

A tesztfejlesztés és tesztadaptáció irányelvei

Elfogadta a Magyar Pszichológiai Társaság Tesztbizottsága
2020. december 31-én

Bevezető

Az MPT Tesztbizottsága (TB) 2020 áprilisában több, mint hétszáz Magyarországon használt pszichológiai vizsgáló eszköztől / tesztől (a továbbiakban: tesztől) szóló publikációt tart nyilván. A kiválasztásra, diagnosztizálásra és kutatásra használt tesztek jogi státusza és szakmai minősége részben történelmi okokból meglehetősen vegyes képet mutat. A TB állásfoglalása szerint különösen a kiválasztásra és diagnosztizálásra használt tesztek hivatalos jogi és szakmai minősítése elengedhetetlen. A tesztek minősítésének kritériumait és minősítési eljárását a TB kidolgozta és közzétette¹, de a minősítési folyamat egy szakmai társaság kereteiben nem valósítható meg. Átmeneti megoldásként a TB jelen irányelveiben röviden megfogalmazza, hogy bármely saját fejlesztésű vagy más nyelvekről adaptált pszichológiai tesztnek milyen kritériumoknak² kell megfelelnie ahhoz, hogy szakmailag elfogadható módon alkalmazható legyen. A kritériumok teljesülésére vonatkozó információk a tesztkönyvben, vagy más publikált anyagban, tanulmányban, folyóiratcikkben stb. a felhasználók számára hozzáférhetőnek kell lenniük. Ha ezek az információk hiányoznak, vagy nem hozzáférhetők, akkor a teszt minősége nem megítélhető, így alkalmazása vagy nem ajánlott, vagy csak megszorításokkal, amelyeket minden, a teszten alapuló szakvéleményekben és publikációban fel kell tüntetni.

A TB azt ajánlja, hogy a tesztfejlesztők és tesztadaptálók a továbbiakban e kritériumoknak megfelelően dolgozzák ki, dokumentálják és publikálják tesztjeiket. A TB azt is szorgalmazza, hogy a szakmai szekciók és műhelyek gondoskodjanak arról is, hogy a kiválasztási és diagnosztikai célokra jelenleg is használt tesztek, e kritériumoknak megfelelően frissítsék fel. A TB szorgalmazza továbbá, hogy ezen felülvizsgálati folyamatok megvalósulását pályázatok kiírásával az állami és szakmai szereplők támogassák.

A pszichológiai tesztek

A továbbiakban *tesztnek* a teljes vizsgáló/mérőeszközt nevezzük, amely egy vagy több pszichológiai konstrukciót (feltételezett vonást, képességet, attitűdöt stb.) mérő *tesztskálát*³ tartalmazhat. Nagyobb vizsgáló rendszerek több tesztet, s ezeken belül több tesztskálát is tartalmazhatnak. Ilyen esetekben a felhasználhatóságot mindegyik tesztre és tesztskálára külön-külön kell ellenőrizni. Ahhoz, hogy egy tétel- vagy item-készlet (ami lehet szöveges-, manuális-, képi- vagy akár video-tételek összessége) pszichológiai tesztskálának minősüljön, az alábbiakban meghatározott tesztfejlesztési vagy tesztadaptálási folyamaton kell keresztülmennie, és egy sor, az alábbiakban felsorolt kritériumnak kell megfelelnie.

¹ <http://mpt.hu/a-tarsasagrol/szervezet/bizottsagok/tesztbizottsag/tesztminosites/>

² Az alább felsorolt szempontok jórészt a pszichológus képzés tananyagában szerepelnek (ld. a referencialistát), ezért ezeket nem fejtjük ki részletesen.

³ A hazai szakmai szóhasználatban egyaránt tesztnek nevezik az egy konstruktumot mérő egyskálás mérőeszközöket, akár csak a több skálából álló, több konstruktumot mérő eszközöket. Ez utóbbiakat olykor leltárnak (inventory) vagy battériának is nevezik. Ezek voltaképpen annyi tesztet jelentenek, ahány mérőskálából állnak. A továbbiakban a teszt szót az eszköz megnevezésére használjuk, az egyes konstruktumokra kidolgozott feladat-, állítás- vagy képanyagsorokat pedig tesztskálának nevezzük.

A tesztfejlesztés és tesztadaptálás szakaszai

- I. Definiálás (1-5. pont): a teszt beazonosíthatóságát biztosító és a teszt alapvető jellemzőit tartalmazó leírás
- II. A tételek megalkotása, illetve a tételek lefordítása és visszafordítása,
- III. Előtanulmány (pilot study): tételelemzés (skálahomogenitás- vagy belső reliabilitás vizsgálata, tesztkitöltés-elemzés, tételkorrekciók
- IV. Reliabilitás: a teszt megbízhatóságának meghatározása
- V. Validitás: a teszt érvényességének meghatározása
- VI. Standardizálás: a teszt normatív átlagainak meghatározása
- VII. Egyéb tételelemzés: tesztkitöltési validitás, tételnehézség, tételdifferenciáló-képesség stb. ellenőrzése

A pszichológiai tesztekkel szemben támasztott szakmai követelmények (a tesztek kritériumai)⁴

Ahhoz, hogy a pszichológus megfelelő döntést tudjon hozni valamely teszt alkalmazásáról, az alábbi információkkal kell rendelkeznie:

1. A teszt egységes szerkezetű elnevezése
 - a. A teszt eredeti elnevezése: pl. Johnson Number Evaluation Test
 - b. A teszt eredeti rövidítése: pl. JNET
 - c. Adaptált teszt Magyarországon használatos neve:
 - i. a szerző nevének benne foglalásával (kerülendő a -féle alkalmazása): pl. Johnson Számértékelés Teszt (*nem* Johnson-féle Számértékelés Teszt)
 - ii. a szerző neve nélkül: Számértékelés Teszt
 - d. Adaptált teszt magyar nevének rövidítése:
 - i. Az eredeti rövidítés meghagyása -H kiegészítő jelöléssel: pl. JNET-H
 - ii. Amennyiben több magyar verzió készül az adaptáció során, akkor a változatok jelölésére a -H1, -H2 stb. kiterjesztés javasolt: pl. JNET-H1, JNET-H2 stb.
2. Az eredeti, idegen nyelvű teszt azonosítása
 - a. a teszt szerzőjének/szerzőinek neve
 - b. a teszt első megjelenésének/kiadásának éve
 - c. a teszt jelen/adaptált verziója megjelenésének/kiadásának éve
 - d. a teszt jelen/adaptált verziója kiadójának és forgalmazójának neve és száma
 - e. a teszt alternatív (párhuzamos) változata(i) (ha van(nak) ilyen(ek))
3. A teszt magyarországi adaptációját és használatát engedélyező dokumentumra vonatkozó információk
 - a. szerzői engedély: kibocsátója, aláírója és dátuma
 - b. kiadói engedély (licence): az irat kibocsátója, aláírója és dátuma
 - c. engedély hatálya és tartalma
4. A teszt általános leírása
 - a. A teszt által *konkrétan vizsgált/mért konstrukciók (változók) megnevezésén* túl e fogalmi konstrukciók tesztre specifikus fogalmi és empirikus *definícióját*⁵ is szükséges megadni
 - b. A tesztfelhasználás célja: diagnózis (pl. klinikai, tanácsadási, iskolai stb.), munkapszichológiai kiválasztás, kutatás/tesztfejlesztés
 - c. A teszt jellege: kvantitatív (pszichometrikus skála), kvalitatív (projektív teszt, tartalomelemzés)

⁴ Készült az EFPA (2005) Review model for the description and evaluation of psychological tests című tesztértékelő sémája szempontjainak figyelembevételével.

⁵ Pl. a *szorongást* mint mérendő pszichológiai konstrukciót (fogalmat) a tesztalkotó hogyan képzei el (konceptualizálja) és milyen tapasztalati (verbális, viselkedéses és/vagy fizikai) adatok révén véli azonosíthatónak.

- d. A teszt által vizsgált pszichológiai funkciók: intellektus, észlelés, figyelem, verbális képesség, numerikus képesség, téri képesség, mozgás, gondolkodás, kognitív stílus, emlékezet, érzelem, motiváció, érdeklődés, érték, állapot, személyiségvonás, vélekedés-hiedelem, családműködés, csoportműködés, szervezeti működés, munkavégző viselkedés, pszichológiai zavarok, klinikai diagnózis, neuropszichológiai diagnózis, addikció (kémiai), addikció (viselkedéses), iskolai teljesítmény, iskolai oktatás-nevelés, egyéb:
 - e. A teszt alkalmazási területei: klinikum/egészségügy, gyógypedagógia, neuropszichológia, igazságügy, oktatás-nevelés, munka/szervezet, közlekedés, tanácsadás-pályaválasztás, sport-szabadidő, fejlődés, egyéb:
5. A teszt technikai leírása
- a. A tesztfelvétel módja
 - i. egyéni interaktív
 - ii. felügyelt csoportos papír-ceruza
 - iii. felügyelt helyi számítógépre telepített vagy online alkalmazás
 - iv. nem felügyelt/önkitöltős online alkalmazás
 - v. Egyéb:
 - b. A tesztfelvétel instrukciója
 - i. az instrukció szövege legyen rövid, lényegre törő, a magyar köznyelv szabályainak megfelelő, a szaknyelvi kifejezéseket kerülő, a teszt célcsoportja által jól érthető
 - ii. tesztadaptáció esetén az eredeti instrukció fordítandó le, ügyelve a magyar nyelv szabályaira, a nyelvhelyességre és közérthetőségre.
 - iii. a lefordított instrukció – akárcsak a teszt tételei – visszafordítással (back translation) ellenőrizendő
 - 1. a visszafordítást az eredeti szöveget nem ismerő, az eredeti nyelvet anyanyelvi szinten ismerő, lehetőleg kétnyelvű személy végezze
 - 2. a visszafordítás ellenőrzését és a szükséges korrekciókat az eredeti szerzőtől kell kérni. Ha az eredeti szerző nem elérhető, akkor a visszafordítást lehetőleg anyanyelvű pszichológussal kell ellenőriztetni.
 - c. A teszt tételeinek típusa (a válaszadás módja):
 - i. nyílt kérdések
 - ii. többszörös választás (ahol a válaszalternatívák egy skálához tartoznak)
 - iii. ellentétes melléknévpárok (kétpólusú skálához)
 - iv. Likert-típusú válaszkála
 - v. kényszerválasztás (ahol a válaszalternatívák különböző skálákhoz tartoznak)
 - vi. többszörös választás (ahol a válaszalternatívák különböző skálákhoz tartoznak) (ipszatív)
 - vii. melléknév párok (ahol a válaszalternatívák különböző skálákhoz tartoznak),
 - viii. Egyéb:
 - d. A tesztfelvétel/adminisztráció módja:
 - i. szóbeli interjú
 - ii. papír-ceruza
 - iii. manuális
 - iv. számítógépes
 - v. Egyéb:
 - e. A tesztvizsgálat és értékelés időkerete:
 - i. előkészítés ideje
 - ii. tesztfelvétel
 - 1. időkorlát nélküli
 - 2. időkorlátos
 - 3. válaszsebesség mérését igényli
 - iii. kiértékelés ideje
 - iv. értelmezés ideje
 - v. visszajelzés (riport)
 - f. A teszt kiértékelése
 - i. számítógépen adott válaszok közvetlen számítógépes kiértékelése

- ii. papíron kitöltött válaszok kézi vagy szkenneres bevitele és számítógépes kiértékelése
 - iii. egyszerű kézi kiértékelés, megoldó kulccsal – ehhez csak adminisztrátori képesítés szükséges
 - iv. kritérium alapú kézi kiértékelés (jelölés) – ehhez a tesztértékelési kiképzésre van szükség
 - v. teszt-kiértékelési szolgáltatás igénybevétele – pl. a tesztforgalmazó/kiadó végzi a kiértékelést
 - vi. Egyéb:
6. A tesztskálák mérési megbízhatóságának (**reliabilitás**) ellenőrzése
- a. A megbízhatóság-mutatók azt vizsgálják, hogy egyrészt az adott pszichológiai konstruktum (pl. szorongás) mérésére szolgáló tesztskála tételei (pl. állítások vagy feladatok) milyen mértékben mérik ugyanazt a konstruktumot (azaz a mérőskála mennyire *konzisztens*), másrészt a mérés *megisméltése* ugyanazzal az eszközzel mennyire ad hasonló eredményeket, harmadrészt a mérő skálával végzett mérések eredményei mennyire adnak hasonló eredményeket egy más tételből összeállított, de ugyanazt a konstruktumot mérő, ún. *ekvivalens* skálával („párhuzamos” tesztel – ha van ilyen) kapott eredményekkel.
 - b. A tesztskála belső reliabilitás-értéke (pl. Cronbach-alfa) meghatározásának nyilvánvalóan csak az azonos konstruktum mérését célzó tételből alkotott skála és nem a többskálás teszt egészének esetében van értelme.
 - i. A több konstruktumot mérő, összetett tesztek (tesztbattériák) esetében a reliabilitás-értékeket minden egyes skálára külön kell meghatározni.
 - ii. Amennyiben több skálából képezünk valamilyen képlettel összetett mutatót (indexet), akkor az összetett skála reliabilitását külön is meg kell határozni.
 - iii. Tartalmilag széles tartományú vonások vagy összesített skálák esetén a belső konzisztencia értékek helyett inkább az ismételt teszteléses (teszt- reteszt) eljárás releváns.
 - iv. Állapotméréskor viszont az ismételt méréses megbízhatóság helyett nagyobb hangsúly tevődik a belső konzisztencia értékekre.
 - c. Fontos, hogy a reliabilitás- értékek lehetőleg több különböző, de legalább két egymástól független csoportnál/mintában álljanak rendelkezésre. A minták méretének külön-külön 100-200 főt el kell érnie.
 - d. A reliabilitás-együttható elfogadására vonatkozóan nincsenek matematikailag rögzített követelmények. A reliabilitás-értékek elfogadását illetően az alábbi szempontokat kell figyelembe venni:
 - i. A reliabilitás együttható, leggyakrabban a Cronbach-alfa értéke, *általában* nem elfogadható, ha kisebb, mint 0.7, elfogadható, ha 0.7 és 0.79 közötti, jó szintű, ha 0.8 és 0.89 közötti, és kiváló, ha 0.9 feletti. A 0.9 feletti értékek esetén azonban a skálatételek esetlegesen túlzott tartalmi homogenitása ellenőrizendő (l. a iv. pontot)
 - ii. A Cronbach-alfa érték függ a skálát alkotó tételek számától: 15 feletti tétel számú skálák esetén a minimálisan elfogadható érték 0.8. Egészen alacsony (3-4) tétel számú skálák esetén már 0.6 fölötti érték is megengedett, bár az ilyen rövid skálák mérési megbízhatósága alapvetően megkérdőjelezhető.
 - iii. Fontos az adott konstruktumot mérő skálába tartozó tételek tartalmi heterogenitása, azaz a tételek a vizsgált/mért jelenséget több oldalról, több viselkedéses/gondolkodási/érzelmi formában és többféle helyzetben írják le.
 - iv. Irreálisan magas lehet a belsőkonzisztencia-érték (pl. Cronbach-alfa), ha a tételek „túlzottan egysíkúak” (bloated specific), amikor ugyanarra a specifikus tartalomra többször, csak némileg eltérő megfogalmazású tétel (pl. rokonértelmű szavakkal) kérdez rá. (A kérdések fölös megisméltése ráadásul a vizsgált személyeket is fölöslegesen terheli, ami esetlegesen a teszteléssel szembeni ellenállást is kiválthatja.)
 - v. Tesztadaptáció esetén az általános elvárás az, hogy a hazai mintán kapott pszichometriai mutatók megközelítsék az eredeti teszt megfelelő mutatóinak értékét. Igényesebb megoldás, ha ennek statisztikai ellenőrzése is megtörténik.

- vi. Az *ismételt mérés*es vizsgálat eredménye nem elfogadható, ha teszt-reteszt korrelációs együtthatók mértéke kisebb, mint 0.6 (azaz nem éri el a közepes hatásméretet), elfogadható, ha ez az érték 0.6 és 0.69 közötti, jó szintű, ha 0.7 és 0.79 közötti és kiváló, ha 0.8-nál nagyobb.
- e. Kvalitatív eljárások megbízhatóságának ellenőrzésére ugyanazon megfigyelési anyag (szabad verbális és nem verbális viselkedés, pl. rajzok) a párhuzamos, egymástól független pontozói vagy kategorizációs kiértékelések összehasonlítása (inter-rater agreement) alkalmazandó. A pontozók közötti egyetértés mértékének értelmezésére jelenleg nem állnak rendelkezésre annyira egzakt határértékek mint a teszt-tételek reliabilitása (pl. a Cronbach-alfa; ld. fent) esetében, ezért a konkrét értékek interpretálása során különösen fontos a szakirodalmi háttér és a hasonló vizsgálatok bemutatása és eredményeik bevonása az értelmezési folyamatba.

7. A tesztskálák érvényességének (**validitás**) ellenőrzése:

- a. empirikus adatok szolgáltatása arról, hogy a tesztskála tételei valóban azt a konstruktumot (feltételezett látens változó, pl. szorongás vagy intelligencia) mérik-e, amire a tesztskálát tervezték és amire a tesztskála elnevezése is utal. A tesztskála voltaképpen a szóban forgó konstruktum egy lehetséges empirikus/mérési definícióját képezi.
- b. A validitás a tesztadatokat olyan empirikus adatokkal való összefüggéseinek bemutatását jelenti, amelyek a tesztől függetlenül keletkeztek.
- c. Fontos, hogy validitás vizsgálatára elméletileg megalapozott, szisztematikus kutatási terv készüljön.
- d. *Tartalmi validitás*
 - i. A tartalmi validitás annak szisztematikus ellenőrzését igényli, hogy a skála tételei mennyire *reprezentálják* az adott konstruktum viselkedéses, kognitív, motivációs, emocionális megnyilvánulásait. A tesztkönyvnek tartalmaznia kell annak leírását, hogy a tesztalkotók milyen módszerrel biztosították a skálatételek adekvát és reprezentatív kiválogatását.
 - ii. A tartalmi validitás vizsgálatának előfeltétele a konstruktum szakirodalomra alapozott elméleti és empirikus definiálása. Ha új pszichológiai konstruktumról van szó, akkor különösen ügyelni kell a konstruktumot azonosító és megkülönböztető empirikus jegyekre (amelyek alapján az adott konstruktum megnyilvánulása bárki számára felismerhető vagy azonosítható).
 - iii. Amennyiben egy konstruktum elméleti és empirikus definíciója nem áll rendelkezésre és a cél bizonyos specifikus csoportok (pl. szorongók) azonosítása, akkor az *empirikus tesztszerkesztés* elveit kell követni: azokat a tételeket beválogatni a skálába, tekintet nélkül a tételek tartalmára, amelyekre a vizsgált célcsoport (példánkban a tesztől függetlenül diagnosztizált szorongó személyek) eltérően válaszolnak a nem szorongó csoportokhoz képest.
 - iv. A lehetséges tételek készletét (item pool) lehetőleg több szakértőből álló munkacsoport állítsa össze. Fordítás esetén a lefordított tételeket ugyancsak több szakértőből álló munkacsoport ellenőrizze.
 - v. Az elkészült tételsort több szakértő ellenőrizze, hogy kellően heterogén legyen, ne ismétlődjenek megnyilvánulási módok, ugyanakkor ne maradjanak ki fontos megnyilvánulási területek sem a tételek közül. Fordítás esetén olyan tesztet érdemes lefordítani, amelynek a létrehozásakor ezek a feltételek teljesültek.
 - vi. Külön szakmai megfontolás tárgya, hogy a tesztskála tételei mennyire legyenek „átlátszóak”, azaz evidensek a vizsgált személyek számára (ún. látszatvaliditás vagy face validity). Az etikai alapelvek mellett szólnak, hogy a vizsgált személyek számára is világos legyen a tételek jelentése, azaz tudják, hogy mit mér a teszt. Ettől indokolt esetben el lehet térni, amikor a személyek választát/teljesítményét adott helyzetben jelentősen befolyásolhatja vagy torzíthatja a személyek szociális megfelelési igénye, a disszimulációs vagy szimulációs tendenciája, vagy a kutatási cél feltárása.
- e. *Kritérium validitás*: annak ellenőrzése, hogy a teszteredmények milyen hatékonyan képesek a személyek konkrét/valós helyzetben mutatott viselkedését/teljesítményét (kritérium) megbecsülni. Ilyen kritérium lehet bármilyen teljesítmény vagy viselkedés-

mutató: iskolai osztályzat, versenyeredmény, előléptetési/minősítési fokozat, tanári/gondozói minősítés, pszichiátriai vagy gyógypedagógiai diagnózis stb.

- i. Konkurens (egyidejű) validitás
 1. A tesztfelvétel és a kritérium adatai időben közel, de egymástól függetlenül keletkeznek. (Pl. a tesztfelvétel napján végeznek viselkedés-megfigyeléseket, vagy ugyanaznap zajlik egy vizsga, amivel a teljesítményteszt eredményei korrelálthatók stb.)
 - ii. Prediktív (előre jelző) validitás
 1. A tesztfelvétel előbb történik, mint a kritérium adatainak felvétele.
 2. Fontos, hogy a kritérium-mutató mérése „minőségileg” azonos helyzetben történjen. Ha egy tesztmutatót például a munkahelyi beválás sikerességének előrejelzésre kívánunk használni, akkor a tesztet mindenképpen a kiválasztás helyzetében és időpontjában kell felvenni. Tehát nemcsak az számít, hogy a tesztelés és a kritérium mérése között mekkora az időintervallum, hanem a mérés körülményeinek is hasonlóknak kell lenniük a teszt majdani alkalmazásához.
 - iii. Posztdiktív validitás
 1. Ebben az esetben egy régebben fejlesztett teszt újonnan átdolgozott vagy rövidített változatát vetik össze az azonos konstruktmura korábbi és/vagy hosszabb tesztjének eredményeivel.
 - iv. Ellentétes (kontraszt) csoportok módszere
 1. A tesztet olyan csoportokkal vesszük fel, amelyek teszteredményeiben elméletileg megalapozott eltérések várhatók el. Pl. különbség várható egy finom motorikát vizsgáló ügyességi teszt eredményeiben a műszaki rajzoló és az énekesek véletlenszerűen kiválasztott csoportjában.
- f. Az ún. *konstruktum-validitás* annak kimutatására szolgál, hogy a tesztskála milyen mértékben fogadható el valamely elméletileg meghatározott konstruktm mérőeszközüél. Az eljárás az elméleti konstruktm és az azt mérő tesztskála, illetve más elméleti konstruktmok és az azok mérésére kidolgozott eszközök korrelációs elemzésén alapul. Az elemzés lényege, hogy az elméletileg összefüggésbe hozható konstruktmokat mérő eszközökkel kapott korrelációknak tükrözniük kell az elméleti elvárásokat. Amennyiben két elméletileg összefüggő konstruktmot mérő tesztskálák nincsenek korrelációban egymással, akkor megkérdőjelezhető, hogy egyik vagy mindkét a skála megfelelően méri-e a saját konstruktmát. Mivel az sem zárható ki, hogy a két konstruktm összefüggésére vonatkozó elmélet hibás, ezért az eljárás annál megbízhatóbb, minél több konstruktmot és azok mérőeszközeit vonjuk be az elemzésekbe (ún. multi trait, multi method eljárás).
- i. Konvergens validitás: elméletileg összefüggő konstruktmokat mérő skáláknak korrelálniuk kell egymással. Ha az elméleti viszony negatív, akkor negatív korreláció az elvárás. A korreláció elvárható mértéke szempontjából a 7j. pont szerinti határértékeket lehet figyelembe venni (a korrelációs együtthatókat abszolút értéken értve).
 - ii. Diszkriminációs (divergens) validitás: az elméletileg egymástól független konstruktmokat mérő skálák korrelációja *zéró* körüli, azaz nem szignifikáns. (Fontos, hogy *nem* a negatív korreláció, hanem a nulla körüli korreláció az elvárás!)
 - iii. A konstruktm-validitás értékeire vonatkozó irányelvek rugalmasan értelmezendők.
 1. Két hasonló eszköz egyidejűleg nyert adatainak korrelációja esetén a 0.60, vagy annál magasabb érték „elfogadható”.
 2. Azokban az esetekben azonban, amikor a konstruktmok kevésbé hasonlóak, vagy a tesztfelvételek bizonyos időintervallum eltéréssel történnek, ott alacsonyabb korrelációs értéke (0,4) is elfogadható.
 - iv. A konstruktm-validitás értékelésekor a nagyon magas korrelációs értékeket óvatosan kell kezelni. A 0.90 feletti korrelációs értékek arra utalnak, hogy a skálák nagy valószínűséggel pontosan ugyanazt a tényezőt mérik (csak esetleg más néven). Ez akkor nem probléma, ha egy már bevált tesztskálát egy új, azzal ekvivalens skálával szeretnénk helyettesíteni, viszont gondot okozhat abban az

esetben, ha azt gondoljuk, hogy az új skála ettől némileg eltérő, további hasznos információt is szolgáltat, vagyis nem teljesen egybevágó eszközökről van szó.

- g. Fontos, hogy több validitás-érték lehetőleg több különböző, de legalább két egymástól független mintában álljon rendelkezésre. Diagnosztikai célra alkalmazott tesztek esetében egy vagy több nagymintás, de legalább két-három fajta, közepes méretű mintán végzett validitás-vizsgálatra van szükség.
 - h. A validitás vizsgálata során a mintaméretre vonatkozó irányelvnek azt a mintanagyságot tekintjük, mely egy erő-elemzésén alapuló, közepes mértékű validitás (ha egyáltalán létezik) eléréséhez szükséges. A minták méretének minimálisan 100-200 főt el kell érnie. Közepes méretű mintának a 300 főnél nagyobb mintát vehetjük, nagy mintának pedig el kell érnie az 1000 főt.
 - i. Meg kell jelölni, hogy a validitás-vizsgálatban részt vevő minta milyen populációra (demográfiai csoportra) reprezentatív, vagy inkább esetleges, könnyen hozzáférhető, kényelmi mintavétel történt-e. Azt is meg kell említeni, hogy a tesztszerkesztők/adaptálók hogyan gondoskodtak a véletlen (random) mintavételről.
 - j. A tesztek, skálák közötti korreláció bizonyos mértékig függ a mérendő látens konstruktum robusztusságától, stabilitásától is (pl. intelligencia esetén inkább erősebb, míg pl. hangulatok mérésekor inkább gyengébb összefüggés várható), de a lényegében azonos, vagy nagyon hasonló tesztek skálái esetén nagyjából ezeket az értékeket várjuk el:
 - i. Nem elfogadható szintű ($r < 0.55$)
 - ii. Elfogadható szintű ($0.55 < r < 0.65$)
 - iii. Jó szintű ($0.65 < r < 0.75$)
 - iv. Kiváló szintű ($r > 0.75$)
 - k. A teszt skáláinak korrelációja a kritériumokkal:
 - i. Nem elfogadható szintű ($r < 0.2$)
 - ii. Elfogadható szintű ($0.2 < r < 0.35$)
 - iii. Jó szintű ($0.35 < r < 0.50$)
 - iv. Kiváló szintű ($r > 0.50$)
 - l. Mivel a kvalitatív eljárások a mennyiségi (intervallum vagy arányskálák) helyett besoroláson, megállapításon, kategorizáción alapuló adatokat szolgáltatnak, ezért a teszt kategóriáinak tartalmi, kritérium, konstruktum validitásának ellenőrzése nem korrelációs elemzésekkel, hanem a megállapítható adatokra alkalmas kontingencia-táblákon végzett függetlenség-vizsgálatokkal (χ^2) lehetséges.
 - i. A tartalmi validitás ellenőrzése lényegében megegyezik a kvantitatív elemzésekkel (lásd 7. i-vi).
 - ii. A kritériumvaliditás vizsgálata esetén a kvalitatív teszttel végzett besorolásokat vetjük össze a tesztől független kritériumon (pl. viselkedéses megfigyelésen, klinikai/pszichiátriai/gyógyypedagógiai diagnózison) alapuló besorolásokkal (pl. a kvalitatív eljárással szorongónak minősített személyek szorongónak minősülnek-e a viselkedés megfigyelése, a más kvalitatív/kvantitatív tesztek eredményei és/vagy a pszichiátriai vizsgálat alapján szorongónak minősített személyekkel)
 - iii. Konstruktum-validitás vizsgálata esetén azt kell ellenőrizni, hogy a kvalitatív besorolások összhangban vannak-e a kategóriák viszonyára vonatkozó elméleti összefüggésekkel (ha pl. az az elméleti állítás igaz, hogy a szorongás csökkenti a teljesítménymotivációt, akkor pl. TAT alapján megállapított szorongás és teljesítménymotivációs adatoknak tükrözniük kell ezt az összefüggést).
8. **Standardizálás:** normatáblák kidolgozása különböző referenciacsoportokra
- a. A tesztnormák (standardok) arra szolgálnak, hogy a személyek vagy csoportok teljesítményét annak a populációnak becsült teljesítményéhez vagy válaszaihoz viszonyíthassuk, amelyhez a személy vagy a csoport tartozik.
 - b. A normatív minta (normacsoport) kiválasztásánál a célpopulációra való reprezentativitásra kell törekedni, és kerülni kell az esetleges, könnyen hozzáférhető, kényelmi mintavételt.
 - c. Az összehasonlításra szolgáló norma- vagy standard csoportok szocio-demográfiai jellemzőit, összetételét – a nemek arányát, életkori, iskolázottsági eloszlását, lakhelyének jellegét (pl. város, falu), etnikumát stb. – pontosan dokumentálni kell.
 - d. Az általános populációra vonatkozó, reprezentatív normák esetén a vizsgálati mintának kellően nagyoknak, és a vizsgálni kívánt populációra reprezentatívnak kell lennie. (A minta nagysága önmagában nem elégséges garancia a reprezentativitáshoz).

- i. A 150 főnél kisebb mintaméret túl kicsinek számít, mivel így az eloszlásgörbe szélső értékeire túl kevés eset kerül.
 - ii. Országos reprezentatív minta min. 1000 fő.
 - iii. Szűkebb csoportnormák, pl. foglalkozás-specifikus normák esetében a legkisebb mintaméret csoportonként érje el a 150-300 főt.
 - iv. Amennyiben a teszteredményekben a különféle szociodemográfiai szegmentációk (pl. nemek, etnikumok stb.) mentén szignifikáns és közepes vagy nagy hatásmértékű (effect-size) különbség mutatkozik, akkor minden érintett csoportra külön normatáblákat kell készíteni.
- e. Verbális tesztek adaptációja esetén általában nem használható a külföldi vagy nemzetközi norma és a standardizációt megfelelő magyar mintán el kell végezni. Az azonos nyelvű, de másik országból származó normák kisebb mintán való ellenőrzés után felhasználhatók hazai mintán is.
- f. A nem verbális funkciókat mérő tesztek esetében, a külföldi normatáblák - kisebb mintán való ellenőrzést követően - felhasználhatók.
- g. A normatív adatokhoz való hasonlítás és a teszt által mért különböző konstruktumok egymással való összehasonlítása érdekében célszerű vizuálisan szemléletes tesztprofil kialakítani az alábbi standard skálák valamelyikének alkalmazásával. Tesztadaptáció esetén az eredeti standardizációs eljárást célszerű követni.
 - i. Standard skálák
 - 1. z-értékek
 - 2. IQ deviációs kvóciens (pl. a Weschler teszt esetén a középérték=100, szórás=15, míg a Stanford-Binet teszt esetén a szórás = 16)
 - 3. T-pontok (középérték = 50, szórása = 10)
 - 4. Sten-pontok (átlaga = 5,5, szórása = 2);
 - 5. Stanine-pontok (range: 1-9; átlaga = 5, szórása = 2);
 - 6. C-pontok (range: 0-10; átlaga = 5; szórása = 2)
 - ii. Gyakorisági övezetek szerinti felosztása
 - 1. 5-fokú besorolás: 10:20:40:20:10 centilis tartományokkal.
 - 2. Decilisek vagy egyéb percentilis alapú besorolás
 - 3. Szórás vagy valószínűségi hiba (pl. a MAWI esetében)
- h. Kvalitatív eljárások esetén a válaszok és az összkép kategóriákba sorolásához, a diagnózishoz alapvetően nincs szükség normatáblákra, a besorolás elméleti megfontolásokra és klinikai tapasztalatokra épül, és a személy válaszait a saját többi válaszával együtt értelmezzük. Szükséges azonban a normaértékek meghatározása, ha a válaszok kategóriába sorolása gyakorisági adatokon alapul:
 - i. Például egy bizonyos választ (FHD) azért tekintjük a szorongás jelének, mert független szorongó mintában gyakrabban előfordul, mint a nem szorongók csoportjában
 - ii. Vagy egy választ azért tekintünk Vulger-nek, mert a populációban gyakran előfordul. Az már külön elméleti megfontolás kérése, hogy az ilyen gyakori választ milyen elméleti keretben értelmezzük, és milyen diagnosztikus kategóriához kapcsoljuk.
- 9. A tesztskálák tételei nehézségi vagy differenciálási jellemzőinek ellenőrzése (*ajánlott*):
 - a. A differenciális tétel-működés (differential item functioning, DIF) ellenőrzése lényegében azt biztosítja, hogy az egyes tesztskála-tételek (itemek) működése nem torzul számottevően különféle populációs csoportok esetén. A cél az, hogy a teszt által mérendő pszichológiai jellemző szempontjából azonos pozíciójú személyek ugyanakkora valószínűséggel válasszák minden egyes tétel bizonyos válaszlehetőségeit - kulturális, demográfiai csoportba tartozásuktól függetlenül. Vagyis a tételek nehézségi szintje vagy diszkriminációs képessége ne legyen eltérő a különböző (al)csoportoknál (pl. az ugyanolyan készségi szinten lévő csoportok fiúk és lányok számára az adott tétel ugyanolyan nehézségű legyen a tételre adott válaszgyakoriságok alapján). Ez a kritérium tulajdonképpen a teszt „kulturális igazságossági” követelményét szolgálja.
 - b. A tétel-válasz elmélet (item response theory, IRT) alapján a tesztskála-tételek (itemek) valószínűségi változók, amelyek együttesen függenek a tétel paramétereitől és a vizsgált személy jellemzőitől. Egy skálatétel akkor alkalmas mérésre (összehasonlításra), ha a mérendő látens változón („képességen”) azonos pozíciót elfoglaló személyeknél a tételre

adott válasz valószínűsége azonos, hiszen enélkül a tétel eltérő jelentést hordozna a két csoport számára, vagyis a tétel torzítana (,item bias’).

10. Tesztkitöltési validitás ellenőrzése (ajánlott)

- a. Ahhoz, hogy érvényes és megbízható eredményekkel szolgáljon, egy mérőeszköz rendelkezhet olyan beépített indikátorokkal, melyek a válaszadás őszinte jellegével kapcsolatban szolgáltatnak információt. Ezen mutatók változatos módszertanok alkalmazásával vagy tartalmi (pl. benyomás-keltési törekvés mérése) vagy formai (pl. választendencia, vagy válaszdő mérés) alapon igyekeznek kiszűrni azokat a válaszadói torzításokat, melyek akár tudatosan, akár nem-tudatos módon eltorzítanák a felméréndő valódi értékeket. A mérőeszköz adaptálásakor ugyanúgy ellenőrizni szükséges ezen indikátorok működőképességét, kultúrközi eltéréseit, illetve használhatóságát, ahogy a teszt egészénél tesszük.

Tájékoztató irodalom:

Rózsa S., Nagybányai Nagy O., Oláh A. (szerk., 2006). A pszichológiai mérés alapjai. Elmélet, módszer és gyakorlati alkalmazás. Bölcsész Konzorcium. HEFP kézirat

Horváth György (1997). *A kérdőíves módszer*. Pedagógus Könyvek, Műszaki Könyvkiadó Budapest, 1-69.

Horváth György (1997). *A modern tesztmodellek alkalmazása*. Akadémiai Kiadó, Budapest, 1-69. (A Pszichológia a gyakorlatban sorozat 50. kötete)

Szokolszky Ágnes (2004). *A pszichológiai kutatás módszertana*. Osiris, Budapest. 3. és 5. fejezet, 119-190, 271-384.

Bartram, D. (szerk., 1999): International guidelines for the development of test-user performance standards. (A Nemzetközi Tesztbizottság (International Test Commission) kiadványa. Magyarul kiadta a Magyar Pszichológiai Társaság Tesztbizottsága (2001). Megjelent még: Nemzetközi irányelvek a tesztalkalmazás normáinak kialakításához. In Rózsa S., Nagybányai Nagy O., Oláh A. (szerk., 2006)

Az MPT Tesztbizottsága tesztminősítési sémája a pszichológiai vizsgálóeszközök (tesztek) leírására és értékelésére. (2011). <http://mpt.hu/atarsasagrol/szervezet/bizottsagok/tesztbizottsag/tesztminosites/>