

# Testilautakunnan ohjeistus: Psykologisten testien käyttökelpoisuuden arviointi

1.9.2024

## Johdanto

Psykologit ovat ainoa ammattikunta, jonka ammattitaitoon sisältyy laaja psykologian teorioiden, psykometriikan, psykologisen tutkimusmetodologian ja psykologisen tutkimusprosessin kokonaisuuden opiskelu, jotka yhdessä mahdollistavat yksittäisten psykologisten testien tulosten kriittisen ja pätevän tulkinnan ja käytön osana psykologin tutkimusprosessia. Kyky arvioida psykologisten testimenetelmien käyttökelpoisuutta ja soveltuvuutta aiottuun käyttötarkoitukseen on osa psykologin ammatillista pätevyyttä ja myös ammattieettistä vastuuta. [Pohjoismaiden psykologien ammattieettisissä periaatteissa](#) todetaan seuraavaa:

*”Psykologi on tietoinen niistä rajoituksista, jotka liittyvät psykologisiin menetelmiin ja niistä tehtäviin johtopäätöksiin. Psykologi on erityisen varovainen käyttäessään menetelmiä, apukeinoja ja tekniikoita, jotka eivät täytä tavanomaisia menetelmille asetettuja vaatimuksia tai joita hän ei täysin hallitse.”*

Jotta psykologisia testejä käyttävä tai testeistä saatua tietoa hyödyntävä psykologi voisi täyttää edellä mainitut vaatimukset, hänen tulee siis kyetä arvioimaan

- a. täyttääkö jokin psykologiseksi testiksi väitetty menetelmä aidosti psykologiselle testille asetetut kriteerit ja laatuvaatimukset sekä
- b. soveltuuko kyseinen testimenetelmä siihen tarkoitukseen, mihin sitä aiotaan käyttää (esim. onko testi tarkoitettu tietyn ikäisten tutkimiseen, onko testi kehitetty ensisijaisesti kliiniseen käyttöön vai esimerkiksi henkilöarviointiin)

Tämän Psykologiliiton testilautakunnan laatiman ohjeistuksen tavoitteena on tukea psykologeja näiden arvioiden tekemisessä.

## Psykologisen testin määritelmä

Psykologisen testin yksiselitteinen määrittelemineen on haastavaa. Euroopan psykologiliittojen federaatio (EFPA) käyttää omassa [ohjeistuksessaan](#) seuraavaa, [Standards for Educational and Psychological Testing](#) -julkaisusta löytyvää määritelmää:

*“the label **test** is used for any evaluative device or procedure in which a sample of examinee’s behaviour in a specified domain is obtained and subsequently evaluated and scored using a standardized process”*

Vapaasti käännettynä:

*“[psykologiseksi] testiksi kutsutaan mitä tahansa arvioivaa välinettä tai menettelytapaa, jonka avulla tutkittavan käyttäytymisestä tietyllä osa-alueella hankitaan otos, jota sitten arvioidaan ja pisteytetään käyttämällä standardoitua prosessia”*

Olennaisempaa kuin jonkin tietyn menetelmän tarkka luokittelu testiksi tai ei-testiksi on kuitenkin **kyetä arvioimaan psykologisessa tutkimuksessa käytettyjä tai käytettäviksi aiottuja menetelmiä psykologisille testeille asetetuilla laatustandardeilla - eli arvioida täyttyvätkö nämä standardit kunkin menetelmän kohdalla.**

Testi on tarkoitettu ainoastaan psykologien käyttöön, jos testin tekijä, alkuperäislaitoksen kustantaja tai muu testin oikeudenomistaja on näin ilmaissut, tai testin julkiselle levitykselle on asetettu rajoitus, jonka mukaan jälleenmyyjä saa myydä testiä ainoastaan psykologeille. Näiden rajoitusten tarkoituksena on mm. estää testien väärinkäyttö yleispätevinä menetelminä, testitulosten virheellinen yleistäminen ja tulkitseminen sekä testitulosten käyttö päätöksenteossa ilman kokonaisvaltaista tutkimusta.

### **Psykologisille testeille asetetut standardit**

Laajin ja keskeisin psykologisiin testeihin ja psykologisten tutkimusten suorittamiseen liittyvä ohjeisto on jo edelläkin mainittu yhdysvaltalainen [Standards for Educational and Psychological Testing](#), jonka ovat luoneet yhteistyössä American Psychological Association (APA), American Educational Research Association (AERA) ja National Council on Measurement in Education (NCME). Ohjeiston uusin versio on vuodelta 2014.

Vuoden 2014 ohjeisto jakautuu kolmeen osaan, joissa käsitellään

1. testien perusteisiin liittyviä standardeja (validiteetti, reliabiliteetti, oikeudenmukaisuus)
2. testien laatimiseen, käyttöön sekä testattavien/tutkittavien ja testien käyttäjien oikeuksiin ja velvollisuuksiin liittyviä standardeja
3. tarkempia testien käyttöön liittyviä standardeja eri sovellusalueilla (mm. psykologinen tutkimus)

Näihin standardeihin perehtyminen on suositeltavaa kaikille psykologisten testien käyttäjille. Kaikkein keskeisimmät kohdat kyseisistä ohjeistuksista on kuitenkin pyritty tiivistämään tähän dokumenttiin.

### **Psykologisten testien arvioinnissa käytettävät tiedonlähteet**

Euroopan psykologiliittojen federaatio EFPA on testiarviointiohjeistuksessaan [EFPA Review Model for the Description and Evaluation of Psychological and Educational Tests](#) (versio 4.2.6 vuodelta 2013) maininnut seuraavat tiedonlähteet, joita voidaan hyödyntää testimenetelmien käyttökelpoisuuden arvioinnissa. Ohjeistuksesta on tulossa uusi versio vuonna 2025.

1. **Testikustantajan laatima testin käyttäjälle tarkoitettu testikäsi kirja ja/tai testin pohjalta laaditut palauteraportit.** Nämä ovat ydinmateriaaleja psykologisen testin arvioinnin kannalta.

2. **Akateemisesta tai muusta kirjallisuudesta saatavilla oleva avoin tieto**, esim. onko testistä olemassa ja/tai viitataan testin käsikirjassa vertaisarvioituihin tutkimuksiin? Tutkimuskirjallisuuteen tutustuminen on erityisen tärkeää, mikäli testikäsikirja on vaatimaton tai puuttuu kokonaan.
3. Testin kustantajalla oleva tieto, jota ei ole virallisesti julkaistu tai jaettu.
4. Liikesalaisuuksien piiriin kuuluva tieto.

**Käytännössä useimmille psykologisten testien käyttäjille kaksi ensimmäistä kohtaa eli testikäsikirja, testin palauteraportit sekä testiin liittyvät tieteelliset julkaisut ovat kohtuullisimmin saavutettavia lähteitä.** Kaksi viimeistä lähdetyyppiä koskevat lähinnä virallista ja määrämuotoista testin arviointiraportin luomista ja ne on sisällytetty tähän täydellisyyden vuoksi. Virallisia testiarvioiteja tekevät muun muassa [British Psychological Society](#) sekä yhdysvaltalainen [Buros Center for Testing](#). Vaikka esimerkiksi näiden toimijoiden tuottamia maksullisia arviointiraportteja voidaan hyödyntää joidenkin testien englanninkielisten alkuperäisversioiden arvioinnin apuna, tulisi testien Suomessa julkaistuja versioita joka tapauksessa arvioida myös itsenäisinä tuotoksina. Suomessa ei ole toistaiseksi tahoa, joka laatisi vastaavanlaisia arviointeja Suomessa julkaistusta testeistä, joten myös siksi on tärkeää, että suomalainen psykologikunta tuntee arvioinnin perusteet.

## Testistä arvioitavat osa-alueet

1. **Onko testin käyttäjälle saatavilla oleva dokumentaatio riittävää? (esim. käyttäjä- ja tekniset käsikirjat, normitaulukot yms.)**

**Perusajatus:** *Dokumentaatioissa on esitetty loogisesti ja selkeästi mitä testi on suunniteltu mittaamaan ja miksi se on rakennettu niin kuin se on.*

- i. **Kehitystyö:** Onko testin käyttötarkoitus ja sen tarkoitettu kohderyhmä kuvattu selkeästi? Onko testidokumentaatioissa kuvattu testien teoreettinen viitekehys tai keskeiset käsitteet, testiosioiden alkuperä, ärsykemateriaalin kehittämisprosessi, pilotointi, vertailututkimukset ja kehittämisprosessin aikana tehdyt muutokset?
- ii. **Käännös / adaptaatiotyö:** Onko testidokumentaatioissa kuvattu mahdollinen suomalaisen version testikäännös ja/tai adaptaatioprosessi?
- iii. **Standardointi:** Onko testidokumentaatioissa kuvattu selkeästi ja yksityiskohtaisesti standardointiotoksen koko ja koostumus sekä standardointiprosessi?
- iv. **Normit:** Onko testidokumentaatioissa kuvattu selkeästi ja yksityiskohtaisesti normiryhmien koot, normitustapa, normien lähteet, normien edustavuus, arviointiolosuhteet yms.

**Nyrkkisääntönä on, että normiryhmän riittävä koko on klassisessa normituksessa vähintään 200 henkilöä ja vähintään 300 henkilöä kun testitulosten pohjalta tehdään ns. ei-triviaaleja päätöksiä (esim. henkilövalintaa). Yli 1000 henkilön normiryhmiä voidaan pitää koonsa puolesta erinomaisina (kts. EFPA:n arviointi**

kohta 9.1.3)

Erityisen tärkeää on arvioida, miten hyvin normiryhmä vastaa mahdollisesti testillä tutkittavia henkilöitä. Hieman kärjistettynä esimerkiksi pelkästään lapsille standardoitua testiä ei ole perusteltua käyttää aikuisten tutkimiseen.

- v. **Reliabiliteetti:** *Reliabiliteetilla viitataan testitulosten johdonmukaisuuteen testausprosessia toistettaessa - eli siihen että testi antaa suurin piirtein samanlaisen tuloksen eri testauskertojen välillä, mikäli myös testillä mitattavan kohteen voidaan olettaa pysyvän samanlaisena testikertojen välillä. Asian voin esittää myös tilastotieteellisesti siten, että reliabiliteetilla tarkoitetaan sitä kuinka vapaita testitulokset ovat mittausvirheen varianssista.*

Onko dokumentaatioissa kuvattu hyvin reliabiliteetti ja mittauksen keskivirhe (*standard error of measurement, SEM*) sekä kuvattu kattavasti erilaisia sisäisen yhtenäisyyden (*internal consistency*), ajallisen pysyvyyden ja/tai pisteyttäjien/arvioijien välisiä reliabiliteettimittareita ja niihin liittyviä keskivirheitä sekä selitetty niiden merkitys? Onko testitulosten yleistettävyyttä arvioitu?

**Nyrkkisääntönä on, että reliabiliteettikertoimien pitäisi olla vähintään suurempia kuin 0.6 tai 0.7 tehtäessä ei-triviaaleja päätöksiä.** Minimikoko vaihtelee kuitenkin käytetyn reliabiliteettimittarin mukaan. Suosittelemme tutustumaan [EFPA:n arviointimallin](#) kohtaan 10, josta löytyy tarkempia lisätietoja aiheesta.

- vi. **Validiteetti:** *Validiteetilla viitataan siihen, missä määrin tutkimustieto ja teoria tukevat testitulosten pohjalta tehtyjä tulkintoja testin käyttötarkoitusta varten - eli sitä, palveleeko testi käyttötarkoitustaan ja voidaanko sen pohjalta tehdä haluttuja johtopäätöksiä? Toisin sanoen: mittaako testi sitä, mitä sillä pyritään mittaamaan. Validiteettimittareita on useita erilaisia eikä kaikkia mahdollisia mittareita tarvitse olla mainittuna. Tärkeää on pystyä arvioimaan olemassa olevien tietojen pohjalta testin kokonaisvaliditeettia.*

**Rakennevaliditeetti** (mittaako testi niitä asioita, mitä se pyrkii mittaamaan vai jotain aivan muuta?): Onko testidokumentaatioissa kuvattu hyvin testin rakennevaliditeetti sekä viitattu useampaan selkeästi ja totuudenmukaisesti kuvattuun tutkimusjulkaisuun?

**Kriteerivaliditeetti** (ovatko testitulokset tutkitusti yhteydessä haluttuihin tosielämän muuttujiin, esimerkiksi työssä suoriutumiseen?): Onko testidokumentaatioissa kuvattu hyvin testin kriteerivaliditeetti ja viitattu viitattu useampaan selkeästi ja totuudenmukaisesti kuvattuun tutkimusjulkaisuun?

Tarkemmat ohjeet eri validiteettien arviointiin on kuvattu [EFPA:n](#)

[arviointimallin](#) kohdassa 11.

- vii. **Tietokoneella automaattisesti luodut raportit:** Onko testidokumentaatiossa kuvattu selkeästi ja yksityiskohtaisesti automaattisesti luotujen raporttien formaatti, kattavuus, reliabiliteetti ja validiteetti?

## 2. Testikäyttäjälle annettujen testin suorittamisohjeiden laatu

- a. **Testin esittämisohteet:** Onko testin esittämisohteet selitetty selkeästi ja yksityiskohtaisesti askel askeleelta? Annetaanko hyviä, yksityiskohtaisia neuvoja ongelmatilanteisiin ja vastauksia tutkittavilta tuleviin mahdollisiin kysymyksiin?
- b. **Testin pisteytysohjeet:** Onko testin pisteytysohjeet esitetty selkeästi ja yksityiskohtaisesti, onko pisteytysvirheiden riskit ennakoitu ohjeistuksessa? Jos pisteytys tehdään automaattisesti, onko esitetty näyttöä siitä, että pisteytys toimii oikein?
- c. **Testinormien käyttöohjeet:** Onko testin eri normitaulukoiden käyttämisen ohjeet esitetty selkeästi ja yksityiskohtaisesti? Onko normien käyttöön liittyvät virheet ennakoitu ohjeistuksessa? Jos normivertailu tehdään automaattisesti, onko esitetty näyttöä siitä, että pisteet muunnetaan oikein ja oikeaa normiryhmää käyttäen?
- d. **Tulkinta ja raportointiohteet:** Onko dokumentaatiossa mainittu, millaista ammattitaitoa ja osaamista testitulosten tulkinta vaatii? Onko eri pisteiden tulkinta ohjeistettu yksityiskohtaisesti, huomioiden normatiiviset mittarit sekä käsitellen eri skaalojen välisiä suhteita yksityiskohtaisten esimerkkien ja esimerkitapausten kautta? Onko ohjeissa käsitelty epä johdonmukaisen vastaamisen, vastaustyylien, vääristelyn, yms. huomioimista?
- e. **Palautteohjeet:** Onko testipalautteen antamisesta annettu yksityiskohtaiset ohjeet, mukaan lukien automaattisesti luotujen raporttien käyttö (mikäli saatavilla)
- f. **Ohjeet liittyen hyviin käytäntöihin, oikeudenmukaisuuteen ja vääristymiin:** Onko esitetty yksityiskohtaista tietoa sukupuoleen ja etnisyyteen liittyviä vääristymiä käsittelevistä tutkimuksista sekä niihin perustuvia varoituksia liittyen testin käyttöön ja validiteettien yleistettävyyteen?
- g. **Käyttöön liittyvät rajoitukset:** Onko kuvattu selkeästi ketä voidaan ja ketä ei pitäisi arvioida kyseistä testiä käyttämällä, mukaan lukien hyvät perustelut rajoitteille (esimerkiksi vamman tyyppi, kielitaidon taso)
- h. **Ohjelmisto ja tekninen tuki:** Tietokone- tai Internetpohjaisen testauksen tapauksessa: Onko tarvittava tieto selaimen liittyvistä vaatimuksista, tietokoneohjelmiston asentamisesta ja käyttämisestä riittävä (sisältäen mahdolliset virhetilanteet ja eri käyttöjärjestelmät)? Onko teknisen tuen saatavuus selkeästi kuvattu?
- i. **Viitteet ja tukimateriaalit:** Onko materiaaleissa kuvattu yksityiskohtaisia viittauksia relevanttiin tieteelliseen tutkimukseen sekä viittauksia muuhun relevanttiin testimateriaaliin?

## **Psykologisen testin soveltuvuus aiottuun käyttötarkoitukseen**

Kun psykologisen testimenetelmän on todettu täyttävän sille asetetut yleiset laatuvaatimukset, on vielä varmistettava, soveltuuko se aiottuun käyttötarkoitukseen - toisin sanoen, voidaanko testillä saada sellaista tietoa, jota voidaan hyödyntää haluttuihin tutkimuskysymyksiin vastaamiseksi. Psykologi on vastuussa myös siitä, että hän tuntee näytön testin validiteetista sen kulloiseenkin aiottuun käyttötarkoitukseen - tähän vaikuttavat sekä psykologisen tutkimuksen tutkimuskysymykset että tutkittavan henkilön/henkilöiden ominaisuudet.

### *Psykologisen tutkimuksen tutkimuskysymykset*

Psykologinen tutkimusprosessi alkaa määrittämällä mahdollisimman tarkkaan ne syyt, miksi tutkittavaa henkilöä (tai henkilöitä) arvioidaan - eli mitä psykologisella tutkimuksella halutaan selvittää. Tämän pohjalta tulee arvioida, mitä tietoa tulee kerätä ja mitä työkaluja käyttämällä kyseisen tiedon kerääminen onnistuu parhaiten. Tämä vaatii usein psykologilta sovellusala-kohtaista ammattitaitoa sekä yhteistyötä tutkimusta pyytävän tahon kanssa, jotta voidaan selvittää kyseisen tahon tiedon tarve. Testilautakunta muistuttaa myös, että psykologin tulee ensin varmistua tutkimusta pyytävän tahon oikeudesta saada pyytämänsä tieto ennen kuin psykologiseen tutkimukseen voidaan ryhtyä. Hyvin perustellusta syystä tutkimuksessa saatetaan päätyä käyttämään testiä, jota ei ole suoraan tarkoitettu kyseiseen tarkoitukseen tai joka on tarkoitettu erilaiselle tutkittavien ryhmälle. Tällöin testin käytössä ja testitulosten tulkinnessa on noudatettava erityistä varovaisuutta.

Psykologisessa tutkimuksessa käytettävien psykologisten testimenetelmien valinta voidaan tehdä yksilöllisesti jokaisen tutkittavan henkilön kohdalla tai samaa testipatteristoa voidaan käyttää useammalle tutkittavalle tutkimuksen luonteesta riippuen (vrt. yksilön neuropsykologinen tutkimus vs. rekrytointitilanteet, joissa hakijoita tulisi kohdella mahdollisimman tasavertaisesti). Tutkimuksen edetessä voi olla tarkoituksenmukaista muokata alkuperäistä suunnitelmaa esimerkiksi käyttämällä lisätestejä, mikäli se on alustavien tutkimustulosten pohjalta perusteltavissa ja tutkimuksen kontekstin kannalta sopivaa.

Mikäli testeistä on olemassa eri versioita, on usein perusteltua valita testin uusin versio, koska sen eteen on todennäköisesti tehty testin ominaisuuksia parantavaa jatkokehitystyötä. Joissain tapauksissa voi kuitenkin olla syytä käyttää myös testin vanhempaa versiota, esimerkiksi tehtäessä pitkäaikaista tutkimusta tai hyödynnettäessä aiemman version sisältämiä tarkoituksenmukaisia osatestejä, joita uudemmassa versiossa ei ole. Pääasia on, että testivalinnat ovat hyvin perusteltavissa. Testien valinnassa voi myös hyödyntää soveltuvin osin Psykologiliiton ohjeistusta *Henkilöarviointi työelämässä – ohjeistus hyväksi käytännöiksi*, jonka osiossa 6. käsitellään testimenetelmiä ja niiden rajoitteita henkilöarvioinnin näkökulmasta. Kyseistä ohjeistusta voidaan hyödyntää suuntaviivoina myös muissa testien käyttökonteksteissa.

### *Tutkittavan henkilön ominaisuudet*

Tutkimuskysymyksen lisäksi on tärkeä tunnistaa tutkittavan henkilön ominaisuudet suhteessa tutkimuksessa tehtäviin testivalintoihin. Tällaisia ominaisuuksia ovat esimerkiksi

ikä ja kulttuuritausta. Samoin testivalintoihin vaikuttavat tutkittavan henkilön mahdolliset muut tarpeet ja kyvyt kuten lukutaito, näkö, kuulo ja motoriikka. Tällaiset tekijät määrittävät, missä määrin tutkittavan kanssa on tarkoituksenmukaista käyttää tiettyjä testejä.

Kunkin testin validiteetin ja reliabiliteetin lisäksi on tärkeää tuntee alkuperäisen standardointiotoksen demografiset piirteet. Psykologin tulisi valita testi, jonka normiryhmät ovat demografisesti (esimerkiksi ikä, sukupuoli, kulttuuri) ja kliinisesti tarkoituksenmukaisia tutkittavan henkilön kohdalla. Toisin sanoen tulee pyrkiä siihen, että tutkittavan testituloksia vertaillaan sellaisiin henkilöihin, joihin tulosten vertailu on tutkittavan kannalta järkevää. Yhdelle ryhmälle luotu testi ei välttämättä sovellu toisen ryhmän tutkimiseen. Esimerkiksi kansainvälisissä tai monikulttuurisissa sovellustapauksissa on tärkeää varmistaa, että testeillä arvioitavat käsitteet ovat yhteneväiset eri kulttuurisissa konteksteissa. Tästä voit lukea lisää [testilautakunnan ohjeistuksesta monikulttuurisuuden huomioimiseksi psykologisessa arvioinnissa](#).

### **Yhteenveto:**

Ennen testin käyttöä osana psykologista tutkimusta psykologin tulee arvioida testin käyttökelpoisuutta eri testimateriaaleista (esim. testikäsikirja) saatavan tiedon pohjalta. Erityisen tärkeää on, että näissä materiaaleissa käsitellään riittävän yksityiskohtaisella tasolla testin käyttötarkoitus, testin esittämis-/käyttöohjeet, ne ryhmät, joita testillä on tarkoitus tutkia, sekä sellaiset pisteytysten pohjalta tehtävät tulkinnat, joista on olemassa riittävää validiteetti- ja reliabiliteettitietoa. Riittävän käyttökelpoisuuden arvion tekemiseen vaaditun reliabiliteetti- ja validiteettidatan määrä ja laatu riippuu siitä, mikä testin rooli ja painoarvo on osana psykologin tutkimusprosessia sekä tutkimusprosessin mahdollisista seurauksista tutkittavalle.