



Leila Niemi-Murola
LT, dosentti, erikoislääkäri
HYKS, ATeK, Meilahden sairaala
leila.niemi-murola[a]hus.fi

Eurooppalainen erikoislääkäritentti ja monivalintakysymysten lumo

Arviointi ohjaa oppimista, sillä jokainen haluaa läpäistä tentin ensimmäisellä yrityksellä.

Suomalaisilla erikoistuvilla lääkäreillä on nykyisin mahdollisuus osallistua eurooppalaisen erikoislääkäritentin EDAIC:n ensimmäiseen osaan (European Diploma of Anaesthesiology and Intensive Care) osaamista mittaavana välitenttinä (in-training assessment). Helsingissä ja Turussa erikoistuvat lääkärit osallistuvat tähän kuulusteluun työnantajan kustannuksella joka vuosi yliopistosairaalapalvelunsa aikana, mutta muilla osallistuminen on kiinni omasta aktiviteetista. Valmistelut sujuvat tiukan protokollan mukaan, jokaisella osallistujalla on numeroitu paikka ja sijoittelu pitää dokumentoida valokuvan avulla (kuva 1).

Suomalaiseen tenttivalvontaan tottuneesta tämä järjestely tuntuu hieman oudolta, mutta EDAIC-erikoislääkärikuulustelu on juridisesti rinnastettavissa ylioppilaskirjoituksiin ja lääketieteellisten tiedekuntien sisäänpääsykokeeseen. Tavallisessa tentissä epäonnistumisesta seurauksena on sen uusiminen, mikä on harmillista sinänsä, mutta osallistujan tulevaisuuden kannalta sillä ei ole merkitystä. Korkeiden panosten (high-stakes)

kuulustelussa epäonnistuminen on sen sijaan todellinen vastoinkäyminen, joka estää uralla etenemisen ja jolla myös saattaa olla taloudellisia vaikutuksia (1). Tällaisen tentin täytyy olla pätevästi laadittu ja arvostelun niin perusteltua, että niitä pystytään tarvittaessa puolustamaan oikeusistuimen edessä.



Kuva 1. Meilahden sairaalan luentosali 4 odottaa tenttiin (in-training assessment) tulijoita.

Essee jättää tilaa omalle pohdinnalle ja myös tiedollisten aukkojen peittelylle.

Essee vai monivalintakysymys?

Suomessa tavallisin tenttimuoto on essee-kysymyksiin perustuva. Pikainen vilkaisu Finnes-tissa julkaistuihin erikoislääkärikuulustelun kysymyksiin osoittaa, että kysymysten muotoilu on hyvin perinteinen, laadittu kuten ylioppilasaineen otsikko. Tällainen muotoilu ei ohjaa vastaajaa, joka usein kirjoittaa paperille kaiken tietämänsä ilman erityistä struktuuria. Arvostelijalla on edes-sään kova urakka vastauksen oleellisten asioiden paikallistamisessa. Nykyisin kirjallinen viestintä on sähköistä, joten koukeroisen käsialan tulkitseminen on haaste sinänsä.

Vastaajan kannalta essee on hyvä tenttimuoto, sillä se jättää tilaa omalle pohdinnalle ja myös mahdollisten tiedollisten aukkojen peittelylle. Korkeiden panosten tenteissä essee-vastausten arvosteleminen kuitenkin on ongelma, sillä vapaan kerronnan arviointikriteerejä on vaikea saada täsmällisiksi (1, 2, 3). Hyvästä tekniikasta on myös hyötyä. Loogisesti muotoillusta, selvästi jäsenne-llystä ja sujuvasti etenevästä vastauksesta on helppompaa antaa hyviä pisteitä kuin sekavasta, vaikka asiasisältö molemmissa olisi aivan sama. Monivalintatentti saattaa vastaajasta tuntua armottomalta, sillä kysymykset ovat tarkkoja eikä vaihtoehtoja ole mahdollista selventää omin sanoin. Tiedon mittaajina monivalintakysymykset ovat kuitenkin huomattavasti parempia kuin esseet, joissa on paljon kerronnan tuomia sekoittavia tekijöitä (1, 2, 3). Korjaajan kannalta monivalinta on helppo ja nopea varsinkin, jos paperit lähetetään optiseen luentaan ja tulokset tulevat suoraan taulukkumuodossa.

Arvioinnin pätevyys riippuu kysymyksen muotoilusta

Arviointiin pätevät samat säännöt kuin tutkimukseen. Sitä saat mitä tilaat eli jos tutkimuskysymys on huonosti laadittu, ei vastauksiltakaan voi

paljon odottaa. Klassinen esimerkki arvioinnin reliabiliteetista eli luotettavuudesta tai toimintavarmuudesta ja validiteetista eli pätevydestä on ajan ilmoittaminen kelloa katsomalla. Luentosalin kello näyttää kaikille samaa aikaa, kun kuulijat vilkuilevat sitä odottaessaan luennon päättymistä. Jos useat mittaajat saavat saman tuloksen, mittarin reliabiliteetti on hyvä. Jos kello on pysähtynyt, se joko edistää tai jättää, tulos ei ole oikea eli mittarin validiteetti on huono.

Hyvän monivalintakysymyksen laatiminen ei ole helppoa. Dikotominen väittäminen (oikein-väärin) tuntuu ajatuksen tasolla helpolta laatia, mutta lääketieteessä asiat ovat harvoin mustavalkoisia (4). Kysymysten laatijan vääräksi tarkoitama väittäminen on helppo tunnistaa siitä, että siinä on mukana määrite aina, usein, joskus jne. Näistä lisämääreistä huolimatta vaihtoehtoisissa on usein epäselvyyttä, mikä turhauttaa hermostunutta vastaajaa. Vaikeaa on myös laatia kysymyksiä, joissa kaikki vaihtoehdot ovat yhtä painavia. Erään kokeneen lääketieteellisen pedagogiikan asiantuntijan analyysin mukaan useita vaihtoehtoja sisältävissä tehtävissä yksi vaihtoehto on usein päivänselvästi väärä, toinen on hyvin helppo ja valittavaksi jää vain kaksi vaihtoehtoa. Vastaustekniikan osuutta on vaikea minimoida myös monivalintatehtävissä, mutta avainasiakysymykset, paras mahdollinen vastaus- ja yhdistelmätehtävät ovat osoittautuneet toimiviksi (Taulukko 1).

Psykometriikka ja vastausten tilastollinen arviointi

Osallistujien kannalta oikeudenmukaisuus on keskeinen arvioinnin ominaisuus (1, 2, 6). Toisin kuin essee-vastauksia, monivalintatehtäviä pystytään arvioimaan tilastollisesti. Psykometriikka on psykologisissa mittauksissa käytetty teoria ja tekniikka, joka on hyvin käyttökelpoinen laskettaessa

>>

Taulukko 1. Kirjallisten tenttien kysymystyyppit (Epstein 2007).

Kysymystyyppi	Mitä mittaa	Missä käytetään	Rajoitukset	Vahvuudet
Monivalintakysymykset sisältäen paras mahdollinen vastaus ja yhdistelytehtävät	Tieto, ongelmanratkaisukyky	Jakson lopuksi pidettävät päättökokeet ja kansalliset loppututkinnot, joista annetaan todistus	Vaikeita laatia siten, etteivät ne johdattele. Asiayhteydestä irrotettuna saattavat vaikuttaa keinotekoisilta	Voidaan arvioida monia osaa- misalueita lyhyessä ajassa, hyvä reliabiliteetti, arvosanat voidaan laskea tietokoneella
Avainasiakysymykset	Kliininen päättely, ongelmanratkaisu, tiedon soveltaminen	Kansalliset loppututkinnot, joista annetaan todistus	Ei näyttöä siirrettävyydestä päättelyä vaativiin oikean elämän tilanteisiin	Arvioi kliinistä ongelman- ratkaisua, ei johdattelua, arvosanat voidaan laskea tietokoneella
Lyhytessee	Diagnostisten tutkimusten tulkinta, ongelmanratkaisu, kliininen päätöksenteko	Välitentit ja jakson lopuksi pidettävät päättökokeet	Reliabiliteetti riippuu arvioijien koulutuksesta	Ei johdattelua, voidaan arvioida tulkintaa, arvosanat voidaan laskea tietokoneella
Strukturoitu essee	Tiedon synteesi, kirjallisuuden tulkinta	Prekliinisellä kurssilla, sovelletusti myös käytännön harjoittelujen aikana	Korjaus aikaa vievää, arvioijien kriteerien yhtenäistämisen vaatii paljon työtä	Ei johdattelua, kirjoittaja joutuu pohtimaan vastaustaan

tenttitulosten reliabiliteettia ja validiteettia. Psykometriikkaa käytetään myös kyselytutkimuksia analysoitaessa.

Reliabiliteettia voidaan tarkastella tentin stabiliteettina, käytetyn mittarin tasalaatuisuutena ja sisäisenä konsistenssina eli yhtenäisyytenä (4). Edellä esitetty kellovertaus kuvaa hyvin tentin stabiliteettia, mittari ja sen toimintaperiaate pysyy samanlaisena mittaushetkestä riippumatta. Tenttipankin kysymyksistä kootun monivalintatentin tulee olla tasapainoinen ja –laatuinen riippumatta siitä, mitkä kysymykset ovat sattuneet mukaan. Perinteiset pretest – posttest-tutkimukset edellyttävät tasalaatuisen mittarin käyttöä (5). Reliabiliteetti voidaan myös tarkastaa laskemalla arvioijien näkemysten yhteneväisyys (inter-rater reliability), kellovertaus sopii hyvin myös tähän. Kolmas tapa reliabiliteetin osoittamiseksi on sisäinen konsistenssi. Jos mittari erottelee kaikkien osaamat eli helpot kohdat ja vain todellisten huippujen tietämät asiat, sen sisäinen konsistenssi on hyvä. Tämän voi laskea käyttämällä Cronbachin alfaa ja faktorianalyysia (6). Tentin hyvä reliabiliteetti siis tarkoittaa, että mittari on kontrolloitavissa, ennakoitavissa, pysyvä ja tulokset toistettavissa (6).

Validiteetista puhuttaessa voidaan tarkoittaa monta asiaa (6). Sisällön validiteetti on hyvä, jos

mittari mittaa sitä asiaa, mitä oli tarkoituskin. Rakenteen validiteetti on hyvin lähellä reliabiliteettia. Ulkoisella validiteetilla tarkoitetaan sitä, miten hyvin mittarin tulokset ovat yleistettävissä toiseen populaatioon eli ovatko esimerkiksi eri yliopistoissa erikoistumispalveluaan suorittavien henkilöiden tulokset vertailukelpoisia (6). Gaussin käyrää on perinteisesti käytetty tapa arvottaa osallistujia. Se on normiin pohjautuva arviointitapa, jossa tenttijää arvioidaan suhteessa muihin samaan ryhmään kuuluviin (5, 7, 8). Jos erikoistuva osallistuu samaan tenttiin huippuosaajien kanssa, hänen todellisuudessa keskinkertainen suorituksensa vaikuttaa heikolta muiden rinnalla. Jos ryhmän tulokset ovat hyviä, korjaaja tulkitsee kysymysten olleen liian helppoja. Jos sama henkilö olisi osallistunut tenttiin osaamistasoltaan hyvin heikkojen kollegojen kanssa, hänen keskinkertainen suorituksensa olisi noussut esiin hyvänä. Jos koko ryhmän suoritukset ovat olleet huonoja, korjaajalle tulee kiusaus tulkita kysymysten tällä kertaa olleen liian vaikeita. Normipohjainen arviointi soveltuu kuitenkin hyvin tilanteisiin, joissa voidaan päästää läpi vain ennalta sovittu määrä osallistujia, esimerkkinä pääsykokeet (8).

Kansainvälisissä korkeiden panosten tentteissä (esim. EDAIC I ja US Medical Licensure

Examination eli USMLE osat I ja II) käytetään kriteeripohjaista arviointia, jossa arviointi pohjautuu ennalta määriteltyihin kriteereihin (8). Ryhmän tasolla ei siis ole merkitystä yksittäisen suorituksen arvioinnin kannalta. Angoffin menetelmä on suosittu ja suhteellisen yksinkertainen (9, 10). Siinä asiantuntijapaneeli määrittelee, miten hyvin läpäisyn rajamailla oleva tenttijä selvittäisi kysymyksen eli kuinka monta prosenttia näistä tapauksista vastaisi siihen oikein (Taulukko 2). Ebelin menetelmää käytettäessä asiantuntijapaneeli arvioi kunkin kysymyksen vaikeusasteen (helppo, keskitasoa, vaikea) ja relevanssin asteikolla oleellisen tärkeä, tärkeä, hyväksyttävä, kyseenalainen ja sen jälkeen arvioidaan, kuinka moni rajatapauksista osaisi vastata kysymykseen oikein (9).

Tentti palautteena

Eurooppalainen välitentti EDAIC I (in-training assessment) antaa osallistujalle palautteen omasta edistymisestä ja mahdollisista tiedollisista katvealueista. Koska useimmat opiskelevat päästäkseen tentistä läpi, summatiivinenkin eli päättöarviointi antaa palautetta omasta osaamisesta (1, 3, 11). Tenttimisen on myös todettu vahvistavan opittujen asioiden mieleen painamista (11).

Harvalle tulee mieleen, miten arvokasta palautetta koulutuksen vastuuhenkilöt saavat

puolueettoman arvioijan laatimasta tentistä (2). Tenttitulokset näyttävät, miten hyvin annettu koulutus on linjassa eurooppalaisten osaamisvaatimusten kanssa, miten laadukasta opetus on ollut ja että koulutuksen aikana on käyty läpi ne asiat, jotka on ollut tarkoitus oppia (2). EDAIC:n etu on sen yleiseurooppalaisuus eli tentti on kaikille sama, mikä mahdollistaa tulosten vertailun eri maiden ja koulutusyksiköiden välillä.

Kansainväliset tenttitulokset lasketaan niin suurista populaatioista, että niiden avulla on mahdollista saada kiinnostavaa tietoa omasta opetuksesta muihin verrattuna. Yleensä suomalaisten kliinisen vaiheen osaaminen on hyvää, sillä meillä jo lääketieteen perusopiskelijat pääsevät osallistumaan kliiniseen työhön. Muualla Euroopassa erikoistuvat aloittavat seuraamalla seniorien työskentelyä ja vasta opintojen edetessä pääsevät itse osallistumaan potilaan hoitoon. Perusopetuksen painottuminen teoriaopintoihin näkyy myös tenttituloksissa.

Suomi on niin pieni maa, että täällä työskentelevät muodostavat eräänlaisen sisäpiirin. Kaikki tuntevat kaikki yhteisten tuttuja kautta, opiskelupaikkakunnalla tai erikoistumissairaalalla ei ole merkitystä. Muualla maailmassa asia ei ole näin. Yliopistojen ranking-listat uutisoidaan säännöllisesti YLE:n verkkosivuilla ja suurimmissa sanomalehdissä. Kansainvälisten julkaisusarjojen

>>

Taulukko 2. Läpäisyrajan arviointi Angoffin menetelmän avulla. Kukin arvioija esittää näkemyksensä siitä, kuinka moni rajatapauksista vastaa kysymykseen oikein. Jos tentin maksimipistemäärä olisi 30 pistettä, läpäisyraja olisi arvioijien paneelin mukaan 15,81 pistettä (9, 10).

Kysymys	Arvioija I	Arvioija II	Arvioija III	Arvioija IV	Yhteensä
1.	90 %	85 %	95 %	80 %	87,5 %
2.	60 %	50 %	55 %	50 %	53,8 %
3.	20 %	15 %	15 %	10 %	15 %
4.	40 %	35 %	35 %	30 %	35 %
5.	70 %	65 %	70 %	60 %	66,3 %
6.	60 %	65 %	55 %	55 %	58,8 %
Tentin läpäisyraja	pisteet x 0,527				

Hyväkään toistettavuus ja oikeellisuus eivät riitä, jos tentti ei tue oppimista.

artikkelien kirjoittajien nimien yhteydessä näkyy usein listauksia siitä, mistä maineikkaasta sairaalasta asianomainen on saanut diplominsa tai missä hän on palvellut stipendiaattina. Meillä suomalaisilla on hyvä maine ja tiedämme olevamme hyviä, mutta tulevaisuudessa saatamme joutua myös osoittamaan maineemme olevan oikeutettu.

Katse tulevaisuuteen

Epstein (1) määrittelee arvioinnin tavoitteeksi opiskelijoiden pystyvyyden tukemisen siten, että se motivoi ja antaa oppimistavoitteita, suojaa terveyspalvelujärjestelmää epäpäteviltä ammattiharjoittajilta ja antaa pohjan jatkokoulutukseen pääsyyllä. Cees van der Vleutenin mukaan tentin käyttökelpoisuuteen vaikuttavat sen reliabiliteetti, validiteetti, vaikuttavuus, hyväksyttävyyys ja kustannukset (3, 12). Hyväkään toistettavuus ja oikeellisuus eivät riitä, jos tentti ei tue oppimista, osallistujat tai järjestäjät eivät sitä hyväksy, pidä luotettavana tai sen järjestäminen ja osallistuminen tulevat liian kalliiksi (2). Mielenkiintoinen tulevaisuuden optio on sähköinen tentti, jonka osallistujat tekevät omilla laitteillaan. Tämä tietysti edellyttää verkotonta tilaa, jotta lisätiedon haku tentin aikana ei onnistu. Ylioppilaskirjoituksissahan tätä jo kokeillaan.

Tenttiin valmistautuminen, sen järjestäminen ja arvostelu kuluttavat paljon voimavaroja. On kaikkien edun mukaista, että saadut tulokset ovat päteviä ja läpinäkyviä. Etiikan lisäksi tenttimuodon pohdintaan vaikuttavat myös muut asiat. Riippumatta työvoiman vapaaseen liikkuvuuteen mahdollisesti tulevista rajoituksista, erikoislääkärikoulutuksen

laatua koskevat odotukset tuskin laskevat. Rahoitus on tehokkain Sosiaali- ja terveysministeriön käytössä olevista ohjaukeinoista ja sille odotetaan selvästi näkyvää vastinetta. Me kaikki teemme hyvää työtä, sekä erikoistuvat että heidän kouluttajansa, ja onneksi voimme sen myös osoittaa pätevän arvioinnin avulla. ■

Viitteet:

1. Epstein R. Assessment in medical education. *NJEM* 2007; 356: 387 - 396.
2. Norcini J, Anderson B, Bollela, V, ym. Criteria for good assessment: Consensus statement and recommendations from the Ottawa 2010 Conference. *MedicalTeacher* 2011; 33: 206–214.
3. Schuwirth LWT, van der Vleuten CPM. How to design a useful test: the principles of assessment. Kirjassa *Understanding medical education – Evidence, theory and practice*. 1. painos s. 195-207. Toim. Tim Swanwick, Wiley-Blackwell 2011.
4. Pamphlett R. It takes only 100 true-false items to test medical students: true or false? *Med Teach* 2005; 27: 468-72.
5. Cohen L, Manion L, Morrison K. *Research methods in education*. 7. painos sivut 279-215. Routledge, Lontoo, 2011.
6. Cook DA, Lineberry M. Consequences validity evidence: Evaluating the impact of educational assessments. *Acad Med* 2016; 91: 785 – 95.
7. Jamieson S. Likert scales: how to (ab)use them. *Med Educ* 2004; 38: 1212–18.
8. Bandaranayake RC. Setting and maintaining standards in multiple choice examinations: AMEE Guide No. 37. *Med Teach* 2008; 30: 836 – 845.
9. Homer M, Darling JC. Setting standards in knowledge assessments: Comparing Ebel and Cohen via Rasch. *Med Teach* 2016; 38: 1267-77.
10. Mäkinen M, Aune S, Niemi-Murola L, Herlitz J, Varpula T, Nurmi J, Å. B. Axelsson, A-B. Thorén, Castren M. Assessment of CPR-D skills of nurses in Gothenburg Sweden and Espoo Finland: Leadership makes a difference. *Resuscitation* 2007; 72: 264-269.
11. Larsen DP ym. Repeated testing improves long-term retention relative to repeated study: a randomised controlled trial. *Med Educ* 2009; 43: 1174-81.
12. Van Der Vleuten CPM. The assessment of professional competence: developments, research and practical implications. *Adv Health Sci Educ* 1996; 1: 41-67.