

# Vertaisarvioinnin luotettavuus MOOC-kursseilla

22.6.2023

tags: MOOC-kurssi, vertaisarviointi

*Arho Toikka, vanhempi yliopistonlehtori, valtiotieteellinen tiedekunta, Helsingin yliopisto*  
arho.toikka@helsinki.fi

*Marko Salonen, yliopistonlehtori, yhteiskuntatieteiden tiedekunta, Tampereen yliopisto*  
marko.salonen@tuni.fi

*Liina Sointu, yliopisto-opettaja, yhteiskuntatieteiden tiedekunta, Tampereen yliopisto*  
liina.sointu@tuni.fi



## KEHITTÄMINEN JA KOKEILUT

([https://yliopistopedagogiikka.files.wordpress.com](https://yliopistopedagogiikka.files.wordpress.com/2023/06/yp-osastotunnus-kehittaminen-ja-kokeilut_page_013-4.png)

[/2023/06/yp-osastotunnus-kehittaminen-ja-kokeilut\\_page\\_013-4.png](https://yliopistopedagogiikka.files.wordpress.com/2023/06/yp-osastotunnus-kehittaminen-ja-kokeilut_page_013-4.png))

Verkkokurssien vertaisarvioinnille aukeaa uusia mahdollisuuksia, kun kurssiopiskelijoita on paljon. Tilastolliset kriteerit auttavat havaitsemaan opiskelijoiden poikkeavia suorituksia, arvioita ja arvioitsijoita.

Vertaisarvioinnin luotettavuuden ja tasapuolisuuden takaaminen edellyttää arvioinnin huolellista etukäteissuunnittelua.

Vertaisarvioinnin puoliautomasointi muuttaa opettajan työnkuvaa, mutta ei vähennä työmäärää.

Tässä artikkelissa esittelemme työkaluja vertaisarvioinnin luotettavuuden ja tasapuolisuuden takaamiseksi opiskelijamäärältään suurella verkkokurssilla. Tarkastelukohteenamme on Yhteiskunta tänään -MOOC-kurssi, jonka perusteella Helsingin yliopiston sosiaalitieteiden kandiohjelman ja Tampereen yliopiston yhteiskuntatutkimuksen tutkinto-ohjelman opiskelijoista osa on valittu vuodesta 2020 alkaen, ja vuodesta 2022 alkaen myös Turun yliopiston sosiaalitieteiden ohjelman. Kurssin aloittaa joka kevät noin 1 500 opiskelijaa. Heistä noin tuhat suorittaa kurssin hyväksytysti, ja noin sata saa lopulta opiskelupaikan. Suuren mittakaavansa ansiosta MOOC-kurssi tarjoaa mahdollisuuden vertaisarvioinnin laadun ja luotettavuuden arviointiin, jossa voidaan hyödyntää suurta osallistujamäärää ja tilastollisten analyysien mahdollisuutta.

## Vertaisarvioinnin luotettavuuden arviointi tutkimuksen valossa

Tutkimustiedon valossa vertaisarviointia pidetään hyvänä arviointikeinona: se edistää oppimista (Tenório, Bittencourt, Isotani & Silva, 2016), parantaa metakognitiivisia taitoja (McGarr & Clifford, 2013; Panadero & Alqassab, 2019) ja lisää sitoutumista kurssin suorittamiseen (Formanek, Wenger, Buxner, Impey & Sonam, 2017). Opiskelijoiden näkemykset ovat kuitenkin kriittisiä: opiskelijapalautteessa usein epäillä opiskelijoiden kykyä arvioida tehtäviä sekä vertaisarvioinnin tasapuolisuutta. Myös

opetushenkilökunnan luottamus vertaisarvioinnin laatuun vaihtelee, vaikka erityisesti useamman opiskelijan arvioista koostuva yhdistetty arvio on havaittu hyväksi (Bachelet, Zongo & Bourelle, 2015; Suen, 2014).

Vertaisarvioinnin luotettavuutta on tutkimuksissa lähestytty validiteetin ja reliabiliteetin näkökulmista (esim. Cho, Schunn & Wilson, 2006). Validiteettia on arvioitu vertaamalla yhden tai useamman opiskelijan arvioita yksittäiseen opettajan arvioon (esim. Garcia-Loro, Martin, Ruipérez-Valiente, Sancristobal & Castro, 2020). Tällöin opettajan arvio oletetaan oikeaksi, vaikka todellisuudessa opettajien arvioinnit vaihtelevat verrattain paljon (Bachelet ym., 2015). Reliabiliteettia taas arvioidaan vertaamalla useamman opiskelijan arvioita keskenään (esim. Garcia-Loro ym., 2020). Perinteisessä pienten ryhmien opetuksessa vertaisarvioinnin on todettu vastaavan opettaja-arviointia varsin hyvin (Dochy, Segers & Sluijsmans, 1999; Lindblom-Ylänne, Pihlajamäki & Kotkas, 2006; Sadler & Good, 2006; Topping, 2018). MOOC-konteksti on erilainen muun muassa opiskelijoiden vaihtelevien taustojen takia, ja opiskelijoiden arvioiden välillä on havaittu vaihtelua (Garcia-Loro ym., 2020; Luo, Robinson & Park, 2014).

Tutkimuskirjallisuudessa on esitelty parhaita käytänteitä validiteetin ja reliabiliteetin vahvistamiseksi niin yleisesti korkeakouluopetuksessa kuin isojen verkkokurssien ja MOOC-kurssien yhteydessä. Ensinnäkin vertaisarvioiden määrä arvioitavaa tehtävää kohden on olennainen sekä arvioinnin tarkkuuden että arvioinnin luotettavuuden arvioimisen kannalta. Jokaisesta lisäarviosta on hyötyä niin arvion tarkkuudelle kuin arvioivalle opiskelijalle, mutta molemmissa rajahyöty pienenee nopeasti: arvio ei enää merkittävästi tarkkene ja arvioivan opiskelijan oppimistavoitteet saavutetaan. Riittävänä pidetty arvioijien määrä on 3–5 (Bachelet ym., 2015; Luo ym., 2014).

Toiseksi on tärkeää tunnistaa poikkeavia arvosteluja, arvostelijoita ja poikkeavasti arvioituja tehtäviä sekä luoda toimintatapa sille, miten poikkeamat käsitellään ja arvioinnin oikeellisuus varmistetaan. Isoilla verkkokursseilla arvioita voi kertyä tuhansia, ja siksi niitä voidaan tarkastella tilastollisesti ja algoritmisesti. Vaihtelun avulla voidaan tunnistaa sellaisia tilanteita, joissa on riski, että arviointi on mennyt vikaan. Useammasta raakapisteytyksestä laskettavaan varsinaiseen arvioon on kehitetty myös menetelmiä, kuten bayesläisiä tilastollisia menetelmiä arvioijakohtaisen vääristymän selvittämiseksi (Goldin, 2012), mutta yksinkertainen ja läpinäkyvä vertaisarvioiden keskiarvo on todettu myös riittäväksi (Bachelet ym., 2015). Mediaani on tosin mahdollisesti vielä parempi, koska se poistaa automaattisesti äärimmäisiä poikkeamia (Kulkarni ym., 2013). Tärkeimpiä ovat ehkä kuitenkin ne kriteerit, joilla poikkeava arvio tunnistetaan. Bachelet ja kollegat (2015) tarkentavat sitä, milloin tulee siirtyä automaattisesti opettaja-arviointiin: kun vertaisarvioita on alle kaksi, kun arvioista ei muodostu konsensusta tai kun arviossa annetaan nolla pistettä.

Kolmanneksi vertaisarvioinnin luotettavuuteen vaikuttavat myös arvioitava tehtävä ja arviointikriteerit eli mitä ja miten arvioidaan. Esimerkiksi Freemanin ja Parks (2010) mukaan opiskelija-arviot ovat sitä lähempänä ammattilaisarvioita, mitä yksinkertaisemmasta tiedosta on kyse. Täysin yksimielistä kirjallisuutta ei ole – esimerkiksi Freemanin ja Parks (2010) tutkimuksessa opiskelija-arvioitsijoiden keskiarvo oli hieman korkeampi kuin ammattilaisarvioitsijoiden. Joka tapauksessa tutkimuksessa on tuotu esiin, että vertaisarvioinnin laatua parantaa selkeä ohjeistus, sillä opiskelijoiden välillä voi olla suuria eroja eikä suurilla verkkokursseilla opiskelijoiden käytettävissä ole suoraa opettajan tukea (Meek, Blakemore & Marks, 2017; Suen, 2014).

## Vertaisarvioinnin toteutus ja luotettavuus Yhteiskunta

# tänään -kurssilla

Tarkastelussamme oleva Yhteiskunta tänään -MOOC-kurssi on opetussuunnitelmien mukainen sosiaalitieteiden johdantokurssi, mutta sitä hyödynnetään myös sosiaalitieteiden ja yhteiskuntatutkimuksen opiskelijavalinnassa Helsingin, Tampereen ja Turun yliopistoissa. Kurssista parhaiten suoriutuneet etenevät toiseen hakuvaiheeseen, jossa varsinainen opiskelijavalinta tehdään. Kurssin tarkoituksena on avata opiskelijoille näköaloja sosiaalitieteiden ja yhteiskuntatutkimuksen tutkimusteemoihin ja tutustuttaa yliopisto-opiskelun käytänteisiin.

Yhteiskunta tänään -kurssin arviointi koostuu erilaisista pisteillä arvioitavista tehtävistä. Osa arvioinnista on automatisoituja, osa manuaalisia. Yhden arvioitavan tehtävän pisteytys perustuu vertaisarvioinnille. Kyseinen tehtävä on toteutettu Moodlen työpaja-aktiviteettina, jossa opiskelija tuottaa ohjeistuksen mukaisesti lyhyen esseetekstin. Tehtävänannossa on määritelty tekstin muotoa ja sisältöä – esimerkiksi vaadittu määrittelemään tiettyjä käsitteitä ja esittämään oma soveltava esimerkki. Vertaisarvioinnissa käytetyt kriteerit koskevat sitä, noudattaako teksti ohjeita, mutta ei sitä, täyttyvätkö kriteerit erinomaisesti, hyvin tai välttävästi. Jokainen oman tekstin palauttanut opiskelija arvioi kolmen satunnaisesti arvotun muun opiskelijan tehtävät.

Vertaisarviointitehtävän pedagogisena tavoitteena on tukea akateemisen kirjoittamisen opettelua ja siihen liittyvien odotusten käsittelyä. Osallistujat ovat suorittamassa ensimmäistä yliopistokurssiaan, ja tästä syntyy helposti kokemus, että opiskelijan tulee hallita sellaisia kirjoittamisen käytäntöjä, joita ei vielä ole opetettu. Kun ensimmäinen tehtävä on yksinkertainen vertaisarvioitu kirjoitustehtävä, toisten opiskelijoiden tekstien näkeminen auttaa ymmärtämään, että mitään uutta yliopistollista kirjoitustyyliä ei tarvitse keksiä. Vertaisarviointitehtävässä painottuvat formatiiviset tavoitteet, jotka kohdistuvat sekä tehtävän tekijään että arvioitsijaan (ks. Garcia-Loro ym., 2020). Vertaisarviointi parantaa opiskelijoiden itsearviointitaitoja (Liu & Carless, 2006): artikuloidessaan sitä, miksi ja miten toisen opiskelijan tehtävä täyttää arviointikriteerit, opiskelijan näkemys omasta työstään muuttuu. Kokemuksemme mukaan ennen vertaisarvioitua tehtävää enemmistö opiskelijoiden yhteydenotoista koski akateemisen kirjoittamisen osaamista ja yksityiskohtia, mutta tehtävän jälkeen kysymykset koskivat opetuksen sisältöä.

Vaikka Yhteiskunta tänään -kurssin vertaisarviointitehtävän ensisijainen tavoite on formatiivinen, on sillä myös summatiivista merkitystä: vertaisarviointitehtävästä saa alle viisi prosenttia kurssin kokonaispisteistä. Painoarvo on kurssin kokonaisuudessa pieni, mutta arvioinnin oikeellisuus ja opiskelijoiden luottamus sitä kohtaan on valintakurssiasetelmassa poikkeuksellisen tärkeää. Tehtävän sulkeutumisen jälkeen kurssilla käynnistetään seuraavat toimet, joiden avulla tunnistetaan arvioinnin luotettavuuden kannalta pulmalliset tilanteet, joihin opettajan tarvitsee puuttua. Kurssillamme tämä tarkastelu tehdään tilanteessa, jossa lyhytessettä on noin tuhat ja vertaisarvioita kolmisen tuhatta:

- a) Poikkeava tehtävä: opiskelijan tehtävän vertaisarvioiden keskiarvo on korkea tai matala, käytännössä kaikkien arvosteltujen tehtävien keskiarvosta  $\pm 2$  keskihajontaa.
- b) Poikkeava arvostelija: arvostelijan arvioiden keskiarvo on korkea tai matala, käytännössä kaikkien arvostelijoiden keskiarvosta  $\pm 2$  keskihajontaa.
- c) Poikkeava arvio: yksittäinen arvio on korkea tai matala, esimerkiksi asteikolla 0–10 yli 2 pisteen päässä muista tehtävän arvioista.

#### d) Tekstianalytiikan perusteella poikkeava tehtävä:

teksti on poikkeavan lyhyt, tekstissä on poikkeuksellisen paljon suoria lainauksia kurssimateriaaleista

plagiaatintunnistus vertailuna muualla verkossa oleviin teksteihin yleisillä

plagiaatintunnistustyökaluilla ja vertailuina muihin tehtävän palautuksiin tilastollisella analyysillä – niin sanottu Jaccardin samankaltaisuusindeksi palautettujen tehtävien välillä mittaa tekstien samankaltaisuutta käytettyjen sanojen tasolla ja mahdollistaa luvattoman yhteistyön esille nousemisen

tehtävät, jotka saavat vähemmän kuin kolme vertaisarviota.

Jos tällaisia ongelmia ei havaita, tehtävän pisteytys lasketaan automaattisesti vertaisarvioiden keskiarvosta. Jos pulmia havaitaan, opettajat ottavat sekä arvioidun tekstin että vertaisarviot tarkasteluun ja päättävät, onko tarpeen muuttaa arviointia. Käytännössä opettaja-arvioon päätyy melkein 20 prosenttia tehtävistä, joten esimerkiksi vertaisarviointia käyttämällä opettajan työmäärä ei vähene vaan työnkuva muuttuu: opettajan työ kohdistuu miettimistä vaativiin tapauksiin. Valintakurssilla kriteerit on asetettu tiukasti. Tavallisella kurssilla poikkeavien tehtävien kriteerit voisi asettaa löyhemmin, ja tällöin työmääräkin saattaisi vähetä.

## Lopuksi

Olemme esittäneet toimia, joiden avulla vertaisarvioinnin tarkkuutta ja luotettavuutta voidaan varmistaa MOOC-kurssilla. Näiden toimien perusteella käyttämämme vertaisarvioinnin validiteetti ja reliabiliteetti ovat todennäköisesti yhtä hyviä kuin opettajienkin arviot.

Esittämämme toimet, erityisesti poikkeavien tehtävien, arvioiden, arvioitsijoiden sekä tekstianalytiikan perusteella opettajan arviointia edellyttävien tehtävien tunnistaminen, toimivat hyvin ennakkoon suunniteltujen isojen kurssien arvioinnissa, mutteivät ole suoraan sovellettavissa pienemmille kurseille. Kun luotettavuuden takuuna ovat yksittäisten ongelmatilanteiden tunnistaminen ja niiden ratkaisu, edellyttää se opiskelijatöiden ja vertaisarviointien verrattain suurta kokonaisuutena.

Vaikka vertaisarvioinnin empiirinen luotettavuus ja tarkkuus saataisiin ratkaistua, voi vertaisarviointi herättää myös muita huolia. Negatiivisiin käsityksiin vaikuttavat esimerkiksi kulttuuriset oletukset opettajan ja opiskelijoiden rooleista ja auktoriteettiasemasta (McGarr & Clifford, 2013). Vertaisarvioinnin pedagogisten tavoitteiden avoin julkittuominen ja arvioinnin kriteerien hyvä kirjoittaminen ja selkeä kiinnitys kurssitavoitteisiin tukevat luotettavuuden rakentumista sekä roolien merkitystä luotettavalle arvioinnille.

Suosittellemme vertaisarviointia kaikille opettajille työkaluksi erityisesti MOOC-kontekstiin. Opettajan on kuitenkin tärkeä tehdä näkyväksi sitä, miten mahdollisesti epäonnistunut tai epätarkka vertaisarvio tunnistetaan.

### Lähteet

Bachelet, R., Zongo, D. & Bourelle, A. (2015). *Does peer grading work? How to implement and improve it? Comparing instructor and peer assessment in MOOC GdP*. European MOOCs Stakeholders Summit 2015,

May 2015, Mons, Belgium.

Cho, K., Schunn, C. D. & Wilson, R. W. (2006). Validity and reliability of scaffolded peer assessment of writing from instructor and student perspectives. *Journal of Educational Psychology*, 98(4), 891–901. <https://doi.org/10.1037/0022-0663.98.4.891> (<https://doi.org/10.1037/0022-0663.98.4.891>)

Dochy, F., Segers, M., & Sluijsmans, D. (1999). The use of self-, peer and co-assessment in higher education: A review. *Studies in Higher education*, 24(3), 331–350. <https://doi.org/10.1080/03075079912331379935> (<https://doi.org/10.1080/03075079912331379935>)

Formanek, M., Wenger, M. C., Buxner, S. R., Impey, C. D. & Sonam, T. (2017). Insights about large-scale online peer assessment from an analysis of an astronomy MOOC. *Computers & Education*, 113, 243–262. <https://doi.org/10.1016/j.compedu.2017.05.019> (<https://doi.org/10.1016/j.compedu.2017.05.019>)

Freeman, S. & Parks, J. W. (2010). How accurate is peer grading? *CBE – Life Sciences Education*, 9(4), 482–488. <https://doi.org/10.1187/cbe.10-03-0017> (<https://doi.org/10.1187/cbe.10-03-0017>)

Garcia-Loro, F., Martin, S., Ruipérez-Valiente, J. A., Sancristobal, E. & Castro, M. (2020). Reviewing and analyzing peer review Inter-Rater Reliability in a MOOC platform. *Computers & Education*, 154, 103894. <https://doi.org/10.1016/j.compedu.2020.103894> (<https://doi.org/10.1016/j.compedu.2020.103894>)

Goldin, I. M. (2012). *Accounting for peer reviewer bias with Bayesian models*. Proceedings of the workshop at the 11<sup>th</sup> International Conference on Intelligent Tutoring Systems, June 15, 2012, Chania, Crete, Greece.

Kulkarni, C., Wei, K. P., Le, H., Chia, D., Papadopoulos, K., Cheng, J., Koller, D. & Klemmer, S. R. (2013). Peer and self assessment in massive online classes. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 20(6), 1–31. <https://doi.org/10.1145/2505057> (<https://doi.org/10.1145/2505057>)

Lindblom-Ylänne, S., Pihlajamäki, H. & Kotkas, T. (2006). Self-, peer- and teacher-assessment of student essays. *Active Learning in Higher Education*, 7(1), 51–62. <https://doi.org/10.1177/1469787406061148> (<https://doi.org/10.1177/1469787406061148>)

Liu, N.-G. & Carless, D. (2006). Peer feedback: The learning element of peer assessment. *Teaching in Higher Education*, 11(3), 279–290. <https://doi.org/10.1080/13562510600680582> (<https://doi.org/10.1080/13562510600680582>)

Luo, H., Robinson, A. C. & Park, J.-Y. (2014). Peer grading in a MOOC: Reliability, validity, and perceived effects. *Journal of Asynchronous Learning Networks*, 18(2). <https://doi.org/10.24059/olj.v18i2.429> (<https://doi.org/10.24059/olj.v18i2.429>)

McGarr, O. & Clifford, A. M. (2013). ‘Just enough to make you take it seriously’: Exploring students’ attitudes towards peer assessment. *Higher education*, 65(6), 677–693. <https://doi.org/10.1007/s10734-012-9570-z> (<https://doi.org/10.1007/s10734-012-9570-z>)

Meek, S. E. M., Blakemore, L. & Marks, L. (2017). Is peer review an appropriate form of assessment in a MOOC? Student participation and performance in formative peer review. *Assessment & Evaluation in Higher Education*, 42(6), 1000–1013. <https://doi.org/10.1080/02602938.2016.1221052> (<https://doi.org/10.1080/02602938.2016.1221052>)

Panadero, E. & Alqassab, M. (2019). An empirical review of anonymity effects in peer assessment, peer

feedback, peer review, peer evaluation and peer grading. *Assessment & Evaluation in Higher Education*, 44(8), 1253–1278. <https://doi.org/10.1080/02602938.2019.1600186> (<https://doi.org/10.1080/02602938.2019.1600186>)

Sadler, P. M. & Good, E. (2006). The impact of self-and peer-grading on student learning. *Educational Assessment*, 11(1), 1–31. [https://doi.org/10.1207/s15326977ea1101\\_1](https://doi.org/10.1207/s15326977ea1101_1) ([https://doi.org/10.1207/s15326977ea1101\\_1](https://doi.org/10.1207/s15326977ea1101_1))

Suen, H. K. (2014). Peer assessment for massive open online courses (MOOCs). *The International Review Of Research In Open And Distributed Learning*, 15(3), 312–327. <https://doi.org/10.19173/irrodl.v15i3.1680> (<https://doi.org/10.19173/irrodl.v15i3.1680>)

Tenório, T., Bittencourt, I. I., Isotani, S. & Silva, A. P. (2016). Does peer assessment in on-line learning environments work? A systematic review of the literature. *Computers in Human Behavior*, 64, 94–107. <https://doi.org/10.1016/j.chb.2016.06.020> (<https://doi.org/10.1016/j.chb.2016.06.020>)

Topping, K. (1998). Peer assessment between students in colleges and universities. *Review of Educational Research*, 68(3), 249–276. <https://doi.org/10.2307/1170598> (<https://doi.org/10.2307/1170598>)

from → Yliopistopedagogiikka 2023/1

No comments yet

This site uses Akismet to reduce spam. [Learn how your comment data is processed.](#)

[Pidä blogia WordPress.comissa.](#)