

# Luotettava mittari opiskelijapalautteelle

■ Leo Aarnio

**Visa Tuominen, Juha Rautapuro ja Antero Puhakka ottivat edellisessä numerossa (*Tieteessä tapahtuu* 3/2013, 15–23) sanojensa mukaan jossain määrin kriittisen tai ei-kriittisen kannan opiskelijapalautteen käyttämisestä osana korkeakoulujen rahoitusmallia. Kirjoitus on epämääräisyydessäänkin hyvä ja perusteltu näkökulma monimutkaiseen aiheeseen.**

Seuraavassa en argumentoi heidän varovaista johtopäätöstään vastaan vaan keskityn siihen, mitä kirjoituksesta tulkitsemani leikittelevän kriittinen sävy jättää huomiotta. Yleisenä kannanottona opiskelijapalautteen käyttämisen toivottavuudesta osana korkeakoulujen rahoitusmallia heidän näkökulmansa on turhan yksipuolinen ja valikoiva.

## Opiskelijapalautte yhtenä koulutuksen laadun mittarina

Kuten kirjoittajat jokseenkin itsestään selvästi toteavat (s. 23), ei opiskelijapalautte ole aukoton tapa saada tietoa opetuksen ja koulutuksen laadusta. Kirjoittajat myös ymmärtävät, ettei tästä tosiasiasta seuraa, ettei palautetta tulisi käyttää koulutuksen laadun mittarina osana korkeakoulujen uutta rahoitusmallia. Pätevässä argumentissa opiskelijapalautte kyselymittarin käyttämisen toivottavuudesta koulutuksen laadun indikaattorina tulisikin ottaa huomioon vaihtoehtoiset tavat mitata koulutuksen laatua.

Oletetaan, että koulutuksen laadun käyttämistä osana korkeakoulujen rahoitusmallia pidetään suotavana. Tällöin muutamien varauksin, joihin palaan kirjoituksen lopussa, olisi kenties perusteltua valita koulutuksen laadun indikaattoriksi yleisiltä mittausteoreettisilta ominaisuuksiltaan paras mittari.

Näihin ominaisuuksiin voidaan Niiniluotoa (1984, 183–191) karkeasti mukailen lukea kuuluviksi esimerkiksi reliabiliteetti, validiteetti sekä mittariin liittyvä käytännön hyöty (tai haitta).

Yllä mainittuja yleisiä mittausteoreettisia ominaisuuksia kirjoittajat eivät käsittele systemaattisesti, vertaile tai erittele arvioidessaan opiskelijapalautteen käyttöä koulutuksen arvioinnissa. Seuraavassa keskitynkkin ennen kaikkea niihin reliabiliteetin, tämän alle kuuluvan stabiliteetin, erottelukyvyn ja hyödyllisyyden käsitteisiin, joita he tulevat kirjoituksessaan sivunneeksi (tarkemmin ja kattavammin opiskelijapalautteen todennetuista mittausteoreettisista ominaisuuksista, ks. Richardson 2005).

## Reliabiliteetti, stabiliteetti ja erottelukyky

Tarkastellessaan eroja Joensuun yliopiston koulutusalojen välillä, kirjoittajat päätyvät siihen tulokseen, että osana rahoitusmallia koulutusalojen välille syntyisi kyseisellä mittarilla mitä ilmeisimmin tasapeli ja näin voisi käydä myös YOPALAN kohdalla tämän pyrkiessä erottelemaan yliopisto- ja näiden tarjoaman koulutuksen laadun mukaan. Mutta onko tasapeli välttämättä huono asia?

Tasapeli voidaan nähdä huonona asiana, jos se on seurausta mittarin huonosta erottelukyvystä. Toisin sanoen silloin, jos yliopistojen välillä todella on merkittäviä eroja, mutta ne eivät näyttäyty merkittävinä käytetyn kyselymittarin takia. Näin voi käydä esimerkiksi siksi, 1) että tutkinnon sisällä paljon vaihtelevaa koulutuksen laatua on kokonaisuudessaan vaikea arvioida, minkä seurauksena yksittäisetkin sitä koskevat arviot hakeutuvat asteikon keskelle, tai siksi, 2) että kyselyssä käytetty asteikko tasapäistää vastauksia.

Kohta 1) on ongelmallinen vain, jos vastaajan koulutuksen kokonaisuutta kuvaava keskimääräinen arvio ei kuvaa todellisuutta. Suurelta osin vastausten tasapäisyys kuitenkin selittyy sillä tosiasialla, että keskimääräiset kokemukset todella ovat melko samankaltaisia. Mittarin erottelukyky korkeakoulujen välillä riippuu tällöin otoksen koosta sekä korkeakoulujen välisten erojen suuruudesta ja hajonnasta. Jos tämänkaltaista tilastolliseen päättelyyn perustuvaa merkittävää eroa ei ole löydetävissä, käytännön erottelun mielekkyyttäkin tulisi tietenkin harkita.

Kohta 2) on ongelmallisempi siksi, että keskimäärin positiiviset vastaukset ovat tällöin asteikon, eivät mitattavan asian eli koulutuksen laadun funktioita. Väittäisin, että tässä tapauksessa ongelma koskee kuitenkin ennemmin arvioitua Joensuun-kyselyä neliportaisine asteikkoineen kuin opiskelijakyselymittareita yleensä.

Vastaajia saataneen Joensuun-kyselyä paremmin eroteltua lisäämällä vastausvaihtoehtoja sekä tarkentamalla väittämien muotoiluja. Tutkimukset indikoivatkin, että erottelukyvyltään, reliabiliteetiltaan, validiteetiltaan sekä käyttäjävälisyydeltään noin seitsenportainen vastausasteikko on optimaalinen ja merkittävästi neliportaista parempi (Preston & Colman 2000; Ramsay 1973).

Vaikka vastaajien taipumus suosia heikosti positiivisia vastausvaihtoehtoja onkin melko yleisesti todennettu, tämäkään ei aiheuttane harhaa rahoitusmallin suhteelliseen mittariin. ”Jokseenkin samaa mieltä” vastausta ei tulisiakaan tulkita osoituksena absoluuttisesta tyytyväisyydestä: niin sisällöllisessä tulkinnassa kuin rahoitusmallin osanakin tulee vastauksia arvioida suhteessa muihin annettuihin vastauksiin.

Vaikka kirjoittajien huomiot tutkintoa koskevan kyselymittarin erottelukyvystä ovat hyviä, unohtavat he, miksi koulutusalojen eroja alun perin tarkasteltiin. Jos kyselyn väittämien on tarkoitus mitata yliopistoon eikä ainoastaan koulutusohjelmaan palautuvaa opiskelijakokemusta, voivat pienet erot koulutusalojen välillä laadun mittarissa olla huonon erottelukyvyn lisäksi indikaattoreita myös hyvästä reliabiliteetista.

Kirjoittajat viittaavat toisaalla kirjoituksessaan opiskelijapalautekyselymittarin reliabiliteettiin epäilevään sävyyn vaikka jättävätkin yllä mainitussa tilanteessa asian huomiotta. Luotettavan mittarin tulisi antaa samankaltaisia tuloksia riippumatta mittaustilanteesta tai muiden arvioitavan asian kannalta epäoleellisten taustamuuttujien vaihtelusta. Tällaisten epäoleellisten kolmansien tekijöiden vaikutusten arvioiminen onkin peruseriaate kyselyjä laadittaessa. Kirjoittajat sivuuttavatkin muutamissa asiaa koskevissa negatiivisissa huomioissaan koko liudan tutkimuksia kyselyistä, joiden reliabiliteetti on hyväksi havaittu (ks. edellä mainittu Richardson 2005).

On kuitenkin totta, ettei kaikkia koulutusinstituution toimista riippumattomia kyselyarvioihin vaikuttavia epäoleellisia kolmansia tekijöitä saada otetuksi huomioon tai kontrolloitua kyselytutkimuksen keinoin. Näin ollen on hyvä olla epäileväinen ja osoittaa kritiikkiä rakennettuja kyselymittareita kohtaan, koska se auttaa korjaamaan mittareita luotettavammiksi ja ottamaan huomioon niihin välttämättä jääviä puutteita.

Ei tule kuitenkaan vaatia mahdottomia. Mittarin arvioitu yleiseen mittausteoriaan perustuva laatu tulee suhteuttaa muiden vastaavien mittareiden vastaaviin ominaisuuksiin. Täydellistä koulutuksen tai opetuksen laadun mittaria ei ole olemassakaan. Esimerkiksi suoritettujen tutkintojen määrä on koulutuksen laadun mittarina tarkka eli reliaabeli, muttei validi: sillä ei ole mitään tekemistä laadun kanssa. Opettajien vertaisarviointi on puolestaan havaittu epäluotettavaksi mittariksi opetuksen laadusta. Ei olekaan syytä olettaa opiskelijapalautteen olevan esimerkiksi edellä mainittuja mittareita huonompi, pikemminkin päinvastoin. Hyvät kyselymittarit pystyvät kattavasti ottamaan huomioon koulutuksen sisällöllisen laadun. Ne myös antavat todistetusti samansuuntaisia tuloksia esimerkiksi yleisesti melko luotettavana pidettyjen ulkopuolisten ammattiarvioijien arvioiden kanssa (Richardson 2005, 389–390).

Arvioidessaan Joensuun kyselymittarin erottelukykyä vastausten liukuvan keskiarvon avulla vuosina 2003–08, kirjoittajat tulevat arvioineek-

si, mutteivät eritelleeksi, myös mittarin stabiliteettia. *Ceteris paribus* on nimittäin toivottavaa, etteivät mittarin antamat arvot merkittävästi muutu näinkin lyhyellä ajanjaksolla siirryttäessä mittaustilanteesta toiseen. Siinä missä tuloksen olisi voinut arvioida positiiviseksi mittarin luotettavuuden näkökulmasta, päätyvät kirjoittajat tulkitsemaan vastausten säilyvyyden stabiliteetin sen sijaan yksinomaan sekä huonoksi erottelukyvyyksi että osoitukseksi mittarin käytännöllisestä hyödyttömyydestä opetuksen kehittämisessä.

Palautteen hyödyttömyyttä pohtiessaan kirjoittajat jopa ikään kuin ohimennen viittaavat tutkimukseen, jossa epäonnisesti oli päädytty tekemään opetuksen laatua huonontavia muutoksia kyselymittaria välineenä käyttäen. Toisin kuin he antavat ymmärtää, ei liene yllättävää, että on myös positiivisia esimerkkejä siitä, kuinka kyselymittareita on käytetty onnistuneesti hyödyksi opetuksen laadun parantamisessa (ks. Hativa 1996).

### Koulutuksen laadun eri mittarit

Vaikka mikään opiskelijapalautekysely ei kykenekään täydellisesti vastaamaan sitä kohtaan esitettyyn yleiseen mittausteoreettiseen kritiikkiin, ei tästä seuraa, ettei opiskelijapalautetta tulisi käyttää osana rahoitusmallia.

Opiskelijapalautekyselymittarin käyttämisen toivottavuus osana rahoitusmallia riippuu siitä, kuinka hyvä mittarista onnistutaan rakentamaan. Kirjoittajien esiin nostamat muutamiin kokemuksiin perustuvat huolet ovat perusteltuja ja vakavasti otettavia, mutta liian valikoivia eivätkä täten yleistettävissä opiskelijapalautteen käytön kannatettavuuteen osana rahoitusmallia yleensä.

Jos opiskelijapalautte nähdään yhtenä koulutuksen laadun mittarina muiden joukossa, eikä ole syytä olettaa sen olevan mittausteoreettisilta ominaisuuksiltaan vaihtoehtoisia mittareita

huonompi, ei sen käyttöä kohtaan tulisi myöskään suhtautua penseästi. Pikemminkin useamman vaihtoehtoisen mittarin käyttäminen tällöin todennäköisesti parantaa koulutuksen laadun kokonaismittarin luotettavuutta.

Jos puolestaan nähdään, ettei opiskelijapalautte mittaa koulutuksen laatua täysin samasta näkökulmasta kuin muut koulutuksen laadun mittarit, kuten auditoinnit tai yliopistovertailut, olisi opiskelijapalautteella perusteltu sija osana rahoitusmallia, vaikka koulutuksen laadun mittarina se olisikin hieman muita huonompi. Tällöin on kyse siitä, pidetäänkö opiskelijan korkeakoulukokemusta itsessään niin tärkeänä osana hyvää korkeakoulutusta, että se on hyvä ottaa itsenäisesti huomioon osana rahoitusmallia.

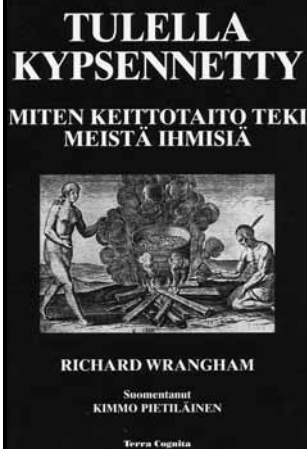
Kunhan varmistetaan, että opiskelijapalautekyselyihin liittyvä kehitystyö on tehty hyvin ja rakennetut mittarit ovat riittävän luotettavia, pidän opiskelijapalautteen käyttämistä osana korkeakoulujen rahoitusmallia kaavailun suuruisena hyvänä asiana.

### Kirjallisuus

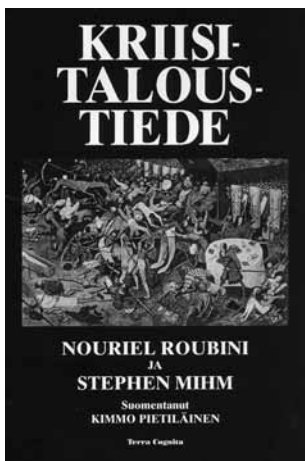
- Hativa, N. (1996) University instructors' ratings profiles: stability over time, and disciplinary differences, *Research in Higher Education*, 37, 341–365.
- Niiniluoto, Ilkka (2002). *Johdatus tieteenfilosofiaan*. Otava.
- Preston, Carolyn C. & Colman, Andrew M. (2000). Optimal number of response categories in rating scales: reliability, validity, discriminating power, and respondent preferences. *Acta Psychologica*, 104 (1), 1–15.
- Ramsay, J. O (1973). The effect of number of categories in rating scales on precision of estimation of scale values. *Psychometrika* 38 (4), 513–532.
- Richardson, John T. E. (2005): Instruments for obtaining student feedback: a review of the literature. *Assessment & Evaluation in Higher Education* 30 (4), 2005, 387–415.

**Kirjoittaja on tutkijana Opiskelun ja koulutuksen tutkimussäätiön Otuksessa ja vastaa ammattikorkeakoulujen valmistumisvaiheen opiskelijapalautekyselyn tutkimuksellisesta tuesta.**

# Parasta suomalaista tietokirjallisuutta



Richard Wrangham  
*Tulella kypsennetty.*  
*Miten keittotaito teki meistä ihmisiä.*  
Ovh. 40,-



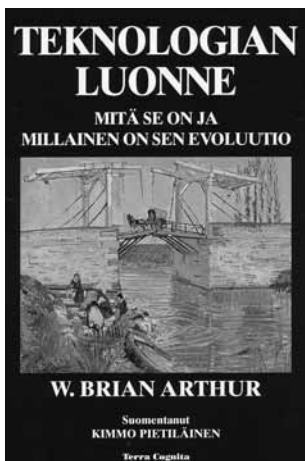
Nouriel Roubini ja  
Stephen Mihn  
*Kriisitaloustiede*  
Ovh. 40,-



Daniel J. Levitin  
*Musiikki ja aivot.*  
*Ihmisen erään pakkomielleen tiedettä*  
Ovh. 40,-



Nicholas Carr  
*Pinnalliset.*  
*Mitä internet tekee aivoillemme*  
Ovh. 40,-



W. Brian Arthur  
*Teknologian luonne.*  
*Mitä se on ja millainen on sen evoluutio*  
Ovh. 40,-



Stuart A. Kauffman  
*Pyhän uudelleen keksiminen.*  
*Uusi näkemys luonnontieteestä, järjestä ja uskonnosta*  
Ovh. 40,-

Kirjakaupasta tai suoraan kustantajalta

TERRA COGNITA OY  
[www.terracognita.fi](http://www.terracognita.fi)