán nem értek egyet = 1 pont; inkább nem értek egyet = 2 pont; közömbös számomra = 3 pont; inkább egyetértek = 4 pont; teljesen egyetértek = 5 pont. Ezeket az állításokat Likert-típusú állításoknak, Likert-tételeknek, Likert-itemeknek lehet nevezni, de ezenkívül számos további elnevezés is használatos, mint például a pontozási skála, értékelő skála, egyetértési skála. A bipoláris válaszformátum az egyetértés, illetve nem egyetértésen kívül vonatkozhat értékelésre (jó-rossz és hasonló ellentétpárokkal), gyakoriságra (soha-mindig) és fontosságra is (egyáltalán nem fontos-nagyon fontos). A 10. táblázat eltérő számú válaszlehetőség lehetséges megszővegezését mutatja be. Ugyanakkor a skála két végpontjának a megszővegezése is elegendő lehet, különösen, ha magas a válaszlehetőségek száma.

A Likert-skála eredetileg egydimenziós volt, tehát ugyanarra a jellemzőre vonatkozott a többféle Likert-típusú állítás, és az azokra adott válaszok összegezhetőek és egyetlen pontszámmal kifejezhetőek voltak. Az eredeti terminológia a gyakorlati használat során azonban két fontos ponton is átalakult: egyrészt manapság a pszichometrián kívül Likert-skála alatt gyakrabban értenek egyetlen Likert-típusú állítást, mint Likert-típusú állítások sorozatát; másrészt, attitűdök mérésén kívül rendkívül sok olyan egyéb alkalmazás jelent meg, ahol a Likert-skála kifejezést használják, és nem azt, hogy a válaszformátum ötfokozatú skála (vagy más számú válaszlehetőséget kínáló) volt. Ha eredeti értelmű Likert-skáláról van szó, akkor az összeállítójáról (személy vagy intézmény), a mért koncepcióról vagy valami másról nevezik el, de többnyire rövidítésükkel hivatkoznak rájuk (pl. Big Five, HEXACO, MMPI, IPI, MSQ, HRSD, BDI, HAD, MADRS, GDS, DACL, GBS, CFQ, AMT stb.). További változás az eredeti alkalmazáshoz képest, hogy Likert-típusú állításokat egymástól függetlenül is beépíthetnek egy felmérésbe, amelynek nincs olyan célja, hogy Likert-skálát képezzenek az egyes tételekből. A Likert-típusú állításoknak ezt a többféle használati módját mutatja be az 1. ábra.

A Likert-skála kettős jelentését, az egyedi állításokat és az állítások összességét a 11. táblázatban is kiemeltem. Ez többnyire nem vezet félreértésekhez, de tisztázását a mérési skálákkal való kapcsolódások miatt fontosnak gondolom. Nem tartom zavarónak, ha az egyszerűség kedvéért, rövidege és megszokottsága miatt Likert-skálán egyetlen állítást értenek, de a továbbiakban az egyértelműség kedvéért a fenti értelemben elkülönítve használom a Likert-skála és a Likert-típusú állítás kifejezéseket. Korábban már említettem, hogy Stevens a skála kifejezést egy ettől merőben eltérő jelentésben alkalmazta és honosította meg, nyelvi zavart okozva azzal, hogy az adattípusra a „mérési skála” kifejezést használta és terjesztette el.

1. ábra

A Likert-típusú állítások (itemek) többféle használata
Different uses of Likert-type statements (items)



Forrás: saját szerkesztés.

11. táblázat

A Likert-skála kettős jelentése
Two meanings of the Likert scale

Likert-skála	
eredeti	szokásos, jelenleg gyakrabban használt
jelentés szerint	
Likert-típusú állítások sorozata ugyanarra a kérdéskörre vonatkozóan, illetve az állításokra adott válaszokból képzett pontszám.	Egyetlen Likert-típusú állítás, amelynél a válaszok ötfokozatú (vagy ettől eltérő számú fokozatot tartalmazó) skálán adhatók meg.
Egymással kapcsolatban lévő kijelentések összessége attitűdök mérésére.	Egyedi kijelentés, bárminek (nemcsak attitűdöknek) a felmérésére.

Forrás: saját szerkesztés.

A Likert-skálához tehát állítások összefüggő rendszerét kell létrehozni. Aki egy pozitívan megfogalmazott állítással egyetért, az a logika szabályai szerint ugyanazon állítás negatív megfogalmazásával nem ért egyet. Ez alapján lehet a válaszok érvényességét ellenőrző állításokat elhelyezni a Likert-típusú kérdések rendszerében. Az azonban problémát jelent, hogy aminek logikailag érvényesnek kellene lennie (a pozitív és negatív megfogalmazások ekvivalenciájának), az pszichológiai szempontból nem feltétlenül érvényesül.

A Likert-skálával és a Likert-típusú állításokkal kapcsolatos módszertani irodalom rendkívül gazdag, és gyakrabban érinti a Likert-típusú állításokat, mint a Likert-skálát. A fontosabb kérdéseket a 12. és a 13. táblázatban foglaltam össze. A táblázatban megfogalmazott vitapontok óriási, évtizedeket átfogó szakirodalmát nem részletezem, csak azokat, amelyek közvetlenebbül érintik a mérési szint kérdéskörét. Néhány esetben a megadott hivatkozott irodalom adhat további iránymutatást. Általánosságban bármilyen mérésekről is van szó, azok korrekt összehasonlíthatóságához a méréseket azonos mérőeszkővel vagy megfelelő módon kalibrált mérőeszkővel szükséges végrehajtani. A felsorolt vitapontokból is jól látható, hogy a Likert-típusú állításoknál ez gyakorlatilag nem kivitelezhető. Hiába ugyanaz az állítás, a válaszadók egyénileg értelmezik azokat, ahogyan a válaszlehetőségek jelentésénél is szerepet játszik a belső értelmezés. Nem beszélhetünk olyan típusú, objektív külső jelentésről, mint a fizikai méréseknél. Egy Likert-típusú állítást aligha lehet azonosan kalibrált mérőeszkőnek tekinteni.

12. táblázat

A Likert-skálával kapcsolatos fontosabb módszertani kérdések
Main methodological issues related to the Likert scale

	A Likert-skálára vonatkozó kérdés	Rövid válasz
1.	Minimum mennyi Likert-típusú állítás szükséges egy Likert-skálához?	Nincs egyértelmű küszöbérték. Egyetlen Likert-típusú állítás kevés a véletlen és egyéb ingadozások miatt, ugyanakkor konkrét és nagyon egyszerű esetekben mégis elég lehet. Minimum négy állítás szükséges a konzisztencia értékeléséhez (<i>Diamantopoulos et al., 2012; Willits–Theodori–Luloff, 2016</i>).
2.	Milyen hatása van a kérdés-sorrendnek?	Kísérletileg sok bizonyíték van arra, hogy a kérdések sorrendje befolyásolja a válaszokat (<i>Brecsok–Németh, 2020</i>).
3.	Az egyes Likert-típusú állítások ugyanazt az egy dimenziót mérik-e?	Sokféleképpen vizsgálható: korreláció, reliabilitás mérőszámai, faktoranalízis (ugyanakkor létezik több dimenziót mérni hivatott felmérés is).
4.	Az egyes Likert-típusú állítások azonos súlyúak-e?	Általában igen, de ettől el lehet térni.
5.	Milyen hatása van a kérdőív felépítésének, elrendezésének, a lekérdezés módjának (személyes, telefonos, önkitöltős, online)?	Évtizedek óta kutatott, van szisztematikus hatás, amit nehéz számszerűsíteni.
6.	Vannak-e kulturális eltérések a válaszok jellegében?	Igen, mind a szélsőséges válaszok arányában, mind az átlagpontszámokban (országok, nemzetek, nyelvek, vallások és egyéb csoportok közötti eltérések) (<i>Chen–Lee–Stevenson, 1995; Watkins–Cheung, 1995; Heine et al., 2002; Smith, 2004</i>).
7.	Mekkora elemszám szükséges az elemzéshez?	Egzaktul nem meghatározható. A nagyobb elemszám előnyösebb, nagyobb érvényű az elemzés, és alcsoportok is elemezhetők.
8.	Mi a mérési szintje a Likert-típusú állításra adott válaszok összegzett értékének?	Néha még az is kérdés, hogy egyáltalán összeadhatók-e az egyedi pontszámok (<i>részletesebben tárgyalom</i>).

Forrás: saját szerkesztés.

13. táblázat

A Likert-típusú állításokkal kapcsolatos fontosabb módszertani kérdések
Main methodological issues related to Likert-type statements

	A Likert-típusú állításra vonatkozó kérdés	Rövid válasz
1.	Mennyi legyen a válaszlehetőségek száma?	Nem egyértelmű, vegyessék az érvek és a tanácsok. Hatékonyan lehet használni a háromfokozatú skálát, és használnak 101 fokozatú (0-tól 100-ig terjedő) skálát is. Az információelméleti érvek a több válaszlehetőség mellett szólnak. A leggyakoribb az 5 és 7 pontos skála, 7 pont felett már nagyon kicsi az információtöbblet (Cox, 1980; Preston–Colman, 2000; Lozano–Carcía–Cueto–Muniz, 2008; Jones–Loe, 2013; Simms et al., 2019).
2.	Meg legyenek-e szövegezve a közbülső válaszlehetőségek, vagy csak a két végponté?	Semlegesebb, ha csak a skála két végpontját nevezik meg. A kérdést vizsgáló kutatások nem mindig találtak eltérést a kétféle szövegezés (csak a végpontoké vagy a közbülső válaszlehetőségeké is) között.
3.	Van-e különbség a pozitívan és a negatívan megszóvegezett állításokra adott válaszokban?	A negatívan és a pozitívan megszóvegezett állítások nem ekvivalensek. Ugyanazon állítás negatív és pozitív változatára eltérő válaszok érkehetnek (ezt számos kísérlet igazolja).
4.	Milyen irányú legyen a skála (bal oldalon a negatív vagy a pozitív válaszlehetőség legyen, bal oldalon az alacsony pontszám legyen vagy a magas)?	A bal oldali válaszlehetőséget kicsit gyakrabban választják a kitöltők. Általában bal oldalon van a negatív, alacsony pontszámú, jobb oldalon pedig a pozitív, magas pontszámú válaszlehetőség. A fordított elrendezés kicsit magasabb pontszámokat eredményezhet (Barnette, 2000; Amoo–Friedman, 2001; Nicholls et al., 2006; Hartley–Betts, 2010).
5.	Milyen hatása van a skála szélső és középső pontjai eltérő számozásának?	Magasabb értékeket eredményez a –5-től +5-ig, –2-től +2-ig terjedő skála, mint a 0-tól 10-ig vagy az 1-től 5-ig terjedő (Armitage–Deeprise, 2004; Schwarz et al., 1991).
6.	Milyen hatása van a kérdések és a válaszok megfogalmazásának, szóhasználatának?	Van hatása, a jelentés azonos nyelven belül is eltérhet egyes személyek és csoportok számára, ez a tesztek fordításánál, eltérő nyelvre adaptálásánál is sok nehézséget okoz (Stegmuller, 2011).
7.	Páros vagy páratlan számú válaszlehetőség legyen?	Páros számú válaszlehetőség esetén nincs lehetőség a középső válaszban megtestesülő semlegesség kifejezésére, ami egy valós lehetőséget zár ki. Ugyanakkor nem különböztethető meg a választás sokféle lehetséges motivációja (származhat ismerethiányból, vélemény hiányából, érdektelenségből és döntés hiányából is), a személyes vélemény elhallgatása is motiválhatja a válaszadót.
8.	A válaszok között mekkora a pszichológiai távolság?	Nem megállapítható. Ha a végpontok közötti lehetőségek nincsenek megnevezve, akkor az egyenletesség feltételezhető. Nincs nagy jelentősége a kérdésnek, jelentősége akkor lenne, ha a válaszok között sok nagyságrendnyi lenne a különbség, ami azonban nem valószínű (pl. nem értek egyet: 1; inkább nem értek egyet: 1,01; közömbös: 1,02; inkább egyetértek: 5; egyetértek: 10000).
9.	Mi a mérési szintje a Likert-típusú állításnak?	Részletesebben tárgyalom.

Forrás: saját szerkesztés.

Ami a Likert-típusú állítások mérési szintjét illeti, a szakirodalom tanulmányozása alapján úgy tűnik, hogy az elmélet-módszertan és a gyakorlat annyiban szétvált, hogy míg a módszertani írások között több az olyan, amely sorrendi szintűnek tartja (és ezzel együtt az elemzési lehetőségek korlátozottabb számát engedélyezné), addig a tényleges elemzési gyakorlatban inkább intervallumszintű változóként kezelik, és átlagot, szórást számítanak, többváltozós módszerekben szerepeltetik a válaszokhoz rendelt számokat, és paraméteres következtető statisztikai eljárásokat használnak. Az állításokhoz kapcsolt pontszámokkal végzett olyan matematikai műveletek, mint a számtani átlag és a szórás számítása ellen felvetett érvek a válaszkategóriák közötti intervallumok nem azonos mértékű, illetve ismeretlen nagyságú különbségein alapulnak. Ez a pszichológiai méréseknek azt az objektíven nem megoldható alapkérdését érinti, amelyet korábban már röviden tárgyaltam. Ha elvetjük a számszerűsítés lehetőségét, akkor a kérdésekre adott válaszkategóriák arányainak a közlése lehetséges, amely ugyan kisebb információ-tartalommal jár, de a mérhetőség alapkérdésére általánosságban érintett, mint a válaszok numerikus kezelése.

A 14. táblázat néhány, az ordinális szinten kezelés mellett szóló véleményt, álláspontot mutat be, amelyet könnyebb találni, mint olyanokat, amelyek megengedik a mennyiségi ismérvként kezelést. Az utolsónak idézett tanulmányban egy intervallumszintű értelmezés is olvasható: „A Likert-skálák esetében szakirodalmi kutatásokra támaszkodva általánosan elmondható, hogy a kutatók a skálaértékeket többek között átlagolják, így élnek azzal a feltételezéssel, hogy intervallumváltozóról van szó, hiszen ebben az esetben van értelme az átlagszámításnak” (*Zerényi, 2016, 470. o.*). A sorrendi változók intervallumszintű kezelését támogató következő általános megjegyzés kiválóan érzékelteti, hogy a téma elméleti kezelésében sincs konszenzus: „Empirikus bizonyítékok alátámasztják, hogy a sorrendi változók kezelhetők úgy, mintha intervallumskálájúak lennének. Bár a sorrendi változók intervallumváltozóként kezelését kis hiba kísérheti, ezt ellensúlyozza az erősebb, érzékenyebb, jobban kidolgozott és egyértelműbben értelmezhető statisztikák használata, ismert mintavételi hibával” (*Labovitz, 1970, 515. o.*). Ez a megközelítés azonban nem is annyira a Likert-típusú állításokra vonatkozik, hanem inkább abba az általános megközelítésbe illeszkedik, amelyet Lord kritikája kapcsán ismerttettem, miszerint „a számok nem emlékeznek arra, hogy honnan jöttek”, így lehet átlagot számolni belőlük, ahogyan az iskolai osztályzatokból is, valamint paraméteres tesztek is lehet velük végezni. A Likert-skálára vonatkozóan egyértelműen fogalmaz *Norman (2010)*, amikor azt írja, hogyha Jamiesonnak és másoknak igazuk van, és nem használhatunk parametrikus módszereket Likert-skálás adatokon, és be kell bizonyítanunk, hogy adataink pontosan normális eloszlásúak, akkor az oktatási, az egészségi állapot és az életminőség értékelésével kapcsolatos kutatásaink mintegy 75%-át gyakorlatilag a kukába dobhatjuk. Norman így zárja

a tanulmányát: „A parametrikus statisztika Likert-adatokkal, kis mintanagysággal, nem egyenlő szórással és nem normális eloszlással is használható, nem kell attól tartani, hogy »rossz következtetésre jutunk«. Ezek a megállapítások összhangban vannak a közel 80 éves empirikus irodalommal. A vita megszűnhet (de valószínűleg nem fog)” (Norman, 2010, 631. o.).

Egyetértek Norman tanulmányának előbb idézett végszavával, miszerint a Likert-adatokkal használható intervallumskálán engedélyezett eljárások. Annak a több évtizedes vitának, amely akörül alakult ki, hogy a Likert-skála és a Likert-típusú állítások vajon ordinális vagy intervallummérési szintűek, a kiindulópontját azonban rossznak tartom, mert a négyfokozatú mérési skálák keretrendszerében értelmezett. Korábban már láthattuk, hogy Stevens mérési skálái sok szempontból hiányosak, különféle adattípusok nem vagy nem egyértelműen helyezhetők el a rendszerben. Ilyen típusú adatok a Likert-típusú állítások is, amelyeknél az összes mérési szint mellett lehetne érvelni:

- Nominális mérési szintű, mert a válaszok kategóriákat jelentenek.
- Ordinális mérési szintű, mert a válaszok sorrendbe helyezhetők.
- Intervallummérési szintű, a válaszok közötti távolságot azonosnak feltételezve.
- Abszolút mérési szintű (ez az arányskála a esete), amennyiben egyes válaszok darabszámát vesszük.

Ezek az érvek az egyértelmű besorolást nem teszik lehetővé, és ugyanúgy alkalmazhatók a Likert-skálára is, vagyis a Likert-típusú állításokat tartalmazó lista összesített értékeire. Az abszolút mérési szint a Likert-skálára még inkább érvényes lehet (amennyiben az egyetértő válaszok darabszámát vesszük csak figyelembe, és eltekintünk a válaszok pontértékeitől), mint egyetlen kérdésre alkalmazva, amely dichotóm változóként sajátosan viselkedik. Az elméletileg leggyakrabban támogatott ordinális mérési szinttel kapcsolatban komoly ellenérvként lehet megemlíteni azt, hogy a Likert-típusú állítások nem a válaszadókat helyezik sorrendbe, nem a válaszadókat rangsorolják, hanem a válaszadók választanak sorrendbe helyezett válaszokból. Az azonos sorrendet választó válaszadók kerülnek egy kategóriába, de, mint írtam, ez nem tekinthető egy azonosan kalibrált mérőeszközzel történt mérés eredményének. A Likert-skála és a Likert-típusú állítások egy sajátos adattípust jelentenek, amely nem helyezhető el a négyfokozatú mérési skálák rendszerében. Az ordinális szinten való értelmezés pont a Likert-típusú állítások használatának fő célját, előnyös kezelési, elemzési lehetőségeit akadályozná.

14. táblázat

**A Likert-típusú állítások ordinális mérési szintjének megfogalmazása
néhány szerzőtől***Determining the ordinal measurement level of Likert-type statements by some authors*

„A módszertani megfontolások alapján általánosan elismert, hogy az attitűdöket mérő skálákat ordinálisnak kell tekinteni. Ennek ellenére sok tanulmány használ kardinális statisztikákat, mint átlagokat, szórásokat, t-teszteket az attitűdadatok elemzésére. A korrekt ordinális megközelítések kisebbségben vannak.” (Göb–McCollin–Ramalhoto, 2007, 602. o.)

„Ebben rejlik a bűn: az ordinális adatok egész számokkal való egyszerű kifejezése nem indokolja a parametrikus statisztika használatát. Ahogyan nem helyes egy adott sebészeti beavatkozás eredményeit rossz, közepes, jó vagy kiváló kategóriába sorolni, és azt állítani, hogy az átlagos eredmény közepes és fél, ugyanúgy nem helyes ugyanezeket az eredményeket 1, 2, 3 vagy 4 értékkel értékelni, és azt állítani, hogy az átlagos eredmény 2,5.” (Kuzon–Urbanček–McCabe, 1996, 266. o.)

„Az értékelő skálás kérdéseket ordinális adatokként kell kezelni, bár nagyon sok példát találhatunk arra, hogy ezt a szabályt megsértik, és a nem parametrikus adatokat parametrikus adatokként kezelik. Ez elfogadhatatlan.” (Cohen–Manion–Morrison, 2011, 390. o.)

„A Likert-skálák az ordinális mérési szinthez tartoznak. Ez azt jelenti, hogy a válaszkategóriák rangsorolva vannak, de az értékek közötti intervallumok nem feltételezhetők egyenlőnek, bár, ahogy Blaikie rámutat, a kutatók gyakran feltételezik, hogy azok. Cohen és társai azonban azt állítják, hogy illegitim arra következtetni, hogy a »határozottan nem értek egyet« és a »nem értek egyet« közötti érzés intenzitása egyenértékű a Likert-skála más, egymást követő kategóriái közötti érzés intenzitásával. A Likert-típusú kategóriák esetében az intervallumskála feltételezésének jogossága fontos kérdés, mivel a megfelelő leíró és következtető statisztikák különböznek az ordinális és az intervallumváltozók esetében, és ha a kutató rossz statisztikai technikát használ, megnő annak az esélye, hogy rossz következtetést von le a kutatása szignifikanciájáról (vagy annak hiányáról).” (Jamieson, 2004, 1217. o.)

„A klasszikus tesztelmélet keretei között a Likert-skálás kérdések esetében a skálaértékek átlagolására nincs mód, hiszen azok nem intervallumskálán helyezkednek el, így legfeljebb az ordinális skála alapján megengedhető statisztikai műveletek elvégzésére és vizsgálati módszerek alkalmazására nyílik lehetőség.” (Zerényi, 2016, 474. o.)

7. Összegzés

A méréssel kapcsolatban tárgyalt kérdések alapvető jelentőségük a statisztikai elemzés minden fázisában, a felmérés előkészítésétől az adatok kiértékelésén át az eredmények értelmezéséig. A Stevens-féle mérési skálákkal és a mérés ahhoz tartozó, kibővített meghatározásával kapcsolatban (elismerve a rendszerezés bizonyos érdemeit is és Stevens nagyobb rugalmasságát, mint követőit) számos olyan probléma merül fel, amely indokolja ennek a statisztikai gyakorlat számára megalapozatlan korlátokat szabó fogalmi és értelmezési keretnek az elvetését és egy

rugalmasabb, gyakorlatiasabb, az adatok sokszínűségét figyelembe vevő felfogás alkalmazását. A mérési skálákkal kapcsolatban megfogalmazott előírások egy része érvényes és nyilvánvaló (kódokból sok mindent nem számolunk), más része azonban nem érvényes, és az elemzési gyakorlatban sokszor nem figyelembe vett, de rendszeresen indokolatlan hivatkozási alapot jelent a korrekt elemzésekkel szembeni bírálatokhoz. A mérés értelmezésével kapcsolatban a részletesebben tárgyalt, fontosabb következtetések az alábbiak voltak:

- A mérésnek rendkívül sok értelmezése létezik.
- A mérés reprezentációs felfogása magával a mérés folyamatával, körülményeivel nem foglalkozik.
- A négy mérési szintet négy különböző adattípusként lehet értelmezni, amelyekben túl sokkal többféle adattípus létezik.
- A Likert-skála nem illeszthető bele a négy mérési szintbe, külön adattípus.
- A sorrendi szinten belül fontos különbséget tenni a sorba rendezett kategóriák és a sorrendbe helyezett megfigyelési egységek között.
- Fel kellene oldani az adattípusok merev rendszerét.
- Ugyanaz az adat egyszerre több kategóriába, adattípusba is tartozhat, értelmezéstől függően.
- A minőségi kategóriákba sorolás alapvető jelentőségű a tudomány és a hétköznapi élet számára is, de nem kellene mérésnek tartani.
- A számok nem tudják, honnan jöttek, de az elemzőnek tudnia kell, és ez alapján kell eldöntenie, milyen elemzési eljárás alkalmazható, és ez alapján tudja értelmezni az eredményeket.
- A szakterületenként/jelenségenként eltérő mérési lehetőségeket el kell fogadni, egyik szakterület normáit nem szabad a másik szakterületre ráerőltetni.
- Következtető statisztikai eljárást csak véletlen jelenségek esetén kellene alkalmazni.

Irodalom

- Amoo, T. – Friedman, H. H. (2001): Do numeric values influence subjects' responses to rating scales? *Journal of International Marketing and Marketing Research*, 26, 41–46.
- Anderson, N. H. (1961): Scales and Statistics: Parametric and Nonparametric. *Psychological Bulletin*, 58(4), 305–316. <https://doi.org/10.1037/h0042576>
- Armitage, C. – Deeproose, C. (2004): Changing student evaluations by means of the numeric values of rating scales. *Psychology Learning and Teaching*, 3, 122–125. <https://doi.org/10.2304/plat.2003.3.2.122>

- Barnett, M. K. (1956): The Development of Thermometry and the Temperature Concept. *Osiris*, 12, 269–341. <https://doi.org/10.1086/368601>
- Barnette, J. J. (2000): Effects of stem and Likert response option reversals on survey internal consistency: If you feel the need, there is a better alternative to using those negatively worded stems. *Educational and Psychological Measurement*, 60, 361–370. <https://doi.org/10.1177/00131640021970592>
- Bárdits A. – Németh R. – Terplán Gy. (2016): Egy régi probléma újra előtérben: a nullhipotézis szignifikanciasztéves gyakorlata. *Statistikai Szemle*, 94(1), 52–75. <https://doi.org/10.20311/stat2016.01.hu0052>
- Benesch, H. (1994): *Pszichológia. SH atlasz*. Springer Hungarica Kiadó, Budapest
- Benyovszki A. (2013): *Leíró statisztika*. Ábel Kiadó, Kolozsvár
- Bigotti, F. – Barry, J. (2022): Introduction. In: Barry, J. – Bigotti, F. (eds.) *Santorio Santori and the Emergence of Quantified Medicine, 1614–1790*. Palgrave Macmillan, Cham, 1–63. https://doi.org/10.1007/978-3-030-79587-0_1
- Boring, E. G. (1923): Intelligence as the tests test it. *New Republic*, 36, 35–37.
- Borsboom, D. (2005): *Measuring the Mind. Conceptual Issues in Contemporary Psychometric*. Cambridge University Press, Cambridge <https://doi.org/10.1017/CBO9780511490026>
- Brecsok A. Á. – Németh R. (2020): A kérdéssorrendből fakadó kontextushatás. *Statistikai Szemle*, 98(3), 191–211. <https://doi.org/10.20311/stat2020.3.hu0191>
- Bridgman, P. (1927): *The logic of modern physics*. Macmillan, New York
- Brown, R. J. C. (2021): Measuring measurement – What is metrology and why does it matter? *Measurement*, 168. <https://doi.org/10.1016/j.measurement.2020.108408>
- Briggs, D. C. (2022): *Historical and Conceptual Foundations of Measurement in the Human Sciences. Credos and Controversies*. Routledge, New York <https://doi.org/10.1201/9780429275326>
- Campbell, N. R. (1920): *Physics, the elements*. Cambridge University Press, London
- Carterette, E. C. – Friedman, M. P. (1974): *Handbook of Perception. Volume II. Psychophysical Judgment and Measurement*. Academic Press, New York
- Chen, C. – Lee, S. Y. – Stevenson, H. W. (1995): Response style and cross-cultural comparisons of rating scales among East Asian and North American students. *Psychological Science*, 6, 170–175. <https://doi.org/10.1111/j.1467-9280.1995.tb00327.x>
- Chrisman, N. R. (1998): Rethinking Levels of Measurement for Cartography. *Cartography and Geographic Information Systems*, 25(4), 231–242. <https://doi.org/10.1559/152304098782383043>
- Christ, C. F. (1994): The Cowles Commission's Contributions to Econometrics at Chicago, 1939–1955. *Journal of Economic Literature*, 32(1), 30–59.
- Cohen, L. – Manion, L. – Morrison, K. (2011): *Research Methods in Education*. Routledge, London
- Comrey, A. L. (1951): Mental Testing and the Logic of Measurement. *Educational and Psychological Measurement*, 11(3), 323–334. <https://doi.org/10.1177/001316445101100301>
- Coombs, C. H. (1952): *A Theory of Psychological Scaling*. University of Michigan Press, Ann Arbor <https://doi.org/10.3998/mpub.9690146>
- Cox, E. P. (1980): The optimal number of response alternatives for a scale: a review. *Journal of Marketing Research*, 17, 407–422. <https://doi.org/10.1177/002224378001700401>

- Cox, N. (1996): Discussion of the Paper by Hand. *Journal of the Royal Statistical Society A*, 159(3), 480–481. <https://doi.org/10.1111/j.1467-985X.1996.tb00768.x>
- De Boer, J. (1995): On the History of Quantity Calculus and the International System. *Metrologia*, 31, 405–429. <https://doi.org/10.1088/0026-1394/31/6/001>
- Diamantopoulos, A. – Sarstedt, M. – Fuchs, C. – Wilczynski, P. – Kaiser, S. (2012): Guidelines for Choosing between Multi-Item and Single-Item Scales for Construct Measurement: A Predictive Validity Perspective. *Journal of the Academy of Marketing Science*, 40(5), 434–449. <https://doi.org/10.1007/s11747-011-0300-3>
- Douglas, V. (1947): The Measurable and Non-measurable. *Nature*, 4078, 908. <https://doi.org/10.1038/160908a0>
- Dusek T. (2006): Területi statisztika, valószínűségszámítás és statisztikai következtetésemélet. *Területi Statisztika*, 46(3), 223–239.
- Dusek T. (2013): On the misuse of probability theory in economics. *Tér Gazdaság Ember*, 1(3), 9–25.
- Euklidész (1983): *Elemek*. (ford.: Mayer Gyula), Gondolat, Budapest
- Fatalin L. (2017): A mérési skálák klasszikus elméletéhez. *Hadmérnök*, 12(2), 310–320.
- Feest, U. (2005): Operationism in Psychology: What the Debate is About, What the Debate Should be about. *Journal of the History of the Behavioral Sciences*, 41(2), 131–149. <https://doi.org/10.1002/jhbs.20079>
- Ferge Zs. – Schnell L. (1975): Az eredmények számszerű értékelése. In: Cseh-Szombaty L. – Ferge Zs. (a kötet összeállítói): *A szociológiai felvétel módszerei*. Közgazdasági és Jogi Könyvkiadó, Budapest, 316–340.
- Ferris, T. L. J. (2004): A new definition of measurement. *Measurement*, 36, 101–109. <https://doi.org/10.1016/j.measurement.2004.03.001>
- Göb, R. – McCollin, C. – Ramalhoto, M. (2007): Ordinal Methodology in the Analysis of Likert Scales. *Quality and Quantity*, 41, 601–626. <https://doi.org/10.1007/s11135-007-9089-z>
- Hartley, J. – Betts, L. R. (2010): Four layouts and a finding: the effect of changes in the order of the verbal labels and numerical values on Likert-type scales. *International Journal of Social Research Methodology*, 13(1), 17–27. <https://doi.org/10.1080/13645570802648077>
- Hay, B. (2020): A brief history of the thermal properties metrology. *Measurement*, 155, 107556. <https://doi.org/10.1016/j.measurement.2020.107556>
- Heine, S. J. – Lehman, D. – Peng, K. – Greenholtz, J. (2002): What's Wrong with Cross-Cultural Comparisons of Subjective Likert Scales?: The Reference-Group Effect. *Journal of Personality and Social Psychology*, 82, 903–918. <https://doi.org/10.1037/0022-3514.82.6.903>
- Hollerbach, T. (2023): *Sanctorius Sanctorius and the Origins of Health Measurement*. Springer, Cham <https://doi.org/10.1007/978-3-031-30118-6>
- Horváth Gy. (1991): *Az értelem mérése*. Tankönyvkiadó, Budapest
- Horváth Gy. (1997): *A modern tesztmodellek alkalmazása*. Akadémiai Kiadó, Budapest
- Hunyadi L. – Vita L. (2008): *Statisztika I*. Aula Kiadó, Budapest
- Hüttl A. (2003a): A gazdasági mérés történetéről. Adatok, elmélet, gazdaságpolitika. *Közgazdasági Szemle*, 50(2), 164–182.
- Hüttl A. (2003b): Hogyan vált a mérés (szinte) kizárólagos eszközzé a közgazdasági megismerésben? In: Bekker Zs. (szerk.): *Tantörténet és közgazdaságtudomány*. 272–286.
- Jamieson, S. (2004): Likert Scales – How to (Ab)Use Them. *Medical Education*, 38(12), 1217–1218. <https://doi.org/10.1111/j.1365-2929.2004.02012.x>

- Jones, W. P. – Loe, S. A. (2013): Optimal Number of Questionnaire Response Categories: More May Not Be Better. *SAGE Open*, April–June, 1–10. <https://doi.org/10.1177/2158244013489691>
- Kehl D. (2011): Skálák és statisztikák: a méréselmélet történetéről. *Statisztikai Szemle*, 78(10–11), 1057–1080.
- Kerékgyártó Gy. – Mundruczó Gy. (1991): *Statisztikai módszerek a gazdasági elemzésben*. Tankönyvkiadó, Budapest
- Korpás A. (1996): *Általános statisztika I.* Nemzeti Tankönyvkiadó, Budapest
- Koyré, A. (1968): *Metaphysics and Measurement*. Chapman & Hall, London
- Krantz, D. H. – Luce, R. D. – Suppes, P. – Tversky, A. (1971): *Foundations of Measurement. Volume I. Additive and Polynomial Representations*. Academic Press, San Diego
<https://doi.org/10.1016/B978-0-12-425401-5.50011-8>
- Kuhn, T. (1961): The Function of Measurement in Modern Physical Science. *Isis*, 52(2), 161–193.
<https://doi.org/10.1086/349468>
- Kun M. – Szegedi M. (szerk.) (1978): *Az intelligencia mérése*. Akadémiai Kiadó, Budapest
- Kuzon, W. M. – Urbanchek, M. G. – McCabe, S. (1996): The Seven Deadly Sins of Statistical Analysis. *Annals of Plastic Surgery*, 37(3), 265–272.
<https://doi.org/10.1097/00000637-199609000-00006>
- Labovitz, S. (1970): The Assignment of Numbers to Rank Order Categories. *American Sociological Review*, 35(3), 515–524. <https://doi.org/10.2307/2092993>
- Leahey, T. H. (1980): The Myth of Operationism. *Journal of Mind and Behavior*, 1(2), 127–143.
- Likert, R. (1932): A Technique for the Measurement of Attitudes. *Archives of Psychology*, 140, 5–55.
- Likert, R. (1934): A simple and reliable method of scoring the Thurstone scales. *Journal of Social Psychology*, 5, 228–238. <https://doi.org/10.1080/00224545.1934.9919450>
- Lord, F. M. (1953): On statistical treatment of football numbers. *American Psychologist*, 8, 750–751.
<https://doi.org/10.1037/h0063675>
- Lozano, L. M. – Carcía-Cueto, E. – Muniz, J. (2008): Effect of the Number of Response Categories on the Reliability and Validity of Rating Scales. *Methodology*, 4(2), 73–79.
<https://doi.org/10.1027/1614-2241.4.2.73>
- Mari, L. (2013): A quest for the definition of measurement. *Measurement*, 46, 2889–2895.
<https://doi.org/10.1016/j.measurement.2013.04.039>
- Mari, L. – Wilson, M. – Mau, A. (2023): *Measurement Across the Sciences. Developing a Shared Concept System for Measurement*. Springer, Cham <https://doi.org/10.1007/978-3-031-22448-5>
- Mau, S. (2019): *The metric society. On the quantification of the social*. Polity Press, Cambridge
- Maxwell, J. C. (1870): Presidential Address to Section A of the British Association. *Nature*, 2, 419–422.
<https://doi.org/10.1038/002419a0>
- Michell, J. (1993): The Origins of the Representational Theory of Measurement: Helmholtz, Hölder, and Russell. *Studies in History and Philosophy of Science Part A*, 24(2), 185–206.
[https://doi.org/10.1016/0039-3681\(93\)90045-L](https://doi.org/10.1016/0039-3681(93)90045-L)
- Michell, J. (1996): Introduction. *Journal of Mathematical Psychology*, 40, 235–237.
<https://doi.org/10.1006/jmps.1996.0023>
- Michell, J. (1997a): Introduction. *Journal of Mathematical Psychology*, 41, 345–346.
<https://doi.org/10.1006/jmps.1997.1178>
- Michell, J. (1997b): Quantitative science and the definition of measurement in psychology. *British Journal of Psychology*, 88, 355–383. <https://doi.org/10.1111/j.2044-8295.1997.tb02641.x>

- Michell, J. (2004): *Measurement in Psychology. A Critical History of a Methodological Concept*. Cambridge University Press, Cambridge
- Michell, J. – Ernst, C. (1996): Axioms of Quantity and the Theory of Measurement. Translated from Part I of Otto Hölder's German Text „Die Axiome der Quantität und die Lehre vom Mass”. *Journal of Mathematical Psychology*, 40, 235–252. <https://doi.org/10.1006/jmps.1996.0023>
- Morgenstern, O. (1963): *On the Accuracy of Economic Observations*. Princeton University Press, Princeton
- Morley, I. (2010): Conceptualising quantification before settlement: Activities and issues underlying the conception and use of measurement. In: Morley, I. – Renfrew, C. (eds.): *Archaeology of Measurement*. Cambridge University Press, Cambridge, 7–18. <https://doi.org/10.1017/CBO9780511760822.004>
- Morley, I. – Renfrew, C. (2010): *The Archaeology of Measurement*. Cambridge University Press, Cambridge <https://doi.org/10.1017/CBO9780511760822>
- Moscati, I. (2019): *Measuring Utility. From the Marginal Revolution to Behavioral Economics*. Oxford University Press, New York <https://doi.org/10.1093/oso/9780199372768.001.0001>
- Mosteller, F. – Tukey, J. W. (1977): *Data analysis and regression: A second course in statistics*. Addison-Wesley, Reading
- Muller, J. Z. (2019): *The Tyranny of Metrics*. Princeton University Press, Princeton
- Nagel, E. (1931): Measurement. *Erkenntnis*, 2, 313–335. <https://doi.org/10.1007/BF02028166>
- Nicholls, M. R. – Orr, C. A. – Okubo, M. – Loftus, A. (2006): Satisfaction guaranteed: The effect of spatial biases on responses to Likert scales. *Psychological Science*, 17, 1027–1028. <https://doi.org/10.1111/j.1467-9280.2006.01822.x>
- Norman, G. (2010): Likert scales, levels of measurement and the „laws” of statistics. *Advances in Health Sciences Education*, 15(5), 625–632. <https://doi.org/10.1007/s10459-010-9222-y>
- Patil, V. T. (2018): *Text Book of Metrology and Quality Control*. Nirali Prakashan, Shivaji Nagar
- Pléh Cs. (1992): *Pszichológiai történet*. Gondolat, Budapest
- Porter, T. M. (1995): *Trust in numbers: The pursuit of objectivity in science and public life*. Princeton University Press, Princeton <https://doi.org/10.1515/9781400821617>
- Preston, C. C. – Colman, A. M. (2000): Optimal number of response categories in rating scales: reliability, validity, discriminating power, and respondent preferences. *Acta Psychologica*, 104, 1–15. [https://doi.org/10.1016/S0001-6918\(99\)00050-5](https://doi.org/10.1016/S0001-6918(99)00050-5)
- Raghavendra, N. V. – Krishnamurthy, L. (2013): *Engineering Metrology and Measurements*. Oxford University Press, New Delhi
- Renfrew, C. – Morley, I. (2010): Introduction: Measure: Towards the construction of our world. In: Morley, I.–Renfrew, C. (eds.): *Archaeology of Measurement*. Cambridge University Press, Cambridge, 1–4. <https://doi.org/10.1017/CBO9780511760822.002>
- Rogers, T. B. (1989): Operationism in Psychology: A Discussion of Contextual Antecedents and an Historical Interpretation of its Longevity. *Journal of the History of the Behavioral Sciences*, 25, 139–153. [https://doi.org/10.1002/1520-6696\(198904\)25:2<139::AID-JHBS2300250204>3.0.CO;2-E](https://doi.org/10.1002/1520-6696(198904)25:2<139::AID-JHBS2300250204>3.0.CO;2-E)
- Rozeboom, W. W. (1966): Scaling Theory and the Nature of Measurement. *Synthese*, 16, 170–233. <https://doi.org/10.1007/BF00485356>
- Russell, B. (1903): *The principles of mathematics*. Cambridge University Press, Cambridge

- Schryvers, P. (2020): *Bad Data. Why We Measure the Wrong Things and Often Miss the Metrics That Matter*. Prometheus Books, Guilford
- Schwarz, N. – Knauper, B. – Hippler, H.-J. – Noelle-Neumann, E. – Clark, L. (1991): Rating scales: Numeric values may change the meaning of scale labels. *Public Opinion Quarterly*, 55, 570–582. <https://doi.org/10.1086/269282>
- Simms, L. J. – Zelazny, K. – Williams, T. F. – Bernstein, L. (2019): Does the Number of Response Options Matter? Psychometric Perspectives Using Personality Questionnaire Data. *Psychological Assessment*, 31(4), 557–566. <https://doi.org/10.1037/pas0000648>
- Smith, P. B. (2004): Acquiescent Response Bias as an Aspect of Cultural Communication Style. *Journal Of Cross-Cultural Psychology*, 35(1), 50–61. <https://doi.org/10.1177/0022022103260380>
- Stegmuller, D. (2011): Apples and Oranges? The Problem of Equivalence in Comparative Research. *Political Analysis*, 19(4), 471–487. <https://doi.org/10.1093/pan/mpr028>
- Stevens, S. S. (1935a): The Operational Basis of Psychology. *American Journal of Psychology*, 43, 323–330. <https://doi.org/10.2307/1415841>
- Stevens, S. S. (1935b): The Operational Definition of Psychological Concepts. *Psychological Review*, 42(6), 517–527. <https://doi.org/10.1037/h0056973>
- Stevens, S. S. (1946): On the Theory of Scales of Measurement. *Science*, 103, 677–680. <https://doi.org/10.1126/science.103.2684.677>
- Stevens, S. S. (1951): Mathematics, measurement, and psychophysics. In: Stevens, S. S. (eds.): *Handbook of Experimental Psychology*. Wiley, New York, 1–49.
- Stevens, S. S. (1957): On the Psychophysical Law. *The Psychological Review*, 64(3), 153–181. <https://doi.org/10.1037/h0046162>
- Stevens, S. S. (1959): Measurement, Psychophysics and Utility. In: Churchman, C. W. – Ratoosh, P. (eds.): *Measurement: Definition and Theories*. Wiley, New York
- Stevens, S. S. (1968): Ratio Scales of Opinion. In: Whitla, D. K. (eds.): *Handbook of Measurement and Assessment in Behavioral Sciences*. Addison-Wesley, Reading
- Swistik, P. (1990): Paradigms of measurement. *Theory and Decision*, 29(1), 1–18.
- Thorndike, E. L. (1918): The nature, purposes, and general methods of measurements of educational products. In: Whipple, G. M. (eds.) *Seventeenth yearbook of the national society for the study of education*. Vol. 2. Public School Publishing, 16–24. <https://doi.org/10.1177/016146811801900702>
- Thorndike, E. L. – Bregman, E. O. – Cobb, M. V. – Woodyard, E. (1926): *The Measurement of Intelligence*. Teachers College, Columbia University, New York <https://doi.org/10.1037/11240-000>
- Thurstone, L. L. (1928): The Absolute Zero in Intelligence Measurement. *The Psychological Review*, 35(3), 175–197. <https://doi.org/10.1037/h0072902>
- Torgerson, W. S. (1958): *Theory and Methods of Scaling*. Wiley, New York
- Velleman, P. F. – Wilkinson, L. (1993): Nominal, ordinal, interval, and ratio typologies are misleading. *American Statistician*, 47(1), 65–72. <https://doi.org/10.1080/00031305.1993.10475938>
- Vessonen, E. (2019): Operationalism and realism in psychometrics. *Philosophy Compass*, 14(2), e12624. <https://doi.org/10.1111/phc3.12624>
- Watkins, D. – Cheung, S. (1995): Culture, gender and response bias. *Journal of Cross-Cultural Psychology*, 26, 490–504. <https://doi.org/10.1177/0022022195265003>

- Weyl, H. (1949): *Philosophy of Mathematics and Natural Science*. Princeton University Press, Princeton
- Willits, F. K. – Theodori, G. L. – Luloff, A. E. (2016): Another Look at Likert Scales. *Journal of Rural Social Science*, 31(3), 126–139.
- Yule, G. U. – Kendall, M. G. (1964): *Bevezetés a statisztika elméletébe*. Közgazdasági és Jogi Könyvkiadó, Budapest
- Zerényi K. (2016): A Likert-skála adta lehetőségek és korlátok. *Opus et Educatio: Munka és Nevelés*, 3(4), 470–478. <https://doi.org/10.3311/ope.39>
- Zimmerman, D. W. – Zumbo, B. D. (1993): The Relative Power of Parametric and Nonparametric Statistical Methods. In: Keren, G. – Lewis, C. (eds.) *A Handbook for Data Analysis in the Behavioral Sciences. Methodological Issues*. Psychology Press, New York, 481–518.
- Zumbo, B. D. – Kroc, E. (2019): A Measurement Is a Choice and Stevens' Scales of Measurement Do Not Help Make It: A Response to Chalmers. *Educational and Psychological Measurement*, 79(6), 1184–1197. <https://doi.org/10.1177/0013164419844305>
- Zumbo, B. D. – Zimmerman, D. W. (1993): Is the Selection of Statistical Methods Governed by Level of Measurement? *Canadian Psychology*, 34(4), 390–400. <https://doi.org/10.1037/h0078865>