



**TURUN  
YLIOPISTO**

Matemaattis-luonnontieteellinen  
tiedekunta

# **Maantieteen ylioppilaskokeiden arviointi kriteeriperustaisen arviointimatriisin avulla**

Anni Maula

Maantiede (aineenopettajalinja)

Pro gradu -tutkielma

Laajuus: 20 op

15.5.2025

Turku

Turun yliopiston laatujärjestelmän mukaisesti tämän julkaisun alkuperäisyys on tarkastettu

Turnitin OriginalityCheck -järjestelmällä.

Pro gradu -tutkielma

**Pääaine:** Maantiede

**Tekijä:** Anni Maula

**Otsikko:** Maantieteen ylioppilaskokeiden arviointi kriteeriperustaisen arviointimatriisin avulla

**Ohjaajat:** Sanna Mäki, Tua Nylén

**Sivumäärä:** 60 sivua + liitteet 6 sivua

**Päivämäärä:** 15.5.2025

---

Arviointimatriisi on kriteeriperustainen arviointiväline, joka soveltuu monipuolisen osaamisen ja esimerkiksi ajattelun taitojen arviointiin. Useiden oppiaineiden ylioppilaskokeiden arvioinnissa käytetään arviointimatriisia, mutta maantieteen kokeen pisteitys perustuu pääasiassa listamuotoisiin hyvän vastauksen piirteisiin. Tämän tutkimuksen tavoitteena oli tarkastella matriisiarvioinnin soveltuvuutta maantieteen ylioppilaskokeiden arviointiin pisteittämällä syksyn 2023 maantieteen ylioppilaskokeen tehtävän 7 vastauksia tutkimuksessa luodulla arviointimatriisilla sekä kokeen nykyisillä arviointiohjeilla. Tilastollisilla menetelmillä selvitettiin, miten eri arviointimenetelmillä tuotetut pisteet eroavat toisistaan sekä miten yhdenmukaisesti vastauksia voidaan pisteittää luodun arviointimatriisin avulla. Lisäksi vastausten laadullisella sisällönanalyysillä tarkasteltiin, minkälaisia ominaisuuksia niissä vastauksissa on, joiden pisteitä matriisiarviointi muutti eniten.

Tutkimustulosten mukaan matriisiarviointi tuotti keskimäärin matalampia pisteitä kuin nykyinen arviointimenetelmä. Erot eri arviointimenetelmillä annettujen pisteiden välillä olivat tilastollisesti merkitseviä. Tutkimuksen tulosten mukaan arviointimatriisilla toteutettu arviointi oli varsin yhdenmukaista, eikä arvioinnin yhdenmukaisuudessa havaittu tarkasteltujen arviointimenetelmien välillä suurta eroa. Vastausten laadullisessa tarkastelussa havaittiin, että matriisiarviointi laski eniten erityisesti sellaisten vastausten pisteitä, joissa esiintyi puutteita huomioiden perusteluissa, käsitteiden käytössä, maantieteen näkökulmien hallinnassa tai havaintojen kytkemisessä maantieteellisiin ilmiöihin tai prosesseihin. Ne vastaukset, joiden pisteitä matriisiarviointi nosti eniten, olivat keskimäärin tasaisen laadukkaita jokaisen tarkastellun ominaisuuden osalta.

Tutkimuksen tulosten mukaan tutkimuksessa luotu arviointimatriisi voisi soveltua hyvin maantieteen ylioppilaskokeiden arviointiin, mutta aihetta olisi tärkeää tutkia lisää esimerkiksi monipuolisemmalla ja laajemmalla aineistolla sekä osallistamalla tutkimukseen useampia arvioijia. Tutkimuksen tuloksia voidaan hyödyntää erityisesti maantieteen ylioppilaskokeiden arvioinnin kehittämistarkoituksissa.

---

**Avainsanat:** maantiede, ylioppilaskirjoitukset, arviointimatriisi, kriteeriperustainen arviointi, reliabiliteetti

Master's thesis

**Subject:** Geography

**Author:** Anni Maula

**Title:** Assessment of the geography matriculation exams using a criterion-based scoring rubric

**Supervisors:** Sanna Mäki, Tua Nylén

**Number of pages:** 60 pages + 6 appendices

**Date:** 15.5.2025

---

Scoring rubrics are criterion-referenced assessment tools that are often used, for example, as a tool to support learning and to assess a wide range of skills, such as thinking skills. Although scoring rubrics are used in the assessment of many subjects in the Finnish matriculation exams, the assessment of geography exams is mainly based on lists of what is required in the answers. The aim of this study was to examine the suitability of scoring rubrics for the assessment of the geography matriculation exams by scoring the answers to question 7 of the autumn 2023 geography matriculation exam using the scoring rubric created in the study and the current grading instructions for the exam. Statistical methods were used to investigate how the scores produced by the different assessment methods differ from each other and how consistently the answers can be scored using the scoring rubric. A qualitative content analysis of the answers was used to examine the characteristics of the answers whose scores were most affected by the rubric assessment.

The results of the study showed that the assessment based on the scoring rubric produced lower scores on average than the current assessment method. The differences between the scores produced by the different assessment methods were statistically significant. According to the results of the study, the assessment conducted with the scoring rubric was fairly consistent, and the consistency of the assessment did not differ considerably between the assessment methods examined. A qualitative analysis of the answers showed that assessment using a scoring rubric lowered the scores most particularly for responses with deficiencies in the reasoning of observations, the use of concepts, the proficiency in geographic perspectives or the connecting of observations to geographical phenomena or processes. The answers with the highest score increase in the assessment using the scoring rubric were, on average, of consistent high quality for each of the characteristics considered.

According to the results of this study, the scoring rubric created in the study could be well suited for the assessment of Finnish matriculation exams in geography, although it would be important to study the topic further for example with more diverse and broader data and by involving more raters in the study. The results can be used to improve the assessment of geography in schools, and particularly the assessment of geography in the Finnish matriculation exams.

---

**Key words:** geography, matriculation examination, scoring rubric, criteria-referenced assessment, reliability

# Sisällysluettelo

<b>1</b>	<b>Johdanto</b>	<b>1</b>
<b>2</b>	<b>Tutkimuksen tausta ja teoreettinen viitekehys</b>	<b>3</b>
<b>2.1</b>	<b>Arviointi</b>	<b>3</b>
2.1.1	Arvioinnin tehtävät ja tarkoitus	3
2.1.2	Laadukkaan arvioinnin ominaisuuksia	4
2.1.3	Suhteellinen ja kriteeriperustainen arviointi	6
<b>2.2</b>	<b>Arviointimatriisi arvioinnin välineenä</b>	<b>8</b>
<b>2.3</b>	<b>Ylioppilaskokeiden arviointi</b>	<b>11</b>
2.3.1	Maantieteen ylioppilaskokeet ja niiden arviointi	11
2.3.2	Syksyn 2023 maantieteen ylioppilaskokeen tehtävän 7 pisteitysohjeet	12
<b>3</b>	<b>Aineisto ja menetelmät</b>	<b>15</b>
<b>3.1</b>	<b>Tutkimuksen vaiheet</b>	<b>15</b>
<b>3.2</b>	<b>Aineisto</b>	<b>15</b>
<b>3.3</b>	<b>Menetelmät</b>	<b>17</b>
3.3.1	Aineiston arviointi	17
3.3.2	Arviointimatriisin luominen	18
3.3.3	Tilastollinen tarkastelu	22
3.3.4	Vastausten analysointi ja laadullinen sisällönanalyysi	24
<b>4</b>	<b>Tulokset</b>	<b>26</b>
<b>4.1</b>	<b>Arviointimenetelmän vaikutus pistemäärään</b>	<b>26</b>
4.1.1	Koko tehtävän tarkastelu	26
4.1.2	Osatehtäväkohtainen tarkastelu	28
<b>4.2</b>	<b>Arviointimenetelmien arvioitsijareliabiliteetit</b>	<b>28</b>
4.2.1	Arviointimatriisin arvioitsijareliabiliteetti	28
4.2.2	Nykyisten arviointiohjeiden arvioitsijareliabiliteetti	32
<b>4.3</b>	<b>Arviointimatriisin kriteerikohtainen reliabiliteetti</b>	<b>34</b>
<b>4.4</b>	<b>Ominaisuuksia vastauksissa, joissa arviointimenetelmien välillä esiintyi pisteroja</b>	<b>35</b>
4.4.1	Niiden vastausten ominaisuuksia, joiden pisteitä matriisiarviointi laski eniten	35
4.4.2	Niiden vastausten ominaisuuksia, joiden pisteitä matriisiarviointi nosti eniten	39
<b>5</b>	<b>Keskustelu</b>	<b>43</b>
<b>5.1</b>	<b>Matriisiarviointi tuotti nykyohjeistusta matalampia pisteitä</b>	<b>43</b>

<b>5.2</b>	<b>Arviointimatriisilla toteutetun arvioinnin yhdenmukaisuus vaikuttaa lupaavalta</b>	<b>45</b>
<b>5.3</b>	<b>Matriisiarviointi laski erityisesti listamaisten ja arkikielisten vastausten</b>	<b>47</b>
	<b>pisteitä</b>	
<b>5.4</b>	<b>Tutkimuksen luotettavuus ja jatkotutkimusmahdollisuudet</b>	<b>49</b>
<b>6</b>	<b>Johtopäätökset</b>	<b>53</b>
	<b>Kiitokset</b>	<b>54</b>
	<b>Lähteet</b>	<b>55</b>
	<b>Liitteet</b>	<b>61</b>
	<b>Liite 1. Tehtävän 7 nykyiset arviointiohjeet eli hyvän vastauksen piirteet</b>	<b>61</b>
	<b>Liite 2. Vastausten sisällönanalyysin analyysirunko</b>	<b>65</b>

# 1 Johdanto

Suomalaisen lukiokoulutuksen tehtävänä on vahvistaa opiskelijoiden laaja-alaista yleissivistystä (Lukion opetussuunnitelman... 2019). Lukiokoulutus kehittää monipuolisesti opiskelijoiden tietoja, taitoja ja arvoja, joita opetussuunnitelmassa määritellään sekä yleisellä tasolla että oppiainekohtaisesti. Näiden osaamiselle asetettujen tavoitteiden saavuttaminen tehdään näkyväksi summatiivisella arvioinnilla, jota toteutetaan opintojaksojen lopussa sekä lukiokoulutuksen päätteeksi järjestettävillä ylioppilaskokeilla (Lukion opetussuunnitelman... 2019; Ouakrim-Soivio 2015). Ylioppilaskokeet edustavat niin kutsuttua *high stakes* -arviointia eli arviointimallia, jossa arviointitulokset vaikuttavat merkittävästi esimerkiksi opiskelijoiden koulutuksellisiin tavoitteisiin (ks. Acosta ym. 2020). Ylioppilastutkinnon arvosanat toimivat korkeakoulujen opiskelijavalintojen perusteena (Ouakrim-Soivio 2015), minkä vuoksi niiden merkitys ylioppilaskokelaiden tulevaisuuteen on kiistaton. *High stakes* -arvioinnilta edellytetään sen seurausten merkittävyyden vuoksi useita asioita, kuten korkeaa reliabiliteettia ja validiteettia (French 2023). Ylioppilaskokeiden arviointiin kohdistetaan siis perustellusti lukuisia vaatimuksia, joten on tärkeää, että kokeiden arvioinnissa käytettävällä arviointimenetelmällä pystytään vastaamaan näihin odotuksiin parhaalla mahdollisella tavalla.

Arvioinnin vaatimukseen voidaan vastata eri tarkoituksiin soveltuvilla arviointimenetelmillä, ja arviointimatriisit ovat esimerkkejä suoritusten laadulliseen arviointiin kehitetyistä ratkaisuksista (Panadero ym. 2023). Arviointimatriisit eli rubriikit ovat kriteeriperustaisia arviointivälineitä, joilla on lukuisia käyttötarkoituksia: niiden hyödyt on tunnistettu summatiivisessa, formatiivisessa sekä diagnostisessa arvioinnissa, ja niiden on todettu usein parantavan esimerkiksi arvioinnin reliabiliteettia ja läpinäkyvyyttä sekä kehittävän opiskelijoiden itsearvioinnin taitoja (Hadibarata ym. 2024; Brookhart 2013; Cantera ym. 2021; Jonsson & Svingy 2007; Panadero & Jonsson 2013). Ylioppilaskokeiden arvioinnin kannalta on kuitenkin erityisen mielenkiintoista tarkastella arviointimatriisien käytettävyyttä monipuolisen osaamisen, kuten ajattelun taitojen, arvioinnissa. Aiempien tutkimuksien mukaan arviointimatriisit voivat toimia ajattelun taitojen arvioinnissa erinomaisesti, minkä lisäksi ne voivat edistää ajattelun taitojen kehittymistä jo oppimisprosessin aikana (Al-Salmani & Thacker 2021; Reynders ym. 2020).

Lukion opetussuunnitelman perusteiden (2019) mukaan maantieteen oppiaineessa on tarkoituksena arvioida tiedollisen osaamisen lisäksi myös maantieteellistä ajattelua sekä maantieteellisten taitojen osaamista ja soveltamista, ja koska vastaavan osaamisen arvioiminen arviointimat-

riisilla on todettu aiemmassa tutkimuskirjallisuudessa onnistuneeksi, on arviointimatriisiin soveltumista maantieteen ylioppilaskokeiden arviointiin mielenkiintoista tutkia tarkemmin. Arviointimatriiseja hyödynnetään useiden oppiaineiden, kuten historian, yhteiskuntaopin ja terveystiedon, ylioppilaskokeiden arvioinnissa (Hyvän... Historia 2025; Hyvän... Yhteiskuntaoppi 2025; Hyvän... Terveystieto 2025). Maantieteen hyvän vastauksen piirteet koostuvat sen sijaan yleisistä arvioinnille asetetuista lähtökohdista sekä tehtäväkohtaisista pisteitysohjeista, jotka ovat muodoltaan pääasiassa listamaiset (Hyvän... Maantiede 2025). Tässä tutkimuksessa tarkastellaan, olisiko myös maantieteen ylioppilaskokeita mahdollista arvioida onnistuneesti arviointimatriisin avulla sekä selvittää, voitaisiinko arviointimatriisiin perustuvalla arvioinnilla pisteittää erilaista osaamista kuin nykyisillä arviointiohjeilla.

Arviointimatriisien käyttöä on tutkittu paljon (Dawson 2017). Tutkimuskirjallisuus matriisien soveltamisesta maantieteen oppiaineen arviointiin on kuitenkin vähäistä, eikä arviointimatriisien käytöstä maantieteen ylioppilaskokeiden arvioinnissa ole tehty aiemmin tutkimusta. Tämän tutkimuksen päätavoitteena onkin tarkastella matriisiarvioinnin soveltuvuutta maantieteen ylioppilaskokeiden arviointiin. Tutkimuksessa luotiin arviointimatriisi, jolla pyrittiin huomioimaan vastauksissa esitetty monipuolinen osaaminen ja ajattelun taidot nykyisin käytössä olevia arviointiohjeita paremmin. Matriisin soveltuvuutta testattiin arvioimalla aitoja ylioppilaskoevastauksia ja vertaamalla matriisilla tuotettuja pisteitä nykyisillä arviointiohjeilla tuotettuihin pisteisiin. Tutkimuksessa pyrittiin selvittämään sekä laadullisilla että määrällisillä menetelmillä, millaisia vaikutuksia matriisiarvioinnilla voisi olla maantieteen ylioppilaskokeiden arviointiin. Tutkimuksen tuloksia voidaan hyödyntää maantieteen arvioinnin ja erityisesti maantieteen ylioppilaskokeiden kehittämistarkoituksissa.

Tämän tutkielman tutkimuskysymykset ovat seuraavat:

1. Miten arviointimatriisiin perustuvat pisteet eroavat nykyisiin arviointiohjeisiin perustuvista pisteistä?
2. Miten yhdenmukaisesti eri arvioijat voivat pisteittää maantieteen ylioppilaskoevastauksia kriteeriperustaisen arviointimatriisin avulla?
3. Minkälaisia ominaisuuksia niissä vastauksissa on, joissa erot arviointimatriisiin perustuvien pisteiden ja nykyisiin arviointiohjeisiin perustuvien pisteiden välillä ovat suurimpia?

## 2 Tutkimuksen tausta ja teoreettinen viitekehys

### 2.1 Arviointi

#### 2.1.1 Arvioinnin tehtävät ja tarkoitus

Arvioinnilla tarkoitetaan toimintaa, jossa etukäteen asetettuja tavoitteita ja saavutettuja tuloksia vertaillaan toisiinsa (Ouakrim-Soivio 2015). Arvioinnilla on lukuisia tehtäviä, jotka määräytyvät esimerkiksi sen tarkoituksen, kohteen, arviointitiedon käyttötarkoituksen sekä arviointitavan mukaan. Lukion opetussuunnitelman perusteiden (2019) mukaan arvioinnilla on lukiokoulutuksessa kaksi keskeistä tehtävää: oppimisen tukeminen ja ohjaaminen sekä osaamisen ja oppimisen tavoitteiden saavuttamisen näkyväksi tekeminen. Oppimisen tukemisen tehtävää toteuttaa erityisesti formatiivinen eli oppimisprosessin aikainen arviointi, jonka tarkoituksena on varmistaa oppimisen edistyminen jatkuvan palautteen avulla (Jakku-Sihvonen 2013; Lukion opetussuunnitelman... 2019). Formatiiivisella arvioinnilla pyritään mahdollistamaan oppimiselle asetettujen tavoitteiden saavuttaminen, ja luonteeltaan se on kannustavaa (Lukion opetussuunnitelman... 2019; Luostarinen & Ouakrim-Soivio 2019). Osaamisen arvioinnin tehtävää puolestaan toteutetaan summatiivisella arvioinnilla eli arvioinnilla, joka tapahtuu yleensä oppimiskokonaisuuden tai opintojakson päätyttyä (Lukion opetussuunnitelman... 2019; Ouakrim-Soivio 2015). Sillä kuvataan kaikille arviointitietoa tarvitseville arvioitavan edistyminen ja suoritustaso, ja yleensä se toteutetaan numeroarvosanalla. Osaamisen arviointi on luonteeltaan toteavaa (Luostarinen & Ouakrim-Soivio 2019). Näiden kahden tehtävän lisäksi arvioinnille voidaan nimetä myös useita muita tehtäviä, joita ovat esimerkiksi tiedon ja palautteen antaminen opettajalle sekä opettajan tukeminen opetuksen suunnittelussa (Ouakrim-Soivio 2015).

Arvioinnin tehtävät liittyvät keskeisesti siihen, mikä arvioinnin tarkoitus on (Ouakrim-Soivio 2015). Aiheen tutkimuskirjallisuudesta voidaan nimetä kolme arvioinnin tarkoitusta eli funktiota, ja näitä ovat validointi ja sertifiointi, kontrollointi sekä oppimisen tukeminen (Nieminen 2019). Summatiiviseen arviointiin liittyvällä arvioinnin validointitarkoituksella tarkoitetaan arvosanan muodostamista arvioitavalle tiettyjen kriteerien perusteella. Tämä funktio liittyy yleensä *high stakes* -arviointiin eli arviointiin, joka vaikuttaa arvioitavaan merkittävästi. Esimerkiksi valtakunnalliset kokeet, joista saadut arvosanat toimivat kriteerinä esimerkiksi arviotavien akateemiselle etenemiselle tai koulujen tuloksellisuuden mittaamiselle, edustavat *high stakes* -arviointia (Acosta ym. 2020). Validointi- ja sertifiointifunktio palvelee ensisijaisesti

koulutuksellisia instituutioita, sillä sen tarkoituksena on asettaa arvioitavat järjestykseen ja sertifioida heistä vain osa (Nieminen 2019).

Arvioinnin kontrollointitarkoituksella tarkoitetaan sitä, että arvioinnin avulla usein ohjataan, mitä ja miten arvioitavien kuuluu oppia (Nieminen 2019). Tämä funktio liittyy arviointiin lähes aina, sillä formatiivisella arvioinnilla ja palautteella ohjataan oppimista ja oppimisen kohteita jo oppimisprosessin aikana, ja toisaalta myös esimerkiksi summatiivisen arvioinnin arviointikriteerit voivat ohjata opiskelijoiden opiskelua arviointikriteerien osoittamaan suuntaan (Luostarinen & Ouakrim-Soivio 2019; Nieminen 2019; Norton 2004). Arviointi voi vaikuttaa opettamiseen ja oppimiseen usein myös tahattomasti, jolloin puhutaan niin sanotusta takaistusvaikutuksesta (engl. *washback* tai *backwash effect*) (Abdallah ym. 2022). Arviointi voi vaikuttaa esimerkiksi opettajien käyttämiin opetusmenetelmiin sekä kohdistaa opetuksen koskemaan niitä aihealueita, joiden osaamista todennäköisesti mitataan ulkopuolisen tahon laatimassa summatiivisessa arvioinnissa (Rahman ym. 2021; Rind & Mari 2019). Vastaavasti arviointi voi kontrolloida myös opiskelijoiden opiskelukohteita- ja tapoja tahattomasti takaistusvaikutuksen kautta (Alqahtani 2021). Kolmannella arvioinnin funktiolla eli oppimisen tukemisella puolestaan tarkoitetaan Lukion opetussuunnitelman perusteiden (2019) nimeämän arvioinnin tehtävän tavoin arviointia, jonka tarkoituksena on tukea arvioitavan oppimista ja opiskelutaitoja (Nieminen 2019). Oppimisen tukemisen tarkoitus on määritelmällisesti ristiriidassa kahden muun arvioinnin funktion kanssa, sillä se painottaa oppimisprosessia lopputuloksen sijaan.

### 2.1.2 Laadukkaan arvioinnin ominaisuuksia

Arvioinnin laatua voidaan tarkastella useiden laatu-kriteerien avulla, ja Ouakrim-Soivion (2015) mukaan keskeisiä arviointitoiminnalle asetettuja vaatimuksia ovat pyrkiminen arvioinnin reliabiliuteen, validiuteen, objektiivisuuteen sekä oikeudenmukaisuuteen. Kriteerien täyttymistä voidaan arvioida sekä tilastollisesti että laadullisesti (Gerritsen-van Leeuwenkamp ym. 2017). Arvioinnin laatuun vaikuttavat kaikkien arviointikäytäntöjen osa-alueiden, kuten arviointiprosessin sekä arvioitavien tehtävien, laatu. Laadukkaan arvioinnin käsite liittyy aina muun muassa vallitsevaan oppimiskäsitykseen ja arvopohjaan, minkä vuoksi se ei ole käsitteenä täysin neutraali (Nieminen 2019). Esimerkiksi sosiokonstruktivistisessä oppimiskäsityksessä korostetaan arvioinnin merkitystä oppimisen tukemisessa sen sijaan, että arvioinnin laatu liittyisi ainoastaan summatiivisen arvioinnin onnistumiseen (Gerritsen-van Leeuwenkamp ym. 2017).

Reliabiliteetti eli reliabelius kuvaa arvioinnin toistettavuutta, pysyvyyttä ja satunnaisvirheettömyyttä (Ouakrim-Soivio 2015). Se ilmaisee, onko arviointi yhdenmukaista sekä arvioijasta

ja arvioinnin olosuhteista riippumatonta (Luostarinen & Nieminen 2019). Reliabiliteetti koostuu kahdesta osatekijästä: stabiliteetista eli pysyvyydestä ajassa sekä konsistenssista eli yhtenäisyydestä (Ouakrim-Soivio 2015). Stabiliteetti viittaa siihen, että koe tuottaa saman arviointituloksen eri kerroilla esimerkiksi olosuhteista, arvioijasta tai ajankohdasta riippumatta (Biggs & Tang 2011; Ouakrim-Soivio 2015). Konsistenssilla tarkoitetaan puolestaan sitä, että kokeen kaikki osat mittaavat kokonaisuudessaan samaa asiaa yhdenmukaisesti. Reliabiliteettia voidaan tarkastella myös arvioitsijareliabiliteetin (engl. *inter-rater reliability*) sekä yhden arvioijan toteuttaman arvioinnin reliabiliteetin (engl. *intra-rater reliability*) kannalta (Biggs & Tang 2011). Arvioitsijareliabiliteetissa on kyse siitä, että eri arvioijat arvioivat saman suorituksen yhdenmukaisesti niin, että arvioinnin tulokset eivät vaihtele arvioijan mukaan (Biggs & Tang 2011; Ouakrim-Soivio 2015). Yhden arvioijan toteuttaman arvioinnin reliabiliteetti puolestaan tarkoittaa sitä, että sama arvioija arvioi saman suorituksen yhdenmukaisesti eri kerroilla ilman, että arvioinnin tulos vaihtelee. Arvioinnin reliabiliteettia voidaan mitata eri tarkoituksiin sopivilla tilastollisilla tunnusluvuilla (Stemler 2004). Tunnusluvut voivat kuvata esimerkiksi arvioijien yksimielisyyden astetta, arvioijien antamien pisteiden välistä korrelaatiota tai arvioijien välisiä eroja arvioinnin ankaruudessa.

Validiteetti eli *validius* tarkoittaa arvioinnin osuvuutta ja pätevyyttä, eli se kuvaa, mitataanko arvioinnilla sitä, mitä sillä on tarkoituskin mitata (Ouakrim-Soivio 2015). Osuva eli korkean validiteetin omaava arviointi tuottaa tietoa juuri siitä, mistä sen on tarkoituskin (Räkköläinen 2013). Validiteettia voidaan tarkastella ulkoisena ja sisäisenä validiteettina (Ouakrim-Soivio 2015). Ulkoisen validiteetin käsite liittyy pääasiassa tutkimusten luotettavuuden arviointiin, sillä sen avulla mitataan, ovatko tulokset yleistettävissä johonkin perusjoukkoon. Esimerkiksi koulujen arviointityötä suorittavilla opettajilla ei ole perusjoukkoa, johon arviointitulokset kuuluisi yleistää, joten ulkoinen validiteetti ei varsinaisesti liity koulussa toteutettavaan arviointiin. Sisäinen validiteetti puolestaan kuvaa arvioinnin tai tutkimuksen omaa luotettavuutta, ja sitä voidaan tarkastella sisältö-, kriteeri- ja käsitevaliditeetin kautta. Sisällön validiustarkastelun avulla kuvataan, onko arviointimittari kohdealueena olevan sisällön kannalta edustava sekä ovatko käytetyt käsitteet teorian mukaisia ja operationalisoitu eli muutettu empiirisesti mitattavaan muotoon oikein. Kriteerivaliditeetissa arviointimittarilla tuotettua tulosta verrataan johonkin validiuden kriteerinä toimivaan arvoon, ja käsitevaliditeetti kuvaa tutkimuksen kohteen käsitteellistämistä sekä mittauksen kytkeytymistä teoreettiseen viitekehukseen. Koulussa toteutettavan arvioinnin kannalta validiteetti tarkoittaa esimerkiksi sitä, että osaamisen arvioinnissa huomioidaan opetussuunnitelmassa asetetut tavoitteet ja sisällöt monipuolisesti, eivätkä asiat

tai ominaisuudet, joita arvioinnissa ei ole tarkoitus mitata, vaikuta arviointiin (Luostarinen & Nieminen 2019).

Arvioinnin objektiivisuus tarkoittaa, että arviointipäätös perustuu täysin arviointivälineeseen arvioijan subjektiivisen mielipiteen sijaan (Keurulainen 2013). Objektiivisuuden periaatteen mukaisesti arviointi ei perustu arvioijan tulkintaan, eivätkä esimerkiksi arvioijan henkilökohtaiset mielipiteet ja näkemykset, havainnointitapa tai mieliala vaikuta arviointiin (Keurulainen 2013; Ten Cate & Regehr 2019). Keurulaisen (2013) mukaan pyrkimys arvioinnin objektiivisuuteen liittyy erityisesti normatiiviseen arviointiin sekä määrälliseen kriteeriperustaiseen arviointiin. Eroja arvioinnin objektiivisuudessa voi esiintyä esimerkiksi eri arviointimenetelmien välillä (Kolaree & Ninčević 2022). Esimerkiksi monivalintakokeiden arvioinnissa arvioinnin lopputulos muodostuu suoraan opiskelijan antamien oikeiden vastausten määrästä, ja näin arvioijan tulkinta ei voi vaikuttaa päätökseen (Keurulainen 2013). Avoimien vastausten arvioinnissa puolestaan objektiivisuuteen pyritään esimerkiksi käyttämällä mallivastauksia arvioinnin tukena.

Laadukkaan ja eettisen arvioinnin ominaisuuksiin kuuluva oikeudenmukaisuus tarkoittaa, että arviointi on johdonmukaista, tasapuolista, läpinäkyvää ja perusteltavissa olevaa (Luostarinen & Ouakrim-Soivio 2019; Ouakrim-Soivio 2015). Oikeudenmukaiset arviointikäytännöt eivät syrji arvioitavia tai aseta heitä eriarvoisiin asemiin, vaan arvioinnissa huomioidaan arvioitavien välinen tasapuolisuus (Atjonen 2007). Kriteeriperustainen arviointi edistää arvioinnin oikeudenmukaisuutta, sillä se on läpinäkyvää ja perustuu opetussuunnitelmaan, minkä lisäksi se tarjoaa kaikille arvioitaville mahdollisuuden onnistua. Ouakrim-Soivion (2015) mukaan oikeudenmukaisuus on yksi arvioinnin olennaisimmista periaatteista, sillä se pitää sisällään myös yhdenvertaisuuden ja tasa-arvon periaatteet.

### 2.1.3 Suhteellinen ja kriteeriperustainen arviointi

Arviointia voidaan toteuttaa suhteellisesti, kriteeriperustaisesti tai näitä menettelytapoja yhdistellen (Ouakrim-Soivio 2015). Suhteellisessa eli normiperustaisessa arvioinnissa arvioitavaa suoritusta verrataan muiden arvioitavien suoritustasoon ja arvosana annetaan suhteessa muihin arvioitaviin (Wallace & Ng 2023). Yleinen normiperustaisen arvioinnin sovellus on sovittaa arvioitavat suoritukset normaalijakaumaan (Ouakrim-Soivio 2015). Näin annetut arvosanat jakautuvat normaalijakauman mukaisesti ryhmän yleisestä suoritustasosta huolimatta, minkä vuoksi ryhmän keskimääräinen osaamistaso vaikuttaa normiperustaisella arvioinnilla annetta-

viin arvosanoihin (Luostarinen & Nieminen 2019; Wallace & Ng 2023). Koska arvioitavat henkilöt asetetaan suhteellisessa arvioinnissa järjestykseen osaamisen perusteella, se mahdollistaa erityisesti yksilöiden välisen vertailun (Lok ym. 2016). Näin se sopii arviointimenetelmäksi esimerkiksi erilaisiin valikointi-, valinta- ja kilpailutilanteisiin (Keurulainen 2013).

Kriteeriperustaisessa arvioinnissa suoritusta ei verrata muihin arvioitaviin suorituksiin vaan ennalta määrättyihin arviointikriteereihin (Sawaki 2016). Ouakrim-Soivio (2015) määrittelee kriteerien olevan tavoitteista johdettuja arvioinnin välineitä, joiden tarkoituksena on kuvata mitattavan osaamisen tasoa. Ryhmän keskimääräinen osaamistaso ei vaikuta kriteeriperustaisella arvioinnilla suoritettavaan arviointiin, vaan opiskelijan arvosana perustuu täysin hänen omaan suoritukseensa (Neil ym. 2010). Kriteeriperustainen arviointi voidaan jakaa edelleen määrälliseen ja laadulliseen kriteeriperustaiseen arviointiin (Ouakrim-Soivio 2015). Määrällisessä kriteeriperustaisessa arvioinnissa suoritusta verrataan johonkin määrälliseen kriteeriin, jolloin arviointipäätös perustuu arviointivälineeseen eikä arvioijan tulkintaan (Keurulainen 2013; Ouakrim-Soivio 2015). Tyypillisiä esimerkkejä tällaisesta arvioinnista ovat esimerkiksi monivalintatehtävät. Laadullisessa kriteeriperustaisessa arvioinnissa puolestaan arvioitavaa osaamista verrataan laadullisiin kriteereihin, ja myös arvioitavat suoritukset ovat pääasiassa laadullisia (Ouakrim-Soivio 2015). Laadullisia kriteerejä käytetään arvioinnin perustana esimerkiksi arviointimatriisin avulla suoritettavassa arvioinnissa, sillä matriisin mittaamaa osaamista arvioidaan laadullisten suoritustasojen kuvausten avulla (Brookhart 2018; Ouakrim-Soivio 2015).

Suhteellisen arvioinnin sekä määrällisen ja laadullisen kriteeriperustaisen arvioinnin periaatteet eroavat toisistaan, minkä lisäksi myös arviointimenetelmien tarkoitukset ja kohteet ovat erilaisia (Biggs & Tang 2011; Keurulainen 2013; Lok ym. 2016). Suhteelliseen arviointiin liitetään esimerkiksi suhteellisuuden, mittaamisen ja määrällisyyden, objektiivisuuden sekä standardoinnin periaatteet (Keurulainen 2013). Suhteellinen arviointi pyrkii asettamaan opiskelijat osaamisen perusteella järjestykseen ja täten se kohdistuu koko arvioitavaan ryhmään (Lok ym. 2016). Määrällisen kriteeriperustaisen arvioinnin periaatteet noudattavat pitkälti suhteellisen arvioinnin periaatteita, mutta suhteellisuuden sijaan siinä korostetaan absoluuttisuutta (Keurulainen 2013). Laadulliseen kriteeriperustaiseen arviointiin liitetään edellisistä arviointimenetelmistä poiketen esimerkiksi tulkinnallisuuden ja kontekstuaalisuuden periaatteet. Kriteeriperustaisen arviointimenetelmien tarkoituksena on kuvata yksilön osaamista ja kehittymistä (Lok ym. 2016).

Usein suhteellista ja kriteeriperustaista arviointia käytetään yhdessä, sillä arviointimenetelmät täydentävät toisiaan ja niiden yhdistäminen voi myös parantaa arvioinnin laatua (Lok ym. 2016). Esimerkiksi ylioppilaskokeiden arviointi perustuu molempiin arviointitapoihin (Ouakrim-Soivio 2015). Kokeiden arviointi suoritetaan yhteisten arvostelukriteerien mukaisesti, jolloin arviointi on kriteeriperustaista, mutta lopulliset arvosanojen pisterajat määräytyvät varsinaisten kokelaiden saamien pisteiden mukaisesti, jolloin arvioinnissa on myös suhteellisen arvioinnin piirteitä (Ouakrim-Soivio 2015; Pisterajat s.a.).

## **2.2 Arviointimatriisi arvioinnin välineenä**

Arviointimatriisi eli rubriikki on kriteeritaulukko, jonka päätarkoituksena on toimia opiskelijoiden osaamisen kriteeriperustaisen arvioinnin apuvälineenä (Brookhart 2013; Dirkx ym. 2019). Brookhartin (2013) mukaan arviointimatriisi koostuu kahdesta pääelementistä: johdonmukaisista kriteereistä eli arvioitavan osaamisen osa-alueista sekä näiden kriteerien erilaisten suoritustasojen kuvauksista. Pophamin (1997) mukaan kriteeristön sekä suoritustasojen kuvausten lisäksi arviointimatriisiin kuuluu sisältää myös suorituksen pisteitysstrategia. Matriisi on siis taulukko, jonka solut sisältävät tyypillisesti vasemmassa sarakkeessa sijaitseville kriteereille konkreettiset laatuluokkien kuvaukset (Aijjawi ym. 2018).

Arviointimatriisien avulla voidaan arvioida tuotoksia, kuten opinnoissa tuotettuja lopputöitä ja koevastauksia, sekä toimintaa, kuten ajattelun taitoja ja esimerkiksi kriittistä ajattelua (Brookhart 2013; Reynders ym. 2020). Niitä käytetään erityisesti monimutkaisten ja käytännönläheisten suoritusten arvioinnissa, eivätkä ne sovellu yksiselitteisesti arvioitavien tehtävien, kuten monivalintakysymysten, arviointiin (Jonsson & Svingby 2007; Popham 1997). Matriisit soveltuvat monipuolisesti erilaisiin arviointi- ja oppimistarkoituksiin, ja erityisen hyvin ne toimivat formatiivisen arvioinnin välineenä niiden avulla annetun laadullisen palautteen sekä arvioinnin läpinäkyvyyden ja palauteprosessin edistämisen vuoksi (Brookhart 2013; Hidayati ym. 2024; Panadero & Jonsson 2013). Ulkopuolisen arvioijan toteuttaman arvioinnin lisäksi arviointimatriisit soveltuvat hyvin myös vertais- ja itsearviointiin, sillä ne voivat muun muassa tukea opiskelijan itsesäätelyä ja minäpystyvyyttä, tarkentaa oppimistavoitteita sekä auttaa opiskelijan oppimisprosessin seuraamista (Brookhart 2013; Hafner & Hafner 2003; Martin-Kniep 2000; Salamanca 2024). Arviointimatriisit voivat tukea myös opettajan työtä, sillä ne voivat esimerkiksi edistää opettajien antamaa ohjeistusta sekä auttaa kohdistamaan opetuksen ja arvioinnin juuri oppimistavoitteisiin (Martin-Kniep 2000).

Arviointimatriisin ensimmäinen pääelementti eli kriteeristö koostuu tekijöistä, joita arvioija tarkastelee arvioitavasta suorituksesta arviointiprosessin aikana (Reddy & Andrade 2010). Kriteerit ovat sen osaamisen osa-alueita, jota tarkasteltavalla suorituksella on tarkoitus mitata, ja ne voivat olla keskenään samanarvoisia tai niitä voidaan painottaa arvioitavan suorituksen pisteityksessä eri tavoin (Brookhart 2013; Popham 1997). Arviointimatriisin validiteetin kannalta kriteerien valinta on tärkeää, sillä laadukkaassa matriisissa kriteerit on valittu tarkoituksenmukaisesti niin, että niiden avulla voidaan arvioida juuri arvioitavan suorituksen indikoimaa osaamista (Moskal & Leydens 2000). Laadukkaassa arviointimatriisissa kriteerit ovat selkeästi määriteltävissä ja havainnoitavissa olevia ominaisuuksia, jotka muodostavat matriisilla arvioitavaa osaamista kattavasti mittaavan kokonaisuuden (Brookhart 2013). Lisäksi kriteerien kuuluu olla sellaisia arvioitavan suorituksen ominaisuuksia, joiden laatu voidaan määrittää sekä hyvin heikolla että hyvin korkealla tasolla: näin laadukkaiden suoritustasojen kuvausten laatiminen on mahdollista.

Suoritusta arvioitaessa kriteerien täyttymistä arvioidaan matriisin toisen pääelementin eli kriteerien suoritustasojen kuvausten avulla (Panadero & Jonsson 2020). Suoritustasojen kuvauksilla määritellään, millainen arvioitava suoritus on tietyn kriteerin osalta jokaisessa matriisin laatuluokassa (Brookhart 2013). Arviointimatriisin toimintaperiaate on, että arvioija valitsee jokaisen kriteerin kohdalla sen suoritustason kuvauksen, joka vastaa arvioitavaa suoritusta parhaiten. Se, että arviointimatriisissa kriteerien toteutumista arvioidaan vertaamalla suoritusta laadullisiin kuvailuihin, erottaa arviointimatriisin perinteisestä arviointitaulukosta, sillä yksinkertaisemmassa arviointitaulukossa kriteerien toteutumista arvioidaan pelkkien arvosanojen tai adjektiivien avulla. Suoritustasojen kuvausten on oltava laadukkaasti määriteltäviä ja niiden kuuluu olla toisistaan selkeästi erotettavissa arviointimatriisin reliabiliteetin säilymiseksi (Moskal & Leydens 2000).

Arviointimatriisit voidaan jakaa analyttisiin ja holistisiin arviointimatriiseihin (Brookhart 2013). Analyttisessä arviointimatriisissa jokainen kriteeri arvioidaan erikseen, kun taas holistisessa arviointimatriisissa on analyttisen matriisin tavoin useita kriteereitä, mutta kriteerien suoritustasojen kuvaukset on yhdistetty ja näin kaikki kriteerit arvioidaan samanaikaisesti. Holistisella arviointimatriisilla koko suoritus arvioidaan siis yhtenä kokonaisuutena ilman, että sen yksittäisiä osia tai ominaisuuksia korostettaisiin arviointiprosessin aikana (Martin-Kniep 2000). Erilaiset arviointimatriisit sopivat erilaisiin tarkoituksiin (Brookhart 2013). Analyttinen arviointimatriisi tuottaa oppijalle hollistista matriisia tarkempaa palautetta, minkä vuoksi se sopii

erityisen hyvin formatiiviseen arviointiin. Holistisen arviointimatriisin avulla puolestaan arviointi on nopeampaa ja se sopii käytettäväksi erityisesti sellaisessa summatiivisessa arvioinnissa, jossa oppija ei saa suorituksestaan muuta palautetta kuin arvosanan.

Arviointimatriisit voidaan jakaa myös yleisiin ja tehtäväkohtaisiin matriiseihin (Panadero & Jonsson 2020). Yleisessä arviointimatriisissa kriteerit ja niiden suoritustasojen kuvaukset on laadittu niin, että matriisilla voidaan arvioida samaa osaamista monista erilaisista suorituksista, kun taas tehtäväkohtaisella arviointimatriisilla on tarkoitus arvioida tiettyä suoritusta ja sen kriteerit vastaavat ainoastaan tietyn suorituksen sisältöä. Tämän vuoksi yleiset arviointimatriisit ovat tehtäväkohtaisia matriiseja monikäyttöisempiä arviointivälineitä (Brookhart 2013). Näiden matriisityyppien välinen merkittävä ero on erityisesti se, että oppijat voivat hyödyntää yleisiä arviointimatriiseja oppimisprosessissaan monella tavalla, kun taas tehtäväkohtaiset matriisit sisältävät suorituksen oikeat ratkaisutavat ja siten niitä ei voida jakaa arvioitaville etukäteen (Brookhart 2013; Martin-Kniep 2000; Popham 1997). Toisaalta myös yleisten arviointimatriisien hyödyntäminen oppimisprosessissa voi olla haastavaa, mikäli niiden kriteerit ja suoritustasojen kuvaukset ovat suorituksesta liian irrallisia (Martin-Kniep 2000). Yleiset arviointimatriisit ovat myös arvioijan kannalta hyödyllisiä, sillä niiden käyttäminen usein kohdentaa arvioinnin koskemaan nimenomaan arvioitavaa osaamista tehtäväkohtaisen suoriutumisen sijaan, minkä lisäksi niitä voidaan käyttää useiden erilaisten suoritusten arvioimisessa helpottaen arvioinnin suunnittelua.

Arviointimatriisien laatua voidaan arvioida tutkimalla esimerkiksi niiden reliabiliteettia ja validiteettia (Brookhart & Chen 2015). Matriisien reliabiliteettia voidaan tarkastella joko arvioinnin suhteellisena tai absoluuttisena yhdenmukaisuutena eri tekijöiden, kuten arvioijien, tilanteiden ja kriteerien, välillä, ja sitä voidaan tutkia erilaisilla laskentamenetelmillä. Esimerkiksi Cohenin kappa -kerroin, sisäkorrelaatiokerroin ja Pearsonin korrelaatiokerroin ovat tyypillisiä reliabiliteettia kuvaavia tilastollisia tunnuslukuja. Myös arviointimatriisien validiteettia voidaan vastaavasti arvioida tilastollisilla menetelmillä, kuten selvittämällä matriisin korrelaatiokerroin muiden arviointivälineiden kanssa (Jonsson & Svingby 2007). Arviointimatriisien reliabiliteettia ja validiteettia voidaan edistää useilla tavoilla (Dawson 2017). Reliabiliteettia voidaan lisätä esimerkiksi kouluttamalla arvioijia matriisien luotettavaan käyttöön sekä täydentämällä matriiseihin esimerkkejä eritasoisista suorituksista (Brookhart 2013; Jonsson & Svingby 2007). Myös matriisityyppi voi vaikuttaa reliabiliteettiin, sillä tehtäväkohtaisissa matriiseissa arvioitsijareliabiliteetti on usein yleisten matriisien reliabiliteettia korkeampi etenkin arviointiproses-

sin alussa, ja analyttisten arviointimatriisien reliabiliteetti on usein holististen matriisien reliabiliteettia korkeampi (Brookhart 2013; Jonsson & Svingby 2007). Matriisien validiteettia puolestaan voidaan pyrkiä lisäämään esimerkiksi hyödyntämällä asiantuntijalausuntoja, jo olemassa olevia matriiseja sekä sidosryhmien palautteita matriisien laadinnassa (Dawson 2017; Jonsson & Svingby 2007).

## **2.3 Ylioppilaskokeiden arviointi**

### **2.3.1 Maantieteen ylioppilaskokeet ja niiden arviointi**

Ylioppilastutkintolautakunnan laatimilla ylioppilaskokeilla pyritään selvittämään, onko ylioppilaskokelas saavuttanut lukion opetussuunnitelman perusteiden mukaisen osaamisen sekä riittävän kypsyyden aineen hallinnassa (Reaaliaineiden... 2024; Valtioneuvoston asetus ylioppilastutkinnosta 612/2019). Kokeiden tehtävät perustuvat lukiokoulutuksen oppimäärään kuuluviin pakollisiin ja valtakunnallisiin valinnaisiin opintoihin (Valtioneuvoston asetus lukiokoulutuksesta 810/2018; Valtioneuvoston asetus ylioppilastutkinnosta 612/2019; Yleiset määräykset ja ohjeet 2024). Kokeet sisältävät myös oppiainerajat ylittäviä tehtäviä, jotka voivat perustua Lukion opetussuunnitelman perusteiden (2019) mukaiseen laaja-alaiseen osaamiseen.

Ylioppilaskoevastausten arviointiprosessi koostuu kahdesta vaiheesta. Koesuoritusten alustavan arvostelun toteuttaa kyseisen aineen opettaja (Laki ylioppilastutkinnosta 502/2019; Yleiset määräykset ja ohjeet 2024). Alustava arvostelu suoritetaan Ylioppilastutkintolautakunnan koepäivänä julkaisemien alustavien hyvän vastauksen piirteiden avulla, ja alustavan arvostelun mukaiset tulokset ilmoitetaan kokelaille. Opettajat, opettajajärjestöt ja muut tahot voivat lähettää lautakunnalle alustaviin hyvän vastauksen piirteisiin liittyviä huomioita tai ehdotuksia niiden julkaisemisen jälkeen. Lopullisen koesuoritusten arvostelun suorittavat Ylioppilastutkintolautakunnan sensorit, jotka pisteittävät kaikki vastaukset ainejaoksissa päätettyjen lopullisten arvostelukriteerien mukaisesti. Lopulliset hyvän vastauksen piirteet julkaistaan tulosten julkaisemisen yhteydessä (Hyvän vastauksen piirteet: FI – Maantiede 2023; Yleiset määräykset ja ohjeet 2024).

Maantieteen ylioppilaskoe järjestetään toisena reaaliaineiden koepäivänä, joiden järjestys voi vaihdella eri tutkintokerroilla (Reaaliaineiden... 2024; Yleiset määräykset ja ohjeet 2024). Lukiokoulutuksen maantieteen oppimäärään kuuluu yksi pakollinen moduuli ja kolme valtakunnallista valinnaista moduulia (Lukion... 2019). Maantieteen oppiaineen tavoitteena on kehittää

opiskelijan maantieteellistä ajattelua, tukea maantieteellisten ilmiöiden ja prosessien ymmärtämistä sekä vahvistaa maantieteellisten taitojen hallintaa ja soveltamista. Maantieteen ylioppilaskoe koostuu kolmesta osiosta (Hyvän vastauksen piirteet: FI – Maantiede 2023). Ensimmäisessä osiossa on yksi pakollinen tehtävä, jonka enimmäispistemäärä on 20. Toisessa osiossa on neljä 20 pisteen tehtävää, joista kokelas vastaa kahteen. Kolmannessa osiossa on neljä 30 pisteen tehtävää, joista vastaavasti vastataan kahteen tehtävään. Kokeen enimmäispistemäärä on siis 120 pistettä. Maantieteen ylioppilaskokeella arvioidaan, miten hyvin kokelas hallitsee itsenäisesti maantieteellisiä tietoja ja taitoja sekä kuinka hyvin kokelas osaa soveltaa niitä erilaisten tehtävien kontekstissa (Hyvän vastauksen piirteet: FI – Maantiede 2023).

Maantieteen koesuoritusten arvioinnissa noudatetaan Ylioppilastutkintolautakunnan antamia reaaliaineiden kokeiden määräyksiä ja ohjeita sekä maantieteen hyvän vastauksen piirteitä (Hyvän vastauksen piirteet: FI – Maantiede 2023; Reaaliaineiden... 2024). Reaaliaineiden kokeiden määräyksissä ja ohjeissa määritellään muun muassa tekijöitä, joihin suoritusten arvioinnissa on kiinnitettävä erityistä huomiota (Reaaliaineiden... 2024). Kypsyttä osoittavia tekijöitä reaaliaineiden koesuorituksissa ovat esimerkiksi monipuoliset tiedonkäsittelytaidot, väitteiden perusteleminen, annettujen työkalujen asianmukainen käyttö sekä tehtäviin liittyvien aineistojen tarkoituksenmukainen käyttö. Suorituksen arvoa alentavia tekijöitä puolestaan ovat esimerkiksi asiavirheet, vastauksen rakentuminen mielipiteiden varaan, tehtävänannon käsittäminen väärin tai tehtävänannossa annetun merkkimäärän ylittäminen. Maantieteen hyvän vastauksen piirteet eli koesuorituksen arvosteluohjeet puolestaan koostuvat eri suorituskertojen kokeille yhteisestä yleisestä osasta sekä tehtäväkohtaisista pisteitysohjeista (Hyvän vastauksen piirteet: FI – Maantiede 2023). Yleisessä osassa kuvataan esimerkiksi vastausten arvioinnin lähtökohtia sekä kokeen tehtäviä yleisellä tasolla. Tehtäväkohtaisissa pisteitysohjeissa kerrotaan ensin yleisemmällä tasolla tehtävän olennaisimmista arvioinnin kohteista, minkä jälkeen vastauksen sisältöjen pisteittäminen ohjeistetaan tarkemmin.

### 2.3.2 Syksyn 2023 maantieteen ylioppilaskokeen tehtävän 7 pisteitysohjeet

Syksyn 2023 ylioppilaskokeen tehtävä 7 kuuluu kokeen kolmanteen osioon (Hyvän vastauksen piirteet – Maantiede 2023). Tehtävän otsikko on Kuusamon alueen ratahanke, ja siinä tarkastellaan liikennemuotoja ja niiden vaikutuksia yleisesti sekä Kuusamon alueen kontekstissa. Tehtävä on jaettu kolmeen osatehtävään, joiden yhteispistemäärä on 30, ja joista jokaiseen kuuluu vastata esseevastauksella (taulukko 1). Tehtävässä on käytettävänä viisi kartta-aineistoa. Tehtävän yleiskuvauksen mukaan siinä arvioidaan vastaajan perustietoja liikennemuodoista,

minkä lisäksi tehtävässä mitataan myös monipuolista yleis- ja teemakarttojen lukutaitoa, kykyä tarkastella tehtävässä esitellyn infrastruktuurihankkeen vaikutuksia monipuolisesti sekä maantieteellistä kirjoitustaitoa. Tehtävässä hyödynnetään useita maantieteen opintojen moduuleja, ja siinä vaaditaan vastaajalta monitieteistä ja laaja-alaista ympäristöön liittyvää osaamista (Hyvän vastauksen piirteet – Maantiede 2023; Lukion opetussuunnitelman... 2019).

Taulukko 1. Syksyn 2023 maantieteen ylioppilaskokeen tehtävän 7 ”Kuusamon alueen ratahanke” tehtävänanto ja kokonaispisteiden jakautuminen osatehtävittäin.

Osa-tehtävä	Tehtävänanto	Pistemäärä
7.1	Kuvaile raideliikenteen etuja ja haittoja verrattuna tieliikenteeseen.	6
7.2	Rautatieyhteyttä Kuusamoon on suunniteltu jo 1800-luvun lopulta lähtien. Vuosikymmenien aikana hanke on saanut sekä kannatusta että vastustusta ja radasta on tehty erilaisia kannattavuuslaskelmia. Uuden radan rakentamiskustannukset ovat eräiden arvioiden mukaan noin viisi miljoonaa euroa kilometriä kohden. Valitse kartan 7.A reittivaihtoehdoista A–C mielestäsi järkevin. Arvioi valitsemasi reittivaihtoehdon etuja ja haittoja karttojen 7.A–7.E perusteella vertaamalla sitä kahteen muuhun vaihtoehtoon.	12
7.3	Pohdi uuden ratayhteyden vaikutuksia ihmisen toimintaan ja luontoon Kuusamon alueella pitkällä aikavälillä.	12

Osatehtävässä 7.1 vastauksesta voi saada kaksi pistettä hyvin perustellusta edusta tai haitasta (Hyvän vastauksen piirteet – Maantiede 2023). Mainittu etu tai haitta voi saada vain yhden pisteen, jos se on kuvattu pintapuolisesti tai ilman vertailua. Vastauksessa täytyy käsitellä sekä etuja että haittoja, jotta siitä voi saada täydet pisteet. Lisäksi osatehtävästä voi saada vain puolet saavutetusta pistemäärästä, jos vastauksessa ei ole vertailuasetelmaa. Hyvän vastauksen piirteissä on kuvattu esimerkkisisältöjä, joista vastaukselle voidaan antaa pisteitä (liite 1).

Osatehtävässä 7.2 vastaaja voi saada kaksi pistettä vastauksen kypsyystyypistä ja loogisesti jäsenneilystä rakenteesta (Hyvän vastauksen piirteet – Maantiede 2023). Vastauksesta voi saada yhden pisteen valitun reittivaihtoehdon nimeämisestä, ja enintään yhdeksän pistettä vastauksesta voi saada reittivaihtoehdon perusteluista. Hyvin kuvattua ja muihin reittivaihtoehtoihin vertaavasta perustelusta voi saada kaksi pistettä ja pintapuolisesta perustelusta yhden pisteen. Vastauksessa kuuluu käyttää vähintään kolmea kartta-aineistoa ja kuvata sekä valitun reittivaihtoehdon etuja että haittoja, jotta osatehtävästä voi saada täydet pisteet. Jos vertailuasetelma puuttuu vastauksesta, voi osatehtävästä saada vain puolet vastaukselle annettusta pistemäärästä. Pisteytysohjeissa on kuvattu esimerkkisisältöjä, joista vastaukselle voidaan antaa pisteitä (liite 1).

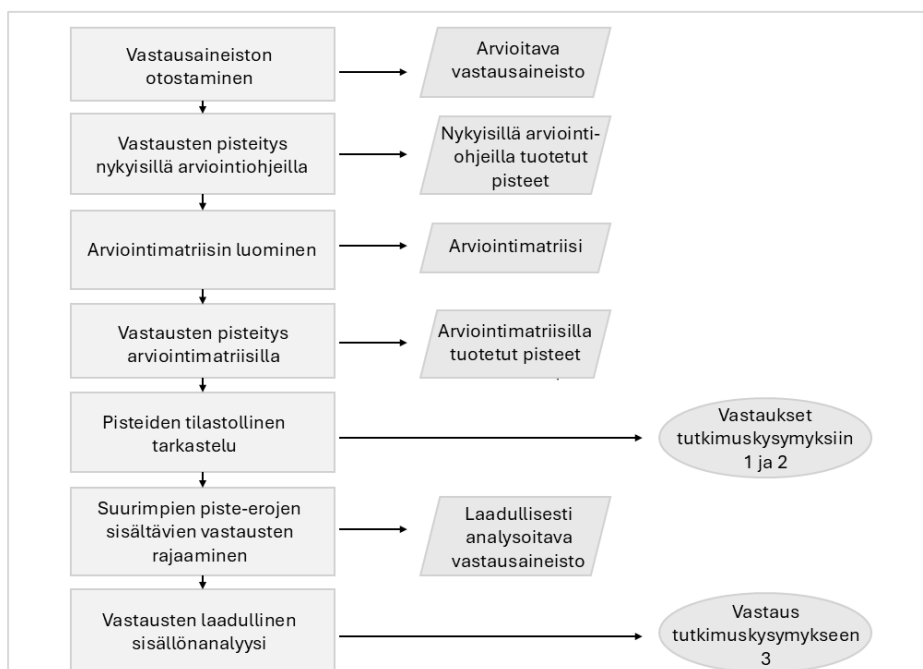
Osatehtävässä 7.3 vastaaja voi edellisen osatehtävän tavoin saada kaksi pistettä kypsästä ja loogisesti jäsenneilystä vastauksesta (Hyvän vastauksen piirteet – Maantiede 2023). Osatehtävästä

voi saada enintään kymmenen pistettä vastauksessa mainituista ratayhteyden vaikutuksista. Hyvin kuvatusta huomiosta voi saada kaksi pistettä ja pintapuolisesti kuvatusta huomiosta yhden pisteen. Vastauksessa kuuluu käsitellä vaikutuksia sekä ihmisen toimintaan että luontoon, jotta osatehtävästä voi saada täydet pisteet: vaikutuksista ihmisen toimintaan voi saada enintään 4–6 pistettä ja vaikutuksista luontoon enintään 4–6 pistettä. Tehtävän pisteitysohjeissa on nimetty esimerkkisisältöjä, joista vastaukselle voidaan antaa pisteitä (liite 1).

### 3 Aineisto ja menetelmät

#### 3.1 Tutkimuksen vaiheet

Tutkimus toteutettiin käyttämällä kvantitatiivisia ja kvalitatiivisia tutkimusmenetelmiä (kuva 1). Tutkimuksessa tarkasteltava vastausaineisto (n=70) pisteitettiin kahdella eri arviointimenetelmällä eli kokeen nykyisillä arviointiohjeilla sekä tutkimuksessa luodulla arviointimatriisilla. Tuotettuja pisteityksiä tarkasteltiin tilastollisilla menetelmillä, joiden avulla saatiin vastaukset ensimmäiseen ja toiseen tutkimuskysymykseen. Vastaukset, joiden pisteissä erot tarkasteltujen arviointimenetelmien välillä olivat suurimmat, analysoitiin laadullisella sisällönanalyysillä, jolla saatiin vastaus kolmanteen tutkimuskysymykseen.



Kuva 1. Tutkimuksen vaiheet.

#### 3.2 Aineisto

Tutkimuksessa tarkasteltavana vastausaineistona käytettiin syksyn 2023 maantieteen ylioppilaskokeen tehtävän 7 ”Kuusamon alueen ratahanke” vastauksia. Kokeen kaikkien tehtävien vastauksista koostuva vastausaineisto saatiin Ylioppilastutkintolautakunnalta tutkimusluvalla OPH-6154-2023. Tehtävän 7 vastausaineisto koostui 1388 vastauksesta. Näistä vastauksista luotiin 70 vastauksen otos, jota käytettiin tutkimuksessa arvioitavana vastausaineistona. Lisäksi

tutkimuksessa käytettiin vastausten alkuperäisessä Ylioppilastutkintolautakunnan suorittamassa arvioinnissa saamia pisteitä, joita nimitetään tässä tutkimuksessa vastauksien alkuperäiseksi sensoripisteiksi.

Otos poimittiin ositetulla otannalla eli otantamenetelmällä, jonka avulla pyritään varmistamaan otoksen edustavuus kaikkien niiden ryhmien osalta, jotka ovat tutkimuksen kannalta merkittäviä (Otos... 2003). Ositetussa otannassa otoksen tarkkuutta ja edustavuutta parannetaan jakamalla perusjoukko sitä koskevien tietojen avulla ositteisiin ennen otoksen luomista (Ositettu... s.a.). Se sopii otantamenetelmäksi silloin, kun jokin ryhmä on niin pieni, että yksinkertaisella satunnaisotannalla ei voida taata kaikkien ryhmien edustavuutta otoksessa (Otos... 2003). Kun tehtävän 7 vastaukset jaettiin luokkiin kokonaispistemäärän mukaan viiden pisteen välein, pisteluokat 0–5 ja 26–30 olivat huomattavasti muita pisteluokkia pienempiä (taulukko 2), ja koska tutkimuksessa haluttiin sen kartoittavan luonteen vuoksi tarkastella kattavasti eritasoisia vastauksia, ositettu otanta varmisti kaikkien pisteluokkien edustavuuden otoksessa ja näin sopi hyvin tutkimuksen otantamenetelmäksi. Otanta tehtiin jakamalla vastausaineisto kuuteen pisteluokkaan viiden pisteen välein, ja jokaisesta pisteluokasta poimittiin suhteellista otosta lähestyvä määrä vastauksia tarkasteluun (taulukko 2). Tutkimuksen otos lähestyy koko aineiston pisteluokkien suhteellista frekvenssijakaumaa, mutta se ei noudata sitä täysin, jotta tarkasteltavaan otokseen saatiin tarpeeksi vastauksia myös ensimmäisestä ja viimeisestä pisteluokasta.

Taulukko 2. Vastausten pistejakauma tutkimuksen otoksessa ja koko vastausaineistossa.

Piste-luokka	Määrä vastausaineistossa	Osuus vastausaineistossa	Määrä otoksessa	Osuus otoksessa
0–5	20	1,4 %	4	5,7 %
6–10	214	15,4 %	11	15,7 %
11–15	446	32,1 %	20	28,6 %
16–20	406	29,3 %	18	25,7 %
21–25	226	16,3 %	11	15,7 %
26–30	76	5,5 %	6	8,6 %
yhhteensä	1388	100,0 %	70	100,0 %

Koko kokeen vastausaineisto saatiin CSV-tiedostona, jota käsiteltiin Microsoft Excel -taulukkolaskentaohjelmassa. Tehtävän 7 vastaukset poimittiin tästä tiedostosta erilliseen tiedostoon, jossa ne luokiteltiin pistemäärien mukaisesti ja josta ne siirrettiin edelleen erillisiin pisteluokkien mukaisiin laskentataulukoihin. Tutkimuksessa käytetyt vastaukset saatiin jokaisesta pisteluokasta Excelissä suoritettulla satunnaisotannalla.

### 3.3 Menetelmät

#### 3.3.1 Aineiston arviointi

Tutkimuksessa toteutettu vastausten arviointi suoritettiin yhteistyössä kahden kokeneen maantieteen sensorin kanssa. Tutkielman tekijään viitataan tästä eteenpäin nimellä arvioija 1 ja tutkielmaan osallistuneisiin maantieteen sensoreihin nimillä arvioija 2 ja arvioija 3.

Tutkimuksessa tarkasteltava vastausaineisto pisteitettiin kahdessa eri vaiheessa. Molemmat arviointivaiheet toteutettiin niin, että vastausten saamat alkuperäiset pisteet eivät olleet arvioijien tiedossa. Ensimmäinen arviointivaihe toteutettiin ennen arviointimatriisin luomista kokeen nykyisillä pisteitysohjeilla eli Ylioppilastutkintolautakunnan laatimien lopullisten hyvän vastauksen piirteiden avulla (Hyvän vastauksen piirteet: FI – Maantiede 2023). Tässä arviointivaiheessa arvioija 1 arvioi 10 vastausta ja arvioijat 2 ja 3 arvioivat kumpikin 40 vastausta (kuva 2). Jokainen arvioija pisteitti vastauksia kaikista pisteluokista, ja otokseen kuului 10 vastausta, jotka kaikki arvioijat pisteittivät.

Pisteluokka	Vastausten arvioijat																			
1-5 p	2	3	3	123																
6-10 p	2	2	2	2	2	3	3	3	3	123	123									
11-15 p	2	2	2	2	2	2	2	2	2	3	3	3	3	3	3	3	3	3	123	123
16-20 p	2	2	2	2	2	2	2	2	3	3	3	3	3	3	3	3	123	123		
21-25 p	2	2	2	2	3	3	3	3	3	123	123									
26-30 p	2	2	2	3	3	123														

Kuva 2. Vastausten jakautuminen arvioijien kesken ensimmäisessä arviointivaiheessa eli nykyisillä arviointiohjeilla suoritettussa arvioinnissa. Vihreä ruutu kuvaa arvioijan 2 ja sininen ruutu kuvaa arvioijan 3 pisteittämää vastausta. Vaaleanpunainen ruutu kuvaa vastausta, jonka kaikki kolme tutkimuksen arvioijaa pisteittivät.

Arvioijat kirjasivat pisteet strukturoituihin Excel-taulukoihin, joihin merkittiin asiat, joista vastaus sai hyvän vastauksen piirteiden mukaisesti pisteitä. Taulukoihin kirjattiin myös huomioita, joita arvioitavista vastauksista tehtiin arvioinnin aikana. Tämän arviointivaiheen jälkeen jokainen arvioija laati esityksen arvioinnin aikana tekemistään havainnoista sekä tunnistamistaan arvioinnin haasteista, kuten siitä, millaisten vastausten ja taitojen arviointi on nykyisillä arviointiohjeilla haastavaa.

Toinen arviointivaihe toteutettiin tutkimuksessa luodun arviointimatriisin avulla (ks. alaluku 3.3.2). Jokainen arvioija pisteitti arviointimatriisilla eri vastaukset kuin ensimmäisessä arviointivaiheessa, jotta edellisen arvioinnin aikana tehdyt tulkinnat vaikuttaisivat pisteitykseen mahdollisimman vähän. Arvioija 1 arvioi tässä arviointivaiheessa 60 vastausta ja arvioijat 2 ja 3

arvioivat kumpikin 30 vastausta (kuva 3). Arviointi toteutettiin niin, että jokaisella vastauksella oli kaksi arvioijaa. Arvioijat pisteittivät aineiston strukturoituihin Excel-laskentataulukoihin, joihin merkittiin vastausten eri kriteereistä saamat prosenttiosuudet sekä mahdolliset vastauksista tehdyt huomiot. Arvioijat eivät keskustelleet antamistaan pisteistä tai matriisin käytöstä arvioinnin aikana, vaan arviointivaihe suoritettiin itsenäisesti.

Pisteluookka	Vastausten arvioijat																	
1-5 p	13	12	12															
6-10 p	13	13	13	13	13	12	12	12	12									
11-15 p	13	13	13	13	13	13	13	13	13	12	12	12	12	12	12	12	12	
16-20 p	13	13	13	13	13	13	13	13	12	12	12	12	12	12	12	12		
21-25 p	13	13	13	13	12	12	12	12										
26-30 p	13	13	13	12	12													

Kuva 3. Vastausten jakautuminen arvioijien kesken toisessa arviointivaiheessa eli arviointimatriisin avulla suoritettussa arvioinnissa. Sininen ruutu kuvaa arvioijien 1 ja 3 pisteittämää vastausta ja vihreä ruutu kuvaa arvioijien 1 ja 2 pisteittämää vastausta. Valkoiset ruudut kuvaavat otokseen kuuluvia vastauksia, joita ei tässä arviointivaiheessa pisteitetty.

### 3.3.2 Arviointimatriisin luominen

Tutkimuksessa käytetty arviointimatriisi luotiin tämän tutkielman tekijän sekä kahden kokeneen maantieteen sensorin yhteistyönä. Arviointimatriisi luotiin tutkimuksen ensimmäisen arviointivaiheen eli nykyisillä arviointiohjeilla suoritettuna pisteityksen jälkeen. Arviointimatriisia suunniteltiin matriisin tekijöiden yhteisissä tapaamisissa sekä itsenäisesti työskennellen. Suunnittelu aloitettiin arviointimatriisin tekijöiden yhteisessä tapaamisessa, jossa sovittiin yhteiset lähtökohdat arviointimatriisin suunnittelulle, eli mitä asioita matriisilla on tarkoitus arvioida ja mihin aineistoihin sen kriteerit perustuvat. Jokainen matriisin tekoon osallistunut henkilö työsti tapaamisen jälkeen itsenäisesti arviointimatriisin, joka jaettiin muille tekijöille. Matriisien toimivuutta testattiin arvioimalla niiden avulla tutkimuksen vastausaineistoon kuulumattomia vastauksia. Näiden matriisien ja niiden toimivuuteen liittyvien kommenttien avulla seuraavassa tekijöiden tapaamisessa keskusteltiin matriisien kehityskohdista, ja tapaamisen jälkeen matriisin tekijät kehittivät matriiseja keskustelun mukaisesti. Näistä itsenäisesti muokatuista matriiseista koostettiin lopullinen arviointimatriisi, jota tutkimuksessa käytettiin.

Nitkon ja Brookhartin (2011) mukaan kaksi keskeisintä arviointimatriisin luomistapaa ovat ylhäältäpäin ohjattu lähestymistapa (engl. *top-down*) sekä alhaaltapäin ohjattu lähestymistapa (engl. *bottom-up*). Ylhäältäpäin ohjatussa eli deduktiivisessa lähestymistavassa arviointimatriisi luodaan arvioitavaan suoritukseen liittyvän käsitteellisen viitekehyksen avulla (Brookhart

2013). Tämä lähestymistapa sopii matriisin tekoon erityisesti sellaisissa tilanteissa, joissa tavoiteltu ja arvioitava osaaminen on kuvattu etukäteen selkeästi esimerkiksi opetussuunnitelmassa. Alhaaltapäin ohjatussa eli induktiivisessa lähestymistavassa puolestaan arviointimatriisi luodaan niiden suoritusten avulla, joissa esitettyä osaamista luodulla matriisilla on tarkoitus arvioida. Brookhart (2013) kuvaa, että alhaaltapäin ohjatussa lähestymistavassa matriisin luominen aloitetaan luokittelemalla arvioitavat suoritukset laatunsa perusteella eri luokkiin. Seuraavaksi jokaiselle suoritukselle luodaan kuvaus siitä, miksi se sijoittuu kyseiseen luokkaan. Suoritusten kuvauksista johdetaan suorituksen laatua määrittäviä ominaisuuksia ja näistä ominaisuuksista muodostetaan arviointimatriisiin kriteerit, joille sitten luodaan osaamista määrittävät suoritustasojen kuvaukset.

Tässä tutkimuksessa arviointimatriisin kriteerien luomisessa mukailtiin sekä ylhäältäpäin ohjattua että alhaaltapäin ohjattua lähestymistapaa (Brookhart 2013). Deduktiivista lähestymistapaa mukailen matriisin kriteerien luomisessa käytettiin pääasiassa kolmea eri aineistoa. Maantieteen ylioppilaskokeiden hyvän vastauksen piirteiden yleisestä osasta käytettiin erityisesti Vastausten arvioinnin lähtökohtia -osiota, jossa määritellään pisteityksen kannalta keskeisiä vastauksen ominaisuuksia (Hyvän vastauksen piirteet: FI – Maantiede 2023). Lukion opetussuunnitelman perusteista (2019) hyödynnettiin maantieteen osiossa kerrottuja maantieteen opetuksen yleisiä tavoitteita sekä osiota, joka käsittelee maantieteen arviointia lukiossa. Reaaliaineiden kokeiden määräyksistä ja ohjeista (2022) huomioitiin erityisesti kokeiden arvostelua käsittelevä osio, jossa määritellään vastauksen kypsyyttä osoittavia tekijöitä. Mainittujen aineistojen lisäksi kriteerien luomisessa hyödynnettiin myös ensimmäisessä arviointivaiheessa tehtyjä havaintoja vastausten arvioinnista ja siihen liittyvistä haasteista (ks. luku 3.3.1). Näin kriteerien luomisessa hyödynnettiin siis myös arvioitavia suorituksia, jolloin kriteerien luomisessa mukailtiin myös Brookhartin (2013) kuvaamaa induktiivista lähestymistapaa.

Myös kriteerien suoritustasojen kuvauksien luomisessa sovellettiin sekä ylhäältä- että alhaaltapäin ohjattua arviointimatriisin luomistapaa (Brookhart 2013). Teoriaperustana kuvauksien luomisessa hyödynnettiin SOLO-taksonomiaa eli oppimisen arviointiin käytettyä viisiportaista mallia (Biggs & Collis 1982). SOLO-taksonomian mukaisesti matriisiin tehtiin viisiportainen vastausten luokitteluasteikko, minkä lisäksi taksonomiaa hyödynnettiin osittain myös kriteerien suoritustasojen kuvausten laatimisessa. Suoritustasojen laatimisessa hyödynnettiin myös muiden reaaliaineiden, kuten filosofian ja psykologian, ylioppilaskokeiden arvioinnissa käytettäviä kriteeritaulukoita ja niissä esitettyjä kuvauksia eri tasoista suorituksista (Hyvän vastauksen

piirteet: FI – Filosofia 2024; Hyvän vastauksen piirteet: FI – Psykologia 2024). Matriisin kriteerien tavoin myös suoritustasojen kuvauksien laadinnassa käytettiin ensimmäisessä arviointivaiheessa tehtyjä havaintoja, jotka liittyivät vastausten arviointiin (ks. alaluku 3.3.1).

Tutkimuksessa luotiin yleinen arviointimatriisi tehtäväkohtaisen arviointimatriisin sijaan, eli luodulla arviointimatriisilla on tarkoitus pystyä arvioimaan kaikkien maantieteen ylioppilaskoetehtävien vastauksia (ks. Panadero & Jonsson 2020). Vaikka siis tutkimuksessa arvioitu vastausaineisto koostui vain yhden ylioppilaskoetehtävän vastauksista, arviointimatriisin luomisessa ei käytetty aineistona kyseisen tehtävän tehtävänantoa tai tehtäväkohtaisia pisteitysohjeita, jotta valmiilla arviointimatriisilla pystyttäisiin arvioimaan mahdollisimman erilaisten tehtävien ja tehtävätyyppien vastauksia.

Tutkimuksessa luodun arviointimatriisin tekijät kehittivät myös arviointimatriisin pisteitysstrategian, eli samanlaista pisteitysstrategiaa ei käytetä esimerkiksi muiden oppiaineiden ylioppilaskokeiden arvioinnissa. Pisteitysstrategia perustuu arviointimatriisin soluihin kiinnitettyihin prosenttiosuuksiin. Arviointimatriisin kaikki viisi arviointikriteeriä on jaettu viiteen suoritustason kuvaukseen. Suoritustasojen kuvauksiin on kiinnitetty heikoimmasta suoritustasosta parhaimpaan prosenttiosuudet 0, 5, 10, 15 ja 20. Vastaukselle annetaan siis jokaisesta kriteeristä tietty prosenttiosuus vastausta kuvaavan suoritustason mukaan, ja vastaukselle on mahdollista antaa tietyistä kriteereistä enintään 20 %. Vastaukselle annettava pistemäärä muodostuu, kun kaikista kriteereistä vastaukselle annetut prosenttiosuudet lasketaan yhteen ja saadulla prosenttiosuudella kerrotaan tehtävän enimmäispistemäärä. Ylioppilaskokeille ominaiseen tapaan arviointimatriisilla voi antaa vain kokonaispisteitä, joten pisteiden pyöristäminen tapahtuu pyöristyssääntöjen mukaisesti.

Luotu arviointimatriisi muodostuu viidestä kriteeristä, näille kriteereille muodostetuista suoritustasojen kuvauksista sekä arvioitavan suorituksen pisteitykseen liittyvistä prosenttiosuuksista (taulukko 3). Arviointimatriisi edustaa analyyttistä matriisityyppiä, eli jokainen kriteeri arvioidaan vastauksen arvioinnin aikana erikseen (Brookhart 2013).

Taulukko 3. Tutkimuksessa luotu arviointimatriisi.

	0 %	5 %	10 %	15 %	20 %
<b>Tehtävänannon noudattaminen:</b> sisältöjen osuvuus, tehtävänannon mukainen tarkastelutapa, vastauksen esitystapa	Vastaus ei vastaa tehtävänantoon; tehtävä on ymmärretty ratkaisevasti väärin. Tehtävänannon mukaista tarkastelutapaa tai esitystapaa ei noudateta.	Vastaus on osin tehtävänannon mukainen, mutta se on heikosti rajattu, epäselvä tai harhaileva. Vastaus- tai tarkastelutapa ei ole täysin tehtävänannon mukainen tai vastaus on tehtävän laajuuteen nähden hyvin suppea.	Vastauksen sisältö sekä tarkastelu- ja vastaustapa ovat pääosin tehtävänannon mukaisia. Vastauksessa voi olla pienissä määrin epäolennaisuuksia tehtävänannon kannalta.	Vastauksessa ilmenee selkeä ymmärrys tehtävänannosta ja sen rajauksesta, ja pääosa vastauksen sisällöistä on olennaisia.	Vastauksessa ilmenee erinomainen ymmärrys tehtävänannon rajauksesta ja se käsittelee kattavasti ja harkitusti olennaiset asiat. Vastauksessa ei ole tehtävänannon kannalta epäoleellisia asioita.
<b>Sisällöt:</b> huomioiden määrä, oikeellisuus ja perustelut	Vastaus sisältää pääasiassa vain virheellistä sisältöä.	Vastauksessa on yksittäisiä oikein kuvattuja sisältöjä, joista osalle saatetaan esittää niukat perustelut. Useita huomioita sisältävässä vastauksessa voi olla muutamia asiavirheitä.	Vastauksessa on muutamia oikein kuvattuja sisältöjä perusteluineen tai useita hatarasti perusteltuja sisältöjä. Vastaus voi sisältää yhden asiavirheen tai jonkin verran epätasallisuutta.	Vastauksessa on muutamia syvällisesti ja virheettömästi käsiteltyjä ja perusteltuja sisältöjä tai useita virheettömiä sisältöjä, joiden perustelut ovat oikein mutta pinnallisempia. Vastauksessa voi olla lievää epätasallisuutta mutta ei varsinaisia virheitä.	Vastauksessa on kattavasti tehtävänannon mukaisia sisältöjä, joista kaikki on virheettömästi, osuvasti ja tarkoituksenmukaisesti perusteltu. Vastauksessa ei ole asiavirheitä eikä epätasallisuutta.
<b>Maantieteellinen tarkastelu- ja jäsentelytapa:</b> maantieteen näkökulmien hallinta, vastauksen looginen eteneminen ja rakenne	Vastauksessa ei ole olennaisia näkökulmia. Rakenne on hajanainen ja sekava. Vastaus voi olla tehtävänantoon nähden hyvin suppea.	Vastaus sisältää joitain olennaisia näkökulmia, mutta se on poukkoileva ja epätasapainoinen. Vastauksen rakenne on puutteellinen ja eteneminen epäloogista.	Vastaus on jäsennelly päänäkökulmien mukaisesti. Rakenne on pääosin selkeä, mutta vastaus ei välttämättä etene loogisesti tai siinä voi olla muita pieniä rakenteellisia puutteita.	Vastaus on jäsennelly tehtävänantoon kuuluvien monipuolisten näkökulmien mukaisesti. Näkökulmia on liitetty toisiinsa jonkin verran. Rakenne on toimiva ja vastaus etenee loogisesti.	Vastaus on jäsennelly johdonmukaisesti niin, että siinä käsitellään monipuolisia näkökulmia sekä niiden välisiä vuorovaikutussuhteita. Vastauksessa otetaan tarvittaessa huomioon myös vaihtoehtoiset lähestymistavat. Vastauksen rakenne on eheä ja tasapainoinen.
<b>Käsitteet:</b> täsmällinen käsitteenkäyttö sekä havaintojen kytkeminen maantieteellisiin ilmiöihin tai prosesseihin tai prosesseihin käsitteellistämällä	Vastauksessa ei käytetä maantieteellisiä käsitteitä; vastaus on arkikielinen. Vastausta ei kytketä lainkaan maantieteellisiin ilmiöihin tai prosesseihin.	Käsitteiden käytössä on selviä puutteita tai virheitä. Vastauksessa saatetaan mainita jokin maantieteellinen ilmiö tai prosessi.	Vastauksessa käytetään maantieteellisiä käsitteitä, mutta käsitteenkäytössä voi olla lieviä puutteita tai yksittäinen virhe. Vastauksessa viitataan pintapuolisesti johonkin maantieteelliseen ilmiöön tai prosessiin.	Vastauksessa käytetään keskeisiä aihepiirin käsitteitä, eikä varsinaisia virheitä ole. Vastauksessa hyödynnetään jotain maantieteellistä teoriaa, mutta vastauksen kytkeminen maantieteellisiin ilmiöihin tai prosesseihin jää kevyeksi.	Käsitteitä käytetään monipuolisesti, täsmällisesti ja huolellisesti. Vastauksessa käsiteltävät asiat on kytketty hyvin maantieteellisiin ilmiöihin ja prosesseihin.
<b>Viestinnän keinot:</b> suomenkielinen asiateksti, paikannimistö, visualisointi, huolellisuus	Vastauksen ilmaisu on vaikeasti ymmärrettävää. Paikannimet on kirjoitettu kokonaan tai pääosin väärin.	Vastaus on osittain puhekielinen, teksti on kömpelöä tai se sisältää paljon kielioppivirheitä. Paikannimistössä on paikoin asia- tai kirjoitusvirheitä. Vastaus voi olla tehtävänantoon nähden hyvin suppea tai viimeistelemätön.	Vastaus on pääosin selkeää asiatekstiä, mutta siinä voi olla paikoitellen joitain kirjoitusvirheitä. Paikannimistössä voi olla pieniä kirjoitusvirheitä. Vastaus voi olla viimeistelemätön tai sisältää huolimattomuusvirheitä.	Vastaus on selkeää ja lähes virheetöntä asiatekstiä. Paikannimet on kirjoitettu pääosin oikein.	Vastaus on kirjoitusasun ja paikannimistön osalta täysin virheetöntä. Vastaus on erinomaista asiatekstiä ja se on esitystavaltaan huoliteltu.

### 3.3.3 Tilastollinen tarkastelu

Vastaukset ensimmäiseen ja toiseen tutkimuskysymykseen selvitettiin kvantitatiivisilla eli määrällisillä tutkimusmenetelmillä, joilla siis tutkittiin, miten eri arviointimenetelmät eroavat toisistaan tuotettujen pisteiden suhteen sekä miten yhdenmukaisesti arviointimatriisilla voidaan pisteittää maantieteen ylioppilaskoevastauksia. Aineiston tilastollinen tarkastelu suoritettiin IBM SPSS Statistics -tilastolaskentaohjelmalla (versio 29.0.0.). Tuloksia havainnollistavat kuvat tuotettiin Microsoft Excel -ohjelmalla sekä IBM SPSS Statistics -ohjelmalla.

Ensimmäiseen tutkimuskysymykseen selvitettiin vastaus kuvailevan tilastoanalyysin (engl. *descriptive statistics*) avulla. Kuvailevan tilastoanalyysin tavoitteena on kuvailla ja tiivistää määrällisen muuttujan jakaumaa tekemättä kuitenkaan yleistyksiä perusjoukkoon (Tilastollinen päättely s.a.). Kuvailevalla tilastoanalyysillä tutkittiin pääasiassa niiden vastausten pisteitä, jotka arvioija 1 pisteitti arviointimatriisilla (n=60). Näistä vastauksista tarkasteltiin koko tehtävän tasolla sekä osatehtäväkohtaisesti niiden saamia alkuperäisiä sensoripisteitä eli nykyisillä arviointiohjeilla annettuja pisteitä ja arvioijan 1 antamia matriisipisteitä. Analyysissa selvitettiin ja vertailtiin esimerkiksi aineiston keski- ja hajontalukuja. Tilastollisten tunnuslukujen lisäksi eri arviointimenetelmien välisiä piste-eroja kuvattiin laatikko-jana-kuvioiden avulla.

Kuvailevan tilastoanalyysin lisäksi tutkimuksessa selvitettiin eri arviointimenetelmillä saatujen pisteiden välisten erojen tilastollinen merkitsevyys. Pisteiden normaalijakautuneisuudet testattiin ensin Shapiro-Wilkin testillä ryhmien välisten erojen tilastollista merkitsevyyttä testaavan testin valintaa varten (Hypoteesien testaus... 2014). Shapiro-Wilkin testin mukaan pisteaineistot noudattivat normaalijakaumaa, joten pisteiden välisten erojen tilastollinen merkitsevyys testattiin käyttämällä parametrissa riippuvien ryhmien t-testiä eli parittaista t-testiä. Tämän tutkimuksen tilastollisten testien riskitasona käytettiin tieteellisessä tutkimuksessa yleisesti hyväksyttyä riskitasoa 0,05 (Vilka 2007).

Toiseen tutkimuskysymykseen selvitettiin vastaus tilastollisten testien ja havainnollistavien kuvaajien avulla, eli arviointimatriisilla toteutetun arvioinnin yhdenmukaisuuden selvittämiseksi tarkasteltiin arvioijien antamien rinnakkaisten pisteiden välistä yhteyttä. Tutkimuksessa laskettiin myös nykyisillä arviointiohjeilla toteutetun arvioinnin yhdenmukaisuutta kuvaavia tilastollisia tunnuslukuja, jotta menetelmien arvioitsijareliabiliteettien keskinäinen vertailu olisi mahdollista. Arviointimatriisilla toteutetun arvioinnin yhdenmukaisuuden tarkastelu aloitettiin muodostamalla arviointimatriisilla tuotetuista pisteistä hajontakaaviot sekä arvioijien 1 ja 2 että arvioijien 1 ja 3 tuottamien pisteiden välille kuvaamaan pisteiden välistä riippuvuutta (Taanila

2022). Hajontakaavioon luotiin myös regressiosuora sekä regressiosuoran yhtälö, joiden avulla voitiin kuvata pisteiden välisen yhteyden suuntaa ja voimakkuutta (Kaakinen & Ellonen s.a.).

Reliabiliteettia voidaan kuvata useilla tilastollisilla tunnusluvuilla (Brookhart & Chen 2017; Stemler 2004), ja tässä tutkimuksessa arviointimatriisilla toteutetun arvioinnin arvioitsijareliabiliteettia tarkasteltiin kolmen tilastollisen testin avulla. Ensin arvioijien antamien pisteiden välille laskettiin sisäkorrelaatiokerroin eli ICC-kerroin (engl. *intraclass correlation coefficient*), joka on tyypillinen arvioitsijareliabiliteettia kuvaava mittari (Graham ym. 2012). Se kuvaa, kuinka suuri osuus varianssista voidaan selittää ylemmän tason tekijöillä, eli tässä tutkimuksessa se kuvaa, kuinka suuri osa pistemäärien varianssista selittyy sillä, kuka on arvioinut vastauksen (Ellonen & Kaakinen s.a.). Kerroin voi saada arvoja väliltä 0–1 niin, että 1 kuvaa täydellistä arvioinnin yhdenmukaisuutta (Graham ym. 2012). Sisäkorrelaatiokerroin sopii kuvaamaan reliabiliteettia erityisen hyvin jatkuvilla mitta-asteikoilla, mutta se soveltuu reliabiliteetin tarkasteluun hyvin myös silloin, kun diskreettejä arvoluokkia on riittävästi eli tyypillisesti vähintään viisi. Tutkimuksessa tarkastellut pistemäärät olivat kokonaislukuja, mutta koska arvoluokkien lukumäärä määräytyi tehtävien enimmäispistemäärien mukaan ja oli jokaisessa osatehtävässä enemmän kuin viisi, sisäkorrelaatiokerroin soveltui käytettäväksi tutkimuksessa. Sisäkorrelaatiokerroin laskettiin myös nykyisillä arviointiohjeilla suoritetulle arvioinnille.

Arvioinnin käyttötarkoitus vaikuttaa siihen, mikä reliabiliteettitaso voidaan hyväksyä riittävän luotettavaksi, eikä millekään reliabiliteettimittarille ole määritetty yksiselitteistä luotettavuuden raja-arvoa (Brookhart & Chen 2015; Graham 2012). Tyypillisesti sisäkorrelaatiokertoimen hyväksyttävä arvo eli arvioinnin riittävästä yhdenmukaisuudesta kertova arvo vaihtelee välillä 0.8–0.9 (Graham 2012). Tässä tutkimuksessa sisäkorrelaatiokertoimen hyväksymisen alarajana käytettiin arvoa 0.8, ja korkeaksi reliabiliteetiksi luettiin arvo 0.9. Grahamin ym. (2012) mukaan nämä arvot ovat asiantuntijoiden tyypillisesti käyttämiä kynnyсарvoja riittävän yhdenmukaiselle arvioinnille. Sisäkorrelaatiokertoimen lisäksi arviointimatriisilla tuotettujen pisteiden välistä korrelaatiota tarkasteltiin Pearsonin korrelaatiokertoimen avulla pisteaineistojen noudattaessa normaalijakaumaa. Pearsonin korrelaatiokerroin mittaa kahden muuttujan välisen lineaarisen yhteyden voimakkuutta, ja se voi saada arvoja välillä -1 ja 1 niin, että 1 kuvaa täydellistä positiivista yhteyttä ja -1 täydellistä negatiivista yhteyttä (Kestilä-Kekkonen s.a.).

Arvioijien tuottamille pisteille laskettiin myös arvioijien välistä yhdenmukaisuutta kuvaava prosenttiluku (engl. *percent of exact agreement*), joka on myös tyypillinen arvioitsijareliabili-

teettä kuvaava tunnusluku (Graham 2012). Sillä kuvataan, kuinka suuri osuus pisteistä saa täysin saman arvon eri arvioijilta (Brookhart & Chen 2017; Graham ym. 2012). Yhdenmukaisuutta kuvaavalla prosenttiluvulla kuvattiin tässä tutkimuksessa täsmälleen täysin saman pistemäärän saaneiden vastausten osuutta sekä sellaisten vastausten osuutta, joissa arvioijien tuottamien pisteiden välinen ero oli korkeintaan yhden pisteen. Prosenttiluku laskettiin SPSS-tilastolaskentaohjelman *Compute variable* -toiminnon avulla. Yhdenmukaisuutta kuvaavat prosenttiluvut laskettiin sekä arviointimatriisin että nykyisten arviointiohjeiden avulla suoritettulle arvioinnille.

### 3.3.4 Vastausten analysointi ja laadullinen sisällönanalyysi

Tutkimuksen kolmannen tutkimuskysymyksen tarkoituksena oli selvittää, millaisissa vastauksissa erot arviointimatriisiin ja nykyisiin arviointiohjeisiin perustuvissa pisteissä ovat suurimpia. Vastaus tutkimuskysymykseen saatiin vertaamalla eri arviointimenetelmillä tuotettuja pisteitä keskenään ja analysoimalla laadullisesti niiden vastausten ominaisuuksia, joissa piste-erot arviointimenetelmien välillä olivat suurimpia.

Vastausten pisteissä esiintyvät arviointimenetelmien väliset erot selvitettiin ensin SPSS-tilastolaskentaohjelman *Compute variable* -toiminnon avulla. Vastauksia tarkasteltiin osatehtäväkohtaisesti, eli piste-ero kuvaa, miten suuri ero osatehtävän vastauksen saamissa pisteissä oli arviointimatriisiin ja nykyisiin arviointiohjeisiin perustuvan pisteityksen välillä. Pisteistä tarkasteltiin vastauksen alkuperäisiä sensoripisteitä eli nykyisillä arviointiohjeilla tuotettuja pisteitä sekä arvioijan 1 arviointimatriisilla tuottamia pisteitä. Koska vastauksia tarkasteltiin osatehtäväkohtaisesti ja arvioija 1 pisteitti arviointimatriisilla 60 kolmesta osatehtävästä koostuvaa vastausta, osatehtäväkohtaisia vastauksia oli tarkastelussa 180. Osatehtävien vastausten eri arviointimenetelmillä saamien pisteiden välinen ero vaihteli välillä 0–4, ja koska tutkimuksen tarkoituksena oli tarkastella suurimman piste-eron saaneita vastauksia, laadulliseen analyysiin otettiin ne osatehtävien vastaukset, joissa piste-ero arviointimenetelmien välillä oli 3 tai 4 pistettä. Vastauksia tarkasteltiin kahdessa osassa eli jakamalla osatehtävien vastaukset niihin, joiden pisteet arviointimatriisilla pisteitettynä laskivat sekä niihin, joissa pisteet nousivat.

Menetelmänä vastausten analysoinnissa käytettiin teorialähtöistä laadullista sisällönanalyysia luokittelun keinoin. Laadullinen sisällönanalyysi on analyysimenetelmä, jonka tarkoituksena on tuottaa tutkittavasta ilmiöstä sanallinen ja selkeä kuvaus (Tuomi & Sarajärvi 2018). Teorialähtöisessä sisällönanalyysissä analyysia ohjaa valmis analyysirunko tai käsitejärjestelmä, ja

analyysin ensimmäinen vaihe on analyysirungon luominen. Tässä tutkimuksessa vastausten tarkastelun analyysirunkona käytettiin muokattua versiota tutkimuksessa luodusta arviointimatriisista (liite 2). Arviointimatriisia muokattiin sisällönanalyysia varten niin, että matriisin kriteerien kuvauksista muodostettiin vastausten analysointia ohjaava ja 12 osa-alueesta koostuva analyysirunko (taulukko 4). Vastausten analyysi toteutettiin luokittelun avulla, jonka voidaan katsoa olevan sisällön teemoihin kohdistuva kvantitatiivinen analyysi (Tuomi & Sarajärvi 2018). Vastaukset luokiteltiin analyysirungon osa-alueiden laatuluokkien (liite 2) mukaisesti ja luokittelun määrällisiä tuloksia tarkasteltiin analyysin tuloksia kuvaavien taulukoiden avulla. Luokittelun tuloksista tarkasteltiin erityisesti sitä, painottuivatko vastaukset jossain analyysirungon osa-alueessa erityisen heikkoihin tai laadukkaisiin laatuluokkiin. Näin selvitettiin, korostuiko jokin ominaisuus niissä tutkimuksen aineistoon kuuluvissa vastauksissa, joissa pisteet eri arviointimenetelmien välillä erosivat toisistaan paljon.

Taulukko 4. Vastausten laadullisessa analyysissä käytetyn analyysirungon osa-alueet. Analyysirungon vasemmassa sarakkeessa on tutkimuksessa luodun arviointimatriisin kriteerit, ja oikeassa sarakkeessa on kriteereistä erotetut osa-alueet. Osa-alueiden laatuluokat esitetään liitteessä 2.

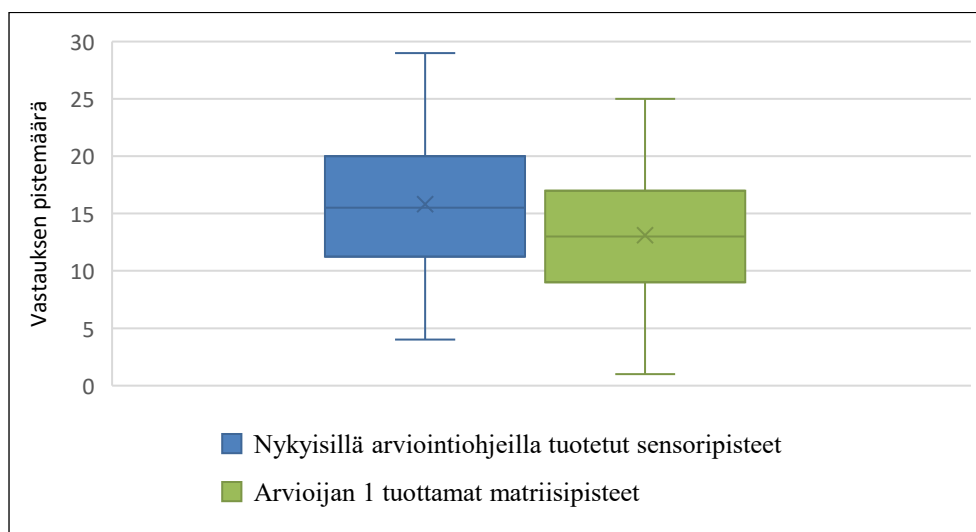
<b>Arviointimatriisin kriteeri</b>	<b>Kriteerin osa-alueet</b>
<b>Tehtävänannon noudattaminen</b>	Sisältöjen osuvuus
	Tehtävänannon mukainen tarkastelutapa
	Vastauksen esitystapa
<b>Sisällöt</b>	Huomioiden määrä
	Huomioiden oikeellisuus
	Huomioiden perustelut
<b>Maantieteellinen tarkastelu- ja jäsentelytapa</b>	Maantieteen näkökulmien hallinta
	Vastauksen looginen eteneminen ja rakenne
<b>Käsitteet</b>	Täsmällinen käsitteiden käyttö
	Havaintojen kytkeminen maantieteellisiin ilmiöihin tai prosesseihin
<b>Viestinnän keinot</b>	Suomenkielinen asiateksti
	Paikannimistö

## 4 Tulokset

### 4.1 Arviointimenetelmän vaikutus pistemäärään

#### 4.1.1 Koko tehtävän tarkastelu

Tutkimuksessa vertailtavat arviointimenetelmät tuottivat toisistaan eroavat pistejakaumat (kuva 5; taulukko 5). Arvioija 1 tuotti arviointimatriisiin perustuvalla pisteitysmenetelmällä systemaattisesti alhaisempia pistemääriä kuin Ylioppilastutkintolautakunnan sensorit tuottivat samoille vastauksille nykyisiin arviointiohjeisiin perustuvalla pisteitysmenetelmällä. Matriisiarvioinnilla tuotettujen pisteiden keskiarvo (13,12) oli nykyisillä arviointiohjeilla tuotettujen pisteiden keskiarvoa (15,82) huomattavasti vähemmän (taulukko 5). Vastaavasti myös pisteaineistojen mediaanien ja moodien vertaaminen osoitti matriisiarvioinnilla tuotettujen pisteiden sijoittuvan nykyisillä arviointiohjeilla tuotettuja pisteitä matalampiin pisteluokkiin. Pisteaineistojen keskihajonnat eivät eronneet toisistaan huomattavasti.



Kuva 5. Nykyisillä arviointiohjeilla tuotettujen sensoripisteiden pistejakauma sekä samoille vastauksille arvioijan 1 arviointimatriisilla tuottamien pisteiden pistejakauma (n=60).

Taulukko 5. Nykyisillä arviointiohjeilla tuotettujen sensoripisteiden ja arvioijan 1 tuottamien matriisipisteiden tilastollisia tunnuslukuja (n=60).

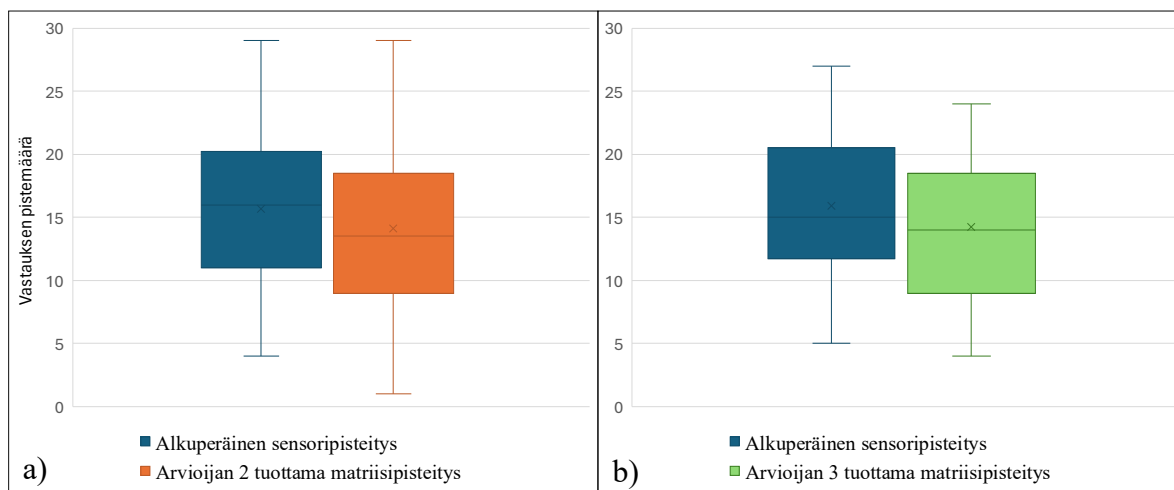
	Nykyisillä arviointiohjeilla tuotetut sensoripisteet	Arvioijan 1 arviointimatriisilla tuottamat pisteet
<b>Aritmeettinen keskiarvo</b>	15,82	13,12
<b>Mediaani</b>	15,50	13,00
<b>Moodi</b>	14	13
<b>Keskihajonta</b>	6,163	6,143
<b>Minimi</b>	4	1
<b>Maksimi</b>	29	25

Parittaisen t-testin p-arvo oli tutkimuksen riskitasoa (0,05) pienempi (taulukko 6). Täten parittaisen t-testin mukaan eri arviointimenetelmillä tuotettujen pisteiden välinen ero oli tilastollisesti merkitsevä. Se tarkoittaa, että tarkastelluilla arviointimenetelmillä tuotettujen pisteiden välinen ero ei todennäköisesti johtunut sattumasta, vaan arviointimatriisilla toteutetulla arvioinnilla saadaan yleisesti matalampia pisteitä kuin nykyisillä arviointiohjeilla.

Taulukko 6. Parittaisen t-testin tulokset eri arviointimenetelmillä tuotettujen pisteiden välisten erojen tilastollisesta merkitsevyydestä (n=60).

	Testimuuttujan arvo (t)	Vapausasteiden lukumäärä (df)	Testin p-arvo (2-suuntainen)
<b>Nykyisillä arviointiohjeilla tuotettujen sensoripisteiden ja arvioijan 1 tuottamien matriisipisteiden parittainen t-testi</b>	7,421	59	<0,001

Myös arvioijien 2 ja 3 arviointimatriisilla tuottamat pisteet erosivat vastausten alkuperäisistä sensoripisteistä eli nykyisillä arviointiohjeilla tuotetuista pisteistä tilastollisesti merkitsevästi (kuva 6). Vastauksille annettiin arviointimatriisiin perustuvalla arvioinnilla keskimäärin vähemmän pisteitä kuin nykyisiin arviointiohjeisiin perustuvalla arvioinnilla. Pistejakaumien mediaanit ja minimiarvot olivat molempien arvioijien matriisipisteissä matalammat kuin alkuperäisten sensoripisteiden vastaavat arvot, ja myös arvioijan 3 tuottamien matriisipisteiden maksimiarvo oli alkuperäisten sensoripisteiden maksimiarvoa matalampi.



Kuva 6a. Arvioijan 2 arviointimatriisilla tuottamien pisteiden pistejakauma (n=30) sekä samoille vastauksille nykyisillä arviointiohjeilla annettujen sensoripisteiden pistejakauma. Kuva 6b. Arvioijan 3 arviointimatriisilla tuottamien pisteiden pistejakauma (n=30) sekä samoille vastauksille nykyisillä arviointiohjeilla annettujen sensoripisteiden pistejakauma.

#### 4.1.2 Osatehtäväkohtainen tarkastelu

Ylioppilastutkintolautakunnan sensorien antamat pisteet eli nykyisillä arviointiohjeilla tuotetut pisteet erosivat arvioijan 1 tuottamista matriisipisteistä myös osatehtäväkohtaisesti tarkasteltuna (taulukko 7). Osatehtävien 7.1 ja 7.2 vastauksille annettiin matriisiarvioinnilla keskimäärin vähemmän pisteitä kuin nykyisiin arviointiohjeisiin perustuvalla arvioinnilla. Ainoastaan osatehtävän 7.3 pisteet nousivat arviointimatriisilla pisteitettynä hieman. Yksikään vastaus ei saavuttanut missään osatehtävässä tehtävän enimmäispistemäärää arviointimatriisilla pisteitettynä, vaikka nykyisillä arviointiohjeilla arvioituna kaikissa osatehtävissä vähintään yksi vastaus saavutti osatehtävän enimmäispistemäärän.

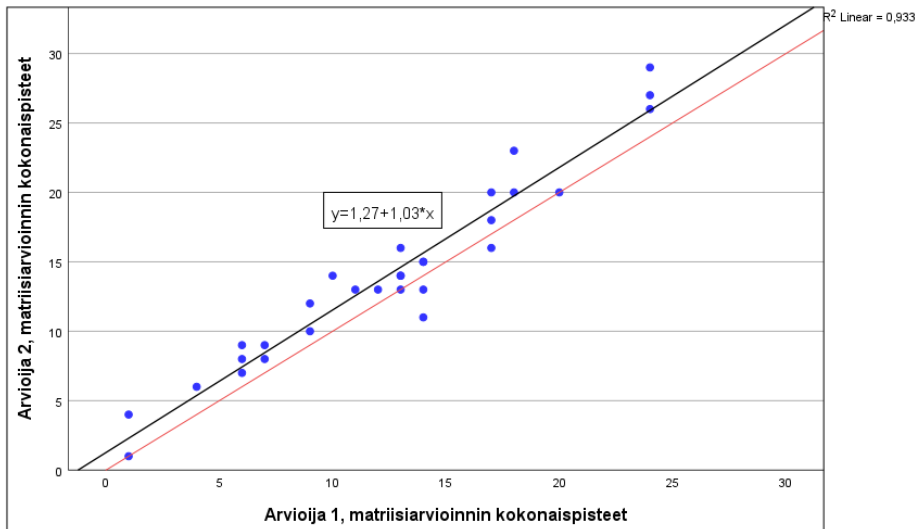
Taulukko 7. Nykyisillä arviointiohjeilla tuotettujen sensoripisteiden ja arvioijan 1 tuottamien matriisipisteiden tilastollisia tunnuslukuja osatehtäväkohtaisesti (n=60).

	Osatehtävä 7.1		Osatehtävä 7.2		Osatehtävä 7.3	
	Alku-peräinen sensori-arviointi	Arvioijan 1 tuottama matriisi-arviointi	Alku-peräinen sensori-arviointi	Arvioijan 1 tuottama matriisi-arviointi	Alku-peräinen sensori-arviointi	Arvioijan 1 tuottama matriisi-arviointi
<b>Keskiarvo</b>	4,05	2,47	6,90	5,68	4,87	4,97
<b>Mediaani</b>	4,00	2,50	7,00	6,00	4,00	5,00
<b>Moodi</b>	3	3	10	5	3	5
<b>Keskihajonta</b>	1,46	1,20	2,89	2,78	2,59	2,81
<b>Minimiarvo</b>	1	0	0	0	1	0
<b>Maksimiarvo</b>	6	5	12	11	12	11

## 4.2 Arviointimenetelmien arvioitsijareliabiliteetit

### 4.2.1 Arviointimatriisin arvioitsijareliabiliteetti

Arvioijien 1 ja 2 arviointimatriisilla antamien tehtävän 7 kokonaispisteiden vertaaminen osoitti, että arvioijat pisteittivät vastauksia melko yhdenmukaisesti, mutta arvioija 1 antoi systemaattisesti hieman matalampia pisteitä kuin arvioija 2 (kuva 7). Arvioijien välinen piste-ero oli riippumaton vastauksen tasosta, eli arvioija 1 antoi keskimäärin arvioijaa 2 matalampia pisteitä sekä heikko- että korkealaatuisissa vastauksissa.



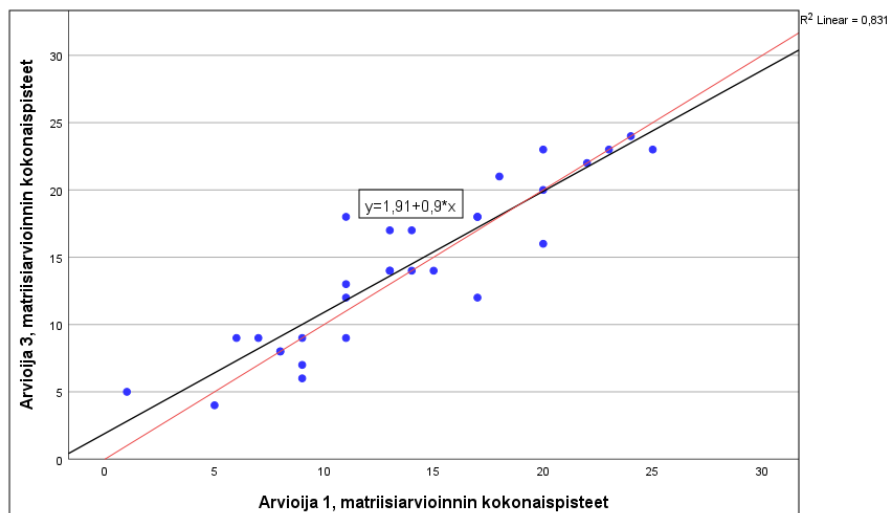
Kuva 7. Arvioijien 1 ja 2 arviointimatriisilla tuottamat pisteet hajontakaaviossa (n=30) sekä kaavioon sovitettu regressiosuora yhtälöineen. Punaisen vertailusuoran yhtälö on  $y=x$  eli se kuvaa tilannetta, jossa arvioijien antamat pisteet vastaisivat toisiaan täysin.  $R^2$  Linear tarkoittaa regressiomallin selitystasetta.

Sisäkorrelaatiokertoimella mitattuna arvioijien 1 ja 2 välinen yhdenmukaisuus oli arviointimatriisilla toteutetussa arvioinnissa pääasiassa erinomaista (taulukko 8). Arvioijien 1 ja 2 välinen sisäkorrelaatiokerroin (0,967) tehtävän kokonaispisteitä tarkasteltaessa kuvaa erinomaista arvioitsijareliabiliteettia arvioijien välillä. Myös osatehtävissä 7.2 ja 7.3 arvioijien välinen arvioitsijareliabiliteetti oli korkea. Sisäkorrelaatiokertoimen matalin arvo (0,843) arvioijien 1 ja 2 välillä oli osatehtävässä 7.1, mutta myös se voitiin laskea riittävän hyväksi sen ylittäessä tutkimukseen valitun raja-arvon 0,8. Sisäkorrelaatiokertoimen luottamusväli oli erityisen suuri osatehtävässä 7.1, mikä viittaa suurempaan pisteiden väliseen vaihteluun kyseisessä osatehtävässä. Arvioijien 1 ja 2 tuottamien kokonaispisteiden välinen Pearsonin korrelaatiokerroin oli 0,966 ja testin p-arvo oli alle 0,001. Testin mukaan arvioijien antamien pisteiden välillä oli siis erittäin vahva ja tilastollisesti merkitsevä positiivinen riippuvuus.

Taulukko 8. Arvioijien 1 ja 2 arviointimatriisilla tuottamien pisteiden välinen sisäkorrelaatiokerroin osatehtäväkohtaisesti ja koko tehtävän tasolla (n=30).

	Arvioijan 1 ja arvioijan 2 tuottamien pisteiden välinen sisäkorrelaatiokerroin	Sisäkorrelaatiokertoimen 95 %:n luottamusväli		Sisäkorrelaatiokertoimen tulkinta
		alaraja	yläraja	
<b>Osatehtävä 7.1</b>	0,843	0,403	0,942	hyväksyttävä
<b>Osatehtävä 7.2</b>	0,943	0,825	0,977	korkea
<b>Osatehtävä 7.3</b>	0,970	0,935	0,986	korkea
<b>Koko tehtävä</b>	0,967	0,787	0,989	korkea

Myös arvioijat 1 ja 3 pisteittivät vastauksia arviointimatriisilla melko yhdenmukaisesti (kuva 8). Arvioija 3 antoi heikkolaatuisille vastauksille keskimäärin hieman enemmän pisteitä ja korkealaatuisille vastauksille hieman vähemmän pisteitä kuin arvioija 1.



Kuva 8. Arvioijien 1 ja 3 arviointimatriisilla tuottamat pisteet hajontakaaviossa (n=30) sekä kaavioon sovitettu regressiosuora yhtälöineen. Punaisen vertailusuoran yhtälö on  $y=x$ , eli se kuvaa tilannetta, jossa arvioijien antamat pisteet vastaisivat toisiaan täysin. R2 Linear tarkoittaa regressiomallin selitystasetta.

Arvioijien 1 ja 3 arviointimatriisilla tuottamien pisteiden välinen sisäkorrelaatiokerroin ylitti jokaisessa osatehtävässä sekä tehtävän kokonaispisteitä tarkasteltaessa tutkimuksessa käytetyn korkean reliabiliteetin kynnyksarvon 0,9 (taulukko 9), joten arvioijien välinen yhdenmukaisuus oli sisäkorrelaatiokertoimella mitattuna erinomaista. Erityisen yhdenmukaista arviointi oli tehtävän kokonaispisteitä tarkasteltaessa, sillä arvioijien tuottamien kokonaispisteiden välinen sisäkorrelaatiokerroin oli 0,953. Arvioijien 1 ja 3 arviointimatriisilla tuottamien pisteiden välinen Pearsonin korrelaatiokerroin oli 0,912 ja testin p-arvo oli alle 0,001. Täten Pearsonin korrelaatioanalyysin mukaan arvioijien pisteiden välillä oli erittäin vahva ja tilastollisesti merkitsevä positiivinen riippuvuus.

Taulukko 9. Arvioijien 1 ja 3 arviointimatriisilla tuottamien pisteiden välinen sisäkorrelaatiokerroin osatehtäväkohtaisesti ja koko tehtävän tasolla (n=30).

	Arvioijan 1 ja arvioijan 3 tuottamien pisteiden välinen sisäkorrelaatiokerroin	Sisäkorrelaatiokertoimen 95 %:n luottamusväli		Sisäkorrelaatiokertoimen tulkinta
		alaraja	yläraja	
<b>Osatehtävä 7.1</b>	0,900	0,752	0,956	korkea
<b>Osatehtävä 7.2</b>	0,906	0,803	0,955	korkea
<b>Osatehtävä 7.3</b>	0,918	0,828	0,961	korkea
<b>Koko tehtävä</b>	0,953	0,903	0,978	korkea

Arvioijien välinen yhdenmukaisuus oli arviointimatriisin avulla toteutetussa arvioinnissa melko matalaa, kun yhdenmukaisuutta tarkasteltiin täsmälleen saman pistemäärän saaneiden vastausten prosenttiosuuden mukaan (taulukko 10). Arvioijat 1 ja 2 antoivat täsmälleen saman pistemäärän 10 %:lle vastauksista tehtävän kokonaispisteitä tarkasteltaessa, ja vastaava luku arvioijien 1 ja 3 välillä oli 26,7 %. Arvioijien 1 ja 2 antamat pisteet vastasivat toisiaan eniten osatehtävässä 7.3, jossa 43,3 % vastauksista sai matriisilla arvioituna täysin saman pistemäärän. Arvioijien 1 ja 3 antamien pisteiden yhdenmukaisuus oli suurinta osatehtävässä 7.1, jossa lähes puolet vastauksista (46,7 %) sai arvioijilta 1 ja 3 täysin saman pistemäärän.

Taulukko 10. Arvioijien välistä yhdenmukaisuutta kuvaavat prosenttiluvut osatehtäväkohtaisesti ja koko tehtävän tasolla. Prosenttiluku kuvaa, kuinka suuri osuus vastauksista sai arvioijilta täysin saman pistemäärän arviointimatriisilla toteutetussa arvioinnissa. Osatehtävässä 7.1 enimmäispistemäärä oli 6 pistettä ja muissa osatehtävissä 12 pistettä.

	Arvioijan 1 ja arvioijan 2 tuottamien pisteiden välinen yhdenmukaisuus (n=30)	Arvioijan 1 ja arvioijan 3 tuottamien pisteiden välinen yhdenmukaisuus (n=30)
<b>Osatehtävä 7.1</b>	33,3 %	46,7 %
<b>Osatehtävä 7.2</b>	23,3 %	20,0 %
<b>Osatehtävä 7.3</b>	43,3 %	30,0 %
<b>Koko tehtävä</b>	10,0 %	26,7 %

Arvioijien yhdenmukaisuutta kuvaavat prosenttiluvut ovat huomattavasti korkeammat, kun arvioijien antamien pisteiden välillä hyväksytään yhden pisteen ero (taulukko 11). Osatehtäväkohtaisesti tarkasteltuna kaikki yhdenmukaisuutta kuvaavat prosenttiluvut ovat vähintään 70,0 %, eli suuressa osassa osatehtävien vastauksia arvioijien antamat pisteet erosivat toisistaan korkeintaan yhdellä pisteellä. Koko tehtävän pisteet erosivat toisistaan korkeintaan yhdellä pisteellä hieman alle puolella (46,7 %) vastauksista arvioijien 1 ja 2 välillä ja tasan puolella (50,0 %) arvioijien 1 ja 3 välillä.

Taulukko 11. Arvioijien välistä yhdenmukaisuutta kuvaavat prosenttiluvut osatehtäväkohtaisesti ja koko tehtävän tasolla, kun pisteiden välillä sallittiin yhden pisteen ero. Prosenttiluku kuvaa, kuinka suuressa osassa vastauksia arvioijien antamat pisteet erosivat toisistaan korkeintaan yhdellä pisteellä. Osatehtävässä 7.1 enimmäispistemäärä oli 6 pistettä ja muissa osatehtävissä 12 pistettä.

	Arvioijan 1 ja arvioijan 2 tuottamien pisteiden välinen yhdenmukaisuus, ero pisteissä $\leq 1$ (n=30)	Arvioijan 1 ja arvioijan 3 tuottamien pisteiden välinen yhdenmukaisuus, ero pisteissä $\leq 1$ (n=30)
<b>Osatehtävä 7.1</b>	86,7 %	96,7 %
<b>Osatehtävä 7.2</b>	80,0 %	70,0 %
<b>Osatehtävä 7.3</b>	90,0 %	80,0 %
<b>Koko tehtävä</b>	46,7 %	50,0 %

#### 4.2.2 Nykyisten arviointiohjeiden arvioitsijareliabiliteetti

Sisäkorrelaatiokertoimella mitattuna arvioijan 2 nykyisillä arviointiohjeilla antamat pisteet olivat jokaisessa osatehtävässä hyväksyttävällä tasolla yhdenmukaiset vastausten alkuperäisten sensoripisteiden kanssa (taulukko 12). Tehtävän kokonaispisteitä tarkasteltaessa yhdenmukaisuus arvioijan 2 antamien pisteiden ja alkuperäisten sensoripisteiden välillä oli erinomaista.

Taulukko 12. Arvioijan 2 nykyisillä arviointiohjeilla tuottamien pisteiden ja alkuperäisten sensoripisteiden välinen sisäkorrelaatiokerroin osatehtäväkohtaisesti ja koko tehtävän tasolla (n=40).

	Arvioijan 2 tuottamien pisteiden ja alkuperäisten sensoripisteiden välinen sisäkorrelaatiokerroin	Sisäkorrelaatiokertoimen 95 %:n luottamusväli		Sisäkorrelaatiokertoimen tulkinta
		alaraja	yläraja	
<b>Osatehtävä 7.1</b>	0,840	0,698	0,916	hyväksyttävä
<b>Osatehtävä 7.2</b>	0,886	0,786	0,940	hyväksyttävä
<b>Osatehtävä 7.3</b>	0,887	0,785	0,940	hyväksyttävä
<b>Koko tehtävä</b>	0,935	0,879	0,966	korkea

Arvioijan 3 nykyisillä arviointiohjeilla antamien pisteiden ja alkuperäisten sensoripisteiden välinen arvioitsijareliabiliteetti oli pääasiassa korkea (taulukko 13). Sisäkorrelaatiokerroin oli korkein tehtävän kokonaispisteitä tarkasteltaessa. Matalin sisäkorrelaatiokertoimen arvo arvioijan 3 antamien pisteiden ja alkuperäisten sensoripisteiden välillä oli osatehtävässä 7.1 (0,854), mutta myös se voitiin hyväksyä tarpeeksi luotettavaksi tutkimuksen kynnysarvojen mukaan.

Taulukko 13. Arvioijan 3 nykyisillä arviointiohjeilla tuottamien pisteiden ja alkuperäisten sensoripisteiden välinen sisäkorrelaatiokerroin osatehtäväkohtaisesti ja koko tehtävän tasolla (n=40).

	Arvioijan 3 tuottamien pisteiden ja alkuperäisten sensoripisteiden välinen sisäkorrelaatiokerroin	Sisäkorrelaatiokertoimen 95 %:n luottamusväli		Sisäkorrelaatiokertoimen tulkinta
		alaraja	yläraja	
<b>Osatehtävä 7.1</b>	0,854	0,725	0,922	hyväksyttävä
<b>Osatehtävä 7.2</b>	0,937	0,881	0,967	korkea
<b>Osatehtävä 7.3</b>	0,930	0,869	0,963	korkea
<b>Koko tehtävä</b>	0,952	0,910	0,975	korkea

Nykyisillä arviointiohjeilla suoritetun arvioinnin yhdenmukaisuus oli melko samankaltainen kuin matriisiarvioinnin yhdenmukaisuus, kun yhdenmukaisuutta tarkasteltiin täsmälleen saman pistemäärän saaneiden vastausten prosenttiosuuden mukaan (taulukko 14). Arvioijan 2 tuottamat pisteet olivat yhdenmukaisimmat alkuperäisten sensoripisteiden kanssa osatehtävissä 7.1 ja 7.2, joissa 35,0 % vastauksista sai arvioijilta täysin saman pistemäärän. Arvioijan 3 pisteet

vastasivat alkuperäisiä sensoripisteitä eniten osatehtävässä 7.3, jossa 42,5 % vastauksista sai arvioinneista täysin saman pistemäärän. Tehtävän kokonaispisteet vastasivat molemmilla arvioijilla alkuperäisiä sensoripisteitä täysin viidessä vastauksessa (12,5 %).

Taulukko 14. Arvioijien välistä yhdenmukaisuutta kuvaavat prosenttiluvut osatehtäväkohtaisesti ja koko tehtävän tasolla. Prosenttiluku kuvaa, kuinka suuri osuus vastauksista sai nykyisillä arviointiohjeilla toteutetussa arvioinnissa täysin saman pistemäärän verrattuna vastauksen alkuperäisiin sensoripisteisiin. Osatehtävässä 7.1 enimmäispistemäärä oli 6 pistettä ja muissa osatehtävissä 12 pistettä.

	Arvioijan 2 tuottamien pisteiden ja alkuperäisten sensoripisteiden välinen yhdenmukaisuus (n=40)	Arvioijan 3 tuottamien pisteiden ja alkuperäisten sensoripisteiden välinen yhdenmukaisuus (n=40)
<b>Osatehtävä 7.1</b>	35,0 %	32,5 %
<b>Osatehtävä 7.2</b>	12,5 %	25,0 %
<b>Osatehtävä 7.3</b>	35,0 %	42,5 %
<b>Koko tehtävä</b>	12,5 %	12,5 %

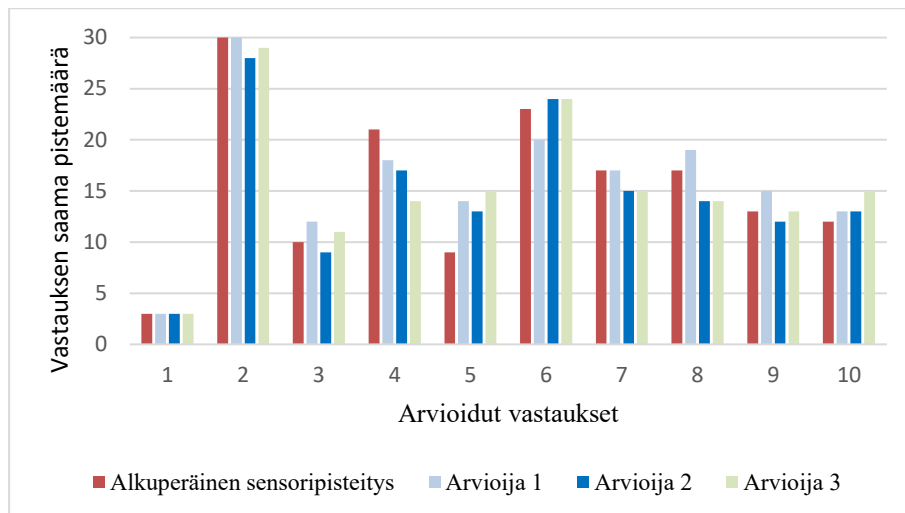
Alkuperäiset sensoripisteet erosivat korkeintaan yhdellä pisteellä arvioijien 2 ja 3 antamista pisteistä yli puolessa vastauksista osatehtäväkohtaisesti tarkasteltuna, kun arviointi suoritettiin nykyisten arviointiohjeiden mukaan (taulukko 15). Eniten pisteet vastasivat toisiaan osatehtävässä 7.1, jossa molempien arvioijien pisteet erosivat alkuperäisistä sensoripisteistä korkeintaan yhdellä pisteellä yli 80 %:ssa vastauksista. Tehtävän kokonaispisteitä tarkasteltaessa arvioijien 2 ja 3 antamat pisteet erosivat yli yhdellä pisteellä vastauksen alkuperäisistä sensoripisteistä yli puolessa vastauksista.

Taulukko 15. Arvioijien välistä yhdenmukaisuutta kuvaavat prosenttiluvut osatehtäväkohtaisesti ja koko tehtävän tasolla, kun pisteiden välillä sallittiin yhden pisteen ero. Prosenttiluku kuvaa, kuinka suuri osuus vastauksista sai nykyisillä arviointiohjeilla toteutetussa arvioinnissa korkeintaan yhdellä pisteellä eroavan pistemäärän verrattuna vastauksen alkuperäisiin sensoripisteisiin. Osatehtävässä 7.1 enimmäispistemäärä oli 6 pistettä ja muissa osatehtävissä 12 pistettä.

	Arvioijan 2 tuottamien pisteiden ja alkuperäisten sensoripisteiden välinen yhdenmukaisuus, ero pisteissä $\leq 1$ (n=40)	Arvioijan 3 tuottamien pisteiden ja alkuperäisten sensoripisteiden välinen yhdenmukaisuus, ero pisteissä $\leq 1$ (n=40)
<b>Osatehtävä 7.1</b>	87,5 %	82,5 %
<b>Osatehtävä 7.2</b>	52,5 %	72,5 %
<b>Osatehtävä 7.3</b>	67,5 %	75,0 %
<b>Koko tehtävä</b>	40,0 %	47,5 %

Tutkimuksen aineistoon kuului 10 vastausta, jotka kaikki tutkimuksen arvioijat pisteittivät nykyisillä arviointiohjeilla. Lähes kaikissa näissä vastauksissa arvioijien antamat pistemäärät erosivat toisistaan ja vastauksen alkuperäisistä sensoripisteistä, kun tarkasteltiin osatehtävien pisteistä koostuvaa tehtävän kokonaispistemäärää (enimmillään 30) (kuva 9). Ainoastaan yhdessä vastauksessa kaikki neljä arvioijaa antoivat vastaukselle saman kokonaispistemäärän. Kolmessa vastauksessa kaikki arvioijat antoivat vastaukselle eri kokonaispistemäärän. Suurin ero

kahden arvioijan tuottamissa kokonaispisteissä oli neljännessä vastauksessa, jossa alkuperäiset sensoripisteet erosivat arvioijan 3 antamasta pistemäärästä seitsemällä pisteellä.



Kuva 9. Kymmenen vastauksen alkuperäiset sensoripisteet ja tutkimuksen arvioijien nykyisillä arviointiohjeilla vastauksille antamat pisteet tehtävän kokonaispistemäärää tarkasteltaessa, kun tehtävän enimmäispistemäärä oli 30.

#### 4.3 Arviointimatriisin kriteerikohtainen reliabiliteetti

Arvioinnin yhdenmukaisuudessa ei ilmennyt huomattavia eroja arviointimatriisin eri kriteerien välillä, ja yhdenmukaisuus vaihteli jokaisessa kriteerissä eri osatehtävien välillä (taulukko 16). Arvioinnin yhdenmukaisuus oli keskimäärin korkeinta kriteerissä ”Maantieteellinen tarkastelu ja jäsentely” sekä arvioijien 1 ja 2 että arvioijien 1 ja 3 välillä.

Taulukko 16. Arvioijien kriteerikohtaista yhdenmukaisuutta kuvaavat prosenttiluvut osatehtäväkohtaisesti. Prosenttiluku kuvaa, kuinka suuri osa osatehtävän vastauksista sai arvioijilta täysin saman arvioinnin arviointimatriisin kriteerissä.

	Arvioijan 1 ja arvioijan 2 yhdenmukaisuus kriteereittäin (n=30)			Arvioijan 1 ja arvioijan 3 yhdenmukaisuus kriteereittäin (n=30)		
	Osa- tehtävä 7.1	Osa- tehtävä 7.2	Osa- tehtävä 7.3	Osa- tehtävä 7.1	Osa- tehtävä 7.2	Osa- tehtävä 7.3
<b>Tehtävänannon noudattaminen</b>	46,7 %	50,0 %	60,0 %	63,3 %	53,3 %	50,0 %
<b>Sisällöt</b>	36,7 %	73,3 %	70,0 %	50,0 %	60,0 %	60,0 %
<b>Maantieteellinen tarkastelu ja jäsentely</b>	50,0 %	70,0 %	76,7 %	63,3 %	50,0 %	70,0 %
<b>Käsitteet</b>	43,3 %	66,7 %	63,3 %	50,0 %	36,7 %	50,0 %
<b>Viestinnän keinot</b>	36,7 %	73,3 %	63,3 %	56,7 %	50,0 %	50,0 %

#### 4.4 Ominaisuuksia vastauksissa, joissa arviointimenetelmien välillä esiintyi piste-eroja

##### 4.4.1 Niiden vastausten ominaisuuksia, joiden pisteitä matriisiarviointi laski eniten

Osatehtävien vastausten (n=180) alkuperäisten sensoripisteiden sekä arvioijan 1 arviointimatriisilla tuottamien pisteiden vertaaminen osoitti, että suuri osa osatehtävien vastauksista eli 114 vastausta sai arviointimatriisilla toteutetussa arvioinnissa vähemmän pisteitä kuin nykyisillä arviointiohjeilla toteutetussa arvioinnissa (taulukko 17). Suurin osatehtäväkohtainen piste-ero arviointimenetelmien välillä oli neljä pistettä, ja laadullisessa analyysissä tarkasteltiin niitä vastauksia, joiden pisteet erosivat eri arviointimenetelmillä pisteitettynä kolmella tai neljällä pisteellä. Laadullisesti tarkasteltiin siis 14 sellaista osatehtävän 7.1 vastausta ja 16 osatehtävän 7.2 vastausta, joiden pisteet laskivat arviointimatriisilla arvioituna. Osatehtävässä 7.3 suurin pistemäärän lasku oli kaksi pistettä (taulukko 17), minkä vuoksi osatehtävän vastaukset jäivät tarkastelun ulkopuolelle.

Taulukko 17. Niiden osatehtävien vastausten lukumäärä pistemäärän laskun mukaisesti, joissa arvioijan 1 antama arviointimatriisiin perustuva pistemäärä oli alhaisempi kuin nykyiseen pisteitysohjeeseen perustuva pistemäärä (n=114). Piste-ero kuvaa, kuinka monta pistettä vähemmän osatehtävän vastaus sai arviointimatriisilla arvioituna kuin nykyisillä arviointiohjeilla arvioituna.

Laskevan pistemäärän piste-ero arviointimenetelmien välillä	Osatehtävän 7.1 vastaukset, joiden pistemäärä laski matriisiarvioinnissa (n=51)	Osatehtävän 7.2 vastaukset, joiden pistemäärä laski matriisiarvioinnissa (n=39)	Osatehtävän 7.3 vastaukset, joiden pistemäärä laski matriisiarvioinnissa (n=24)	Vastausten määrä yhteensä (n=114)
1	21	11	17	49
2	16	12	7	35
3	13	14	0	27
4	1	2	0	3

Puolet vastauksista käsitteli pääosin tai täysin tehtävänannon kannalta olennaisia asioita, kun tarkasteltiin niitä osatehtävien vastauksia, joiden matriisiarvioinnista saama pistemäärä oli vähintään kolme pistettä nykyisiin arviointiohjeisiin perustuvaa pistemäärää alhaisempi (n=30) (taulukko 18). Sen sijaan suuressa osassa vastauksia ei noudatettu tehtävänannon mukaista tarkastelutapaa, mikä tarkoittaa, että vastauksista ei löytynyt esimerkiksi tehtävänannossa edellytettyä vertailuasetelmaa. Vastaukset noudattivat pääasiassa hyvin tehtävissä pyydettyä essemäistä esitystapaa.

Taulukko 18. Niiden vastausten jakautuminen laatuluokkiin osa-alueittain tehtävänannon noudattaminen -kriteerissä, joiden pistemäärä laski eniten eli kolmella tai neljällä pisteellä arviointimatriisilla arvioituna verrattuna nykyisiin arviointiohjeisiin (n=30).

Arviointimatriisin kriteeri	Kriteerin osa-alueet	Osa-alueen laatuluokkien kuvaukset (0, 5, 10, 15, 20)	Lukumäärä
Tehtävänannon noudattaminen	Sisältöjen osuvuus	ei vastaa tehtävänantoon; tehtävä ymmärretty väärin	3
		osin tehtävänannon mukainen; heikosti rajattu, epäselvä tai harhaileva	4
		pääosin tehtävänannon mukainen; pienissä määrin epäolennaisuuksia	8
		pääosa vastauksen sisällöistä on olennaisia	12
		käsittelee kattavasti ja harkitusti olennaiset asiat; ei tehtävänannon kannalta epäoleellisia asioita	3
	Tehtävänannon mukainen tarkastelutapa	annettua tarkastelutapaa ei noudateta	5
		ei täysin tehtävänannon mukainen	6
		pääosin tehtävänannon mukainen	11
		selkeä ymmärrys tehtävänannosta	7
		erinomainen ymmärrys tehtävänannosta	1
	Vastauksen esitystapa	annettua esitystapaa ei noudateta	1
		ei täysin tehtävänannon mukainen	5
		pääosin tehtävänannon mukainen	1
		selkeä ymmärrys esitystavasta	15
		erinomainen ymmärrys esitystavasta	8

Sekä huomioiden määrä että oikeellisuus vaihtelivat tarkastelluissa vastauksissa paljon, ja suurin osa vastauksista oli keskinkertaisia molemmissa osa-alueissa (taulukko 19). Sen sijaan huomioiden perustelu oli vastauksissa hyvin puutteellista, ja suurimmassa osassa vastauksia esitetyt huomiot perusteltiin vain niukasti tai niitä ei perusteltu lainkaan.

Taulukko 19. Niiden vastausten jakautuminen laatuluokkiin osa-alueittain sisällöt-kriteerissä, joiden pistemäärä laski eniten eli kolmella tai neljällä pisteellä arviointimatriisilla arvioituna verrattuna nykyisiin arviointiohjeisiin (n=30).

Arviointimatriisin kriteeri	Kriteerin osa-alueet	Osa-alueen laatuluokkien kuvaukset (0, 5, 10, 15, 20)	Lukumäärä	
Sisällöt	Huomioiden määrä	- (pääasiassa vain virheellistä sisältöä)	2	
		yksittäisiä oikein kuvattuja huomioita	8	
		muutamia laajemmin käsiteltyjä huomioita/useita suppeammin käsiteltyjä huomioita	16	
		muutamia syvällisesti käsiteltyjä huomioita/useita pinnallisemmin käsiteltyjä huomioita	2	
		tehtävänannon mukaisia huomioita kattavasti	2	
	Huomioiden oikeellisuus	pääasiassa vain virheellistä sisältöä	yksittäisiä oikein kuvattuja sisältöjä; jos useita huomioita, voi olla muutamia asiavirheitä	2
			yksi asiavirhe tai jonkin verran epätasällisyyttä	6
			lievää epätasällisyyttä; ei varsinaisia virheitä	12
			ei asiavirheitä tai epätasällisyyttä	7
				3
	Huomioiden perustelut	huomioita ei perusteltu	osalle huomioista voidaan esittää niukat perustelut	6
			muutamalla huomiolla tyydyttävät perustelut/usealla huomiolla hatarat perustelut	13
			muutamia syvällisesti ja virheettömästi perusteltuja/useita pinnallisesti perusteltuja huomioita	9
			huomiot perusteltu osuvasti, virheettömästi ja tarkoituksenmukaisesti	2
				0

Tarkastellut vastaukset olivat pääasiassa heikkoja maantieteellisen tarkastelu- ja jäsentelytavan näkökulmasta (taulukko 20). Maantieteellisten näkökulmien esittäminen oli suuressa osassa vastauksia niukkaa, minkä lisäksi vastaukset olivat rakenteeltaan pääosin puutteellisia tai epäloogisia.

Taulukko 20. Niiden vastausten jakautuminen laatuluokkiin osa-alueittain maantieteellinen tarkastelu- ja jäsentelytapa -kriteerissä, joiden pistemäärä laski eniten eli kolmella tai neljällä pisteellä arviointimatriisilla arvioituna verrattuna nykyisiin arviointiohjeisiin (n=30).

Arviointimatriisin kriteeri	Kriteerin osa-alueet	Osa-alueen laatuluokkien kuvaukset (0, 5, 10, 15, 20)	Lukumäärä	
Maantieteellinen tarkastelu- ja jäsentelytapa	Maantieteen näkökulmien hallinta	ei olennaisia näkökulmia	4	
		joitain olennaisia näkökulmia; epätasapainoinen vastaus	17	
		jäsennelty päänäkökulmien mukaisesti	8	
		jäsennelty tehtävänantoon kuuluvien monipuolisten näkökulmien mukaisesti; näkökulmia liitetty toisiinsa	1	
		monipuoliset näkökulmat ja niiden väliset vuorovaikutussuhteet kuvattu; johdonmukainen jäsentely	0	
	Vastauksen looginen eteneminen ja rakenne	rakenne hajanainen ja sekava/vastaus hyvin suppea	puutteellinen rakenne, eteneminen epäloogista	7
			pääosin selkeä rakenne; joitain pieniä rakenteellisia puutteita tai ei etene täysin loogisesti	8
			toimiva rakenne, etenee loogisesti	14
			eheä ja tasapainoinen rakenne	0

Käsitteiden käyttö oli tarkastelluissa vastauksissa pääasiassa hyvin puutteellista: valtaosassa vastauksia maantieteellisiä käsitteitä ei käytetty lainkaan tai niitä käytettiin joko hyvin vähän tai virheellisesti (taulukko 21). Vastauksissa esitetyt havainnot kytkettiin vain harvoin maantieteellisiin ilmiöihin ja prosesseihin, ja pääasiassa kytkeminen jäi vain maininnan tasolle.

Taulukko 21. Niiden vastausten jakautuminen laatuluokkiin osa-alueittain käsitteet-kriteerissä, joiden pistemäärä laski eniten eli kolmella tai neljällä pisteellä arviointimatriisilla arvioituna verrattuna nykyisiin arviointiohjeisiin (n=30).

Arviointimatriisin kriteeri	Kriteerin osa-alueet	Osa-alueen laatuluokkien kuvaukset (0, 5, 10, 15, 20)	Lukumäärä
Käsitteet	Täsmällinen käsitteiden käyttö	ei käytetä käsitteitä; arkikielinen vastaus	9
		käsitteidenkäytössä selviä puutteita tai virheitä	14
		käsitteitä käytetään; käytössä lieviä puutteita tai yksittäinen virhe	5
		käytetään keskeisiä käsitteitä; ei varsinaisia virheitä	2
		käsitteitä käytetään monipuolisesti, täsmällisesti ja huolellisesti	0
	Havaintojen kytkeminen maantieteellisiin ilmiöihin tai prosesseihin	ei kytketä lainkaan ilmiöihin tai prosesseihin	11
		saatetaan mainita jokin ilmiö tai prosessi	13
		viitataan pintapuolisesti ilmiöön tai prosessiin	4
		hyödynnetään maantieteellistä teoriaa; kytkeminen ilmiöihin ja prosesseihin kevyttä	2
		asiat kytketty hyvin ilmiöihin ja prosesseihin	0

Suomenkielisen asiatekstin laatu vaihteli tarkastelluissa vastauksissa jonkin verran, mutta suuri osa vastauksista oli pääosin selkeää asiatekstiä (taulukko 22). Paikannimistön esittämistä arviointiin vain osatehtävän 7.2 vastauksista, sillä vain kyseisessä osatehtävässä edellytettiin paikannimistön käyttöä. Tarkastelluissa vastauksissa paikannimistöä käytettiin pääasiassa oikein ja asianmukaisesti.

Taulukko 22. Niiden vastausten jakautuminen laatuluokkiin osa-alueittain viestinnän keinot -kriteerissä, joiden pistemäärä laski eniten eli kolmella tai neljällä pisteellä arviointimatriisilla arvioituna verrattuna nykyisiin arviointiohjeisiin.

Arviointimatriisin kriteeri	Kriteerin osa-alueet	Osa-alueen laatuluokkien kuvaukset (0, 5, 10, 15, 20)	Lukumäärä
Viestinnän keinot	Suomenkielinen asiateksti (n=30)	ilmaisu vaikeasti ymmärrettävää	1
		osittain puhekielinen/kömpelö/paljon kielioppivirheitä	8
		pääosin selkeää asiatekstiä; voi olla paikoitellen kirjoitusvirheitä	12
		selkeää ja lähes virheetöntä asiatekstiä	9
		kirjoitusasu täysin virheetön	0
	Paikannimistö (n=16)	paikannimet kirjoitettu kokonaan tai pääosin väärin	1
		paikannimissä paikoitiin asia- tai kirjoitusvirheitä	0
		paikannimissä pieniä kirjoitusvirheitä	2
		paikannimet kirjoitettu pääosin oikein	5
		paikannimet kirjoitettu virheettömästi	8

#### 4.4.2 Niiden vastausten ominaisuuksia, joiden pisteitä matriisiarviointi nosti eniten

Tutkimuksessa tarkasteltuihin osatehtävien vastauksiin (n=180) kuului yhteensä vain seitsemän vastausta, joiden pisteet nousivat kolmella tai neljällä pisteellä arviointimatriisilla arvioituna (taulukko 23). Siten laadullisesti analysoitiin yhtä osatehtävän 7.2 vastausta ja kuutta osatehtävän 7.3 vastausta.

Taulukko 23. Niiden osatehtävien vastausten lukumäärä pistemäärän nousun mukaisesti, joissa arvioijan 1 antama arviointimatriisiin perustuva pistemäärä oli korkeampi kuin nykyiseen pisteitysohjeeseen perustuva pistemäärä (n=28). Piste-ero kuvaa, kuinka monta pistettä enemmän osatehtävän vastaus sai arviointimatriisilla arvioituna kuin nykyisillä arviointiohjeilla arvioituna.

Nousevan pistemäärän piste-ero arviointimenetelmien välillä	Osatehtävän 7.1 vastaukset, joiden pistemäärä nousi matriisiarvioinnissa (n=1)	Osatehtävän 7.2 vastaukset, joiden pistemäärä nousi matriisiarvioinnissa (n=7)	Osatehtävän 7.3 vastaukset, joiden pistemäärä nousi matriisiarvioinnissa (n=20)	Vastausten määrä yhteensä (n=28)
1	1	3	10	14
2	0	3	4	7
3	0	1	5	6
4	0	0	1	1

Vastaukset, joiden pisteet nousivat arviointimatriisilla pisteitettynä eniten, noudattivat tehtävänantoa erinomaisesti (taulukko 24). Kaikissa tarkastelluissa vastauksissa esitettiin siis täysin tai pääasiassa tehtävänannon kannalta olennaisia asioita ja tehtävänannon mukaista tarkastelutapaa noudatettiin hyvin. Erityisesti tehtävänannossa edellytettyä esseemäisitä vastaustapaa noudatettiin vastauksissa erinomaisesti.

Taulukko 24. Niiden vastausten jakautuminen laatuluokkiin osa-alueittain tehtävänannon noudattaminen -kriteerissä, joiden pistemäärä nousi eniten eli kolmella tai neljällä pisteellä arviointimatriisilla arviointina verrattuna nykyisiin arviointiohjeisiin (n=7).

Arviointimatriisin kriteeri	Kriteerin osa-alueet	Osa-alueen laatuluokkien kuvaukset (0, 5, 10, 15, 20)	Lukumäärä
Tehtävänannon noudattaminen	Sisältöjen osuvuus	ei vastaa tehtävänantoon; tehtävä ymmärretty väärin	0
		osin tehtävänannon mukainen; heikosti rajattu, epäselvä tai harhaileva	0
		pääosin tehtävänannon mukainen; pienissä määrin epäolennaisuuksia	0
		pääosa vastauksen sisällöistä on olennaisia	5
		käsittelee kattavasti ja harkitusti olennaiset asiat; ei tehtävänannon kannalta epäoleellisia asioita	2
	Tehtävänannon mukainen tarkastelutapa	annettua tarkastelutapaa ei noudateta	0
		ei täysin tehtävänannon mukainen/vastaus hyvin suppea	0
		pääosin tehtävänannon mukainen	1
		selkeä ymmärrys tehtävänannosta	4
		erinomainen ymmärrys tehtävänannosta	2
	Vastauksen esitystapa	annettua esitystapaa ei noudateta	0
		ei täysin tehtävänannon mukainen	0
		pääosin tehtävänannon mukainen	0
		selkeä ymmärrys esitystavasta	2
		erinomainen ymmärrys esitystavasta	5

Myös sisältöjen osalta vastaukset olivat pääasiassa laadukkaita (taulukko 25). Vastausten huomiot esitettiin pääasiassa ilman asiavirheitä tai epätasällisyyttä. Huomioiden sekä niille esitettyjen perustelujen määrä ja laatu vaihtelivat vastauksissa, mutta kaikissa vastauksissa huomioita esitettiin kohtuullinen määrä eikä yhdessäkään vastauksessa huomioita jätetty täysin perusteluitta.

Taulukko 25. Niiden vastausten jakautuminen laatuluokkiin osa-alueittain sisällöt-kriteerissä, joiden pistemäärä nousi eniten eli kolmella tai neljällä pisteellä arviointimatriisilla arvioituna verrattuna nykyisiin arviointiohjeisiin (n=7).

Arviointimatriisin kriteeri	Kriteerin osa-alueet	Osa-alueen laatuluokkien kuvaukset (0, 5, 10, 15, 20)	Lukumäärä	
Sisällöt	Huomioiden määrä	- (pääasiassa vain virheellistä sisältöä)	0	
		yksittäisiä oikein kuvattuja huomioita	0	
		muutamia laajemmin käsiteltyjä huomioita/useita suppeammin käsiteltyjä huomioita	3	
		muutamia syvällisesti käsiteltyjä huomioita/useita pinnallisemmin käsiteltyjä huomioita	2	
		tehtävänannon mukaisia huomioita kattavasti	2	
	Huomioiden oikeellisuus	pääasiassa vain virheellistä sisältöä	yksittäisiä oikein kuvattuja sisältöjä; jos useita huomioita, voi olla muutamia asiavirheitä	0
			yksi asiavirhe tai jonkin verran epätasaisuutta	1
			lievää epätasaisuutta; ei varsinaisia virheitä	3
			ei asiavirheitä tai epätasaisuutta	3
			huomioita ei perusteltu	0
	Huomioiden perustelut	osalle huomioista voidaan esittää niukat perustelut	muutamia huomioita perusteluineen/useita huomioita hatarilla perusteluilla	0
			muutamia syvällisesti ja virheettömästi perusteltuja/useita pinnallisesti perusteltuja huomioita	3
			huomiot perusteltu osuvasti, virheettömästi ja taroituksenmukaisesti	3
				1
				1

Maantieteellisten näkökulmien hallinta oli vastauksissa pääasiassa onnistunutta, ja kaikki vastaukset olivat vähintäänkin päänäkökulmien mukaisesti jäsenneily (taulukko 26). Vastaukset olivat rakenteeltaan enimmäkseen toimivia ja loogisesti eteneviä.

Taulukko 26. Niiden vastausten jakautuminen laatuluokkiin osa-alueittain maantieteellinen tarkastelu- ja jäsentelytapa -kriteerissä, joiden pistemäärä nousi eniten eli kolmella tai neljällä pisteellä arviointimatriisilla arvioituna verrattuna nykyisiin arviointiohjeisiin (n=7).

Arviointimatriisin kriteeri	Kriteerin osa-alueet	Osa-alueen laatuluokkien kuvaukset (0, 5, 10, 15, 20)	Lukumäärä	
Maantieteellinen tarkastelu- ja jäsentelytapa	Maantieteiden näkökulmien hallinta	ei olennaisia näkökulmia	0	
		joitain olennaisia näkökulmia; poukkoileva ja epätasapainoinen vastaus	0	
		jäsenneily päänäkökulmien mukaisesti	3	
		jäsenneily tehtävänantoon kuuluvien monipuolisten näkökulmien mukaisesti; näkökulmia liitetty toisiinsa	2	
		monipuoliset näkökulmat ja niiden väliset vuorovaikutussuhteet kuvattu; johdonmukainen jäsentely	2	
	Vastauksen looginen eteneminen ja rakenne	rakenne hajanainen ja sekava/vastaus hyvin suppea	puutteellinen rakenne, eteneminen epäloogista	0
			pääosin selkeä rakenne; joitain pieniä rakenteellisia puutteita tai ei etene täysin loogisesti	0
			toimiva rakenne, etenee loogisesti	2
			eheä ja tasapainoinen rakenne	5
				0

Vastaukset sijoittuivat kolmeen parhaaseen laatuluokkaan myös käsitteet-kriteerin osa-alueissa (taulukko 27). Siten vastauksissa siis käytettiin maantieteellisiä käsitteitä pääasiassa asianmukaisesti ja oikein, eikä yksikään tarkastelluista vastauksista ollut täysin arkikielinen. Vastauksissa esitetyt huomiot kytkettiin myös pääasiassa maantieteelliseen teoriaan, vaikka osassa vastauksista viittaaminen jäikin pintapuoliseksi.

Taulukko 27. Niiden vastausten jakautuminen laatuluokkiin osa-alueittain käsitteet-kriteerissä, joiden pistemäärä nousi eniten eli kolmella tai neljällä pisteellä arviointimatriisilla arvioituna verrattuna nykyisiin arviointiohjeisiin (n=7).

Arviointimatriisin kriteeri	Kriteerin osa-alueet	Osa-alueen laatuluokkien kuvaukset (0, 5, 10, 15, 20)	Lukumäärä
Käsitteet	Täsmällinen käsitteiden käyttö	ei käytetä käsitteitä; arkikielinen vastaus	0
		käsitteidenkäytössä selviä puutteita tai virheitä	0
		käsitteitä käytetään; käytössä lieviä puutteita tai yksittäinen virhe	3
		käytetään keskeisiä käsitteitä; ei varsinaisia virheitä	2
		käsitteitä käytetään monipuolisesti, täsmällisesti ja huolellisesti	2
	Havaintojen kytkeminen maantieteellisiin ilmiöihin tai prosesseihin	ei kytketä lainkaan ilmiöihin tai prosesseihin	0
		saatetaan mainita jokin ilmiö tai prosessi	0
		viitataan pintapuolisesti ilmiöön tai prosessiin	3
		hyödynnetään maantieteellistä teoriaa; kytkeminen ilmiöihin ja prosesseihin kevyttä	2
		asiat kytketty hyvin ilmiöihin ja prosesseihin	2

Vastaukset olivat laadukkaita sekä suomenkielisen asiatekstin että paikannimistön käytön osalta (taulukko 28). Lähes kaikki vastaukset olivat selkeää ja lähes virheetöntä asiatekstiä, ja kaikissa vastauksissa paikannimistö oli kirjoitettu joko virheettömästi tai pääosin oikein.

Taulukko 28. Niiden vastausten jakautuminen laatuluokkiin osa-alueittain viestinnän keinot -kriteerissä, joiden pistemäärä nousi eniten eli kolmella tai neljällä pisteellä arviointimatriisilla arvioituna verrattuna nykyisiin arviointiohjeisiin (n=7).

Arviointimatriisin kriteeri	Kriteerin osa-alueet	Osa-alueen laatuluokkien kuvaukset (0, 5, 10, 15, 20)	Lukumäärä
Viestinnän keinot	Suomenkielinen asiateksti	ilmaisu vaikeasti ymmärrettävää	0
		osittain puhekielinen/kömpelö/paljon kielioppivirheitä	0
		pääosin selkeää asiatekstiä; voi olla paikoitellen kirjoitusvirheitä	1
		selkeää ja lähes virheetöntä asiatekstiä	6
		kirjoitusasu täysin virheetön	0
	Paikannimistö	paikannimet kirjoitettu kokonaan tai pääosin väärin	0
		paikannimissä paikoitiin asia- tai kirjoitusvirheitä	0
		paikannimissä pieniä kirjoitusvirheitä	0
		paikannimet kirjoitettu pääosin oikein	2
		paikannimet kirjoitettu virheettömästi	5

## 5 Keskustelu

### 5.1 Matriisiarviointi tuotti nykyohjeistusta matalampia pisteitä

Tässä tutkimuksessa arviointimatriisiin perustuvalla pisteitysmenetelmällä tuotettiin keskimäärin matalampia pisteitä kuin nykyisiin arviointiohjeisiin eli hyvän vastauksen piirteisiin perustuvalla pisteitysmenetelmällä. Kaikki tutkimuksen arvioijat antoivat tarkastelluille vastauksille matriisiarvioinnissa keskimäärin matalammat kokonaispisteet verrattuna vastausten alkuperäisessä sensoriarvioinnissa eli nykyisillä arviointiohjeilla saamiin pisteisiin. Osatehtäväkohtaisesti tarkasteltuna arviointimatriisilla toteutetulla arvioinnilla tuotettiin kahdessa ensimmäisessä osatehtävässä keskimäärin matalammat pisteet kuin nykyisillä arviointiohjeilla, kun taas viimeisen osatehtävän pistemäärää matriisiarviointi nosti keskimäärin hieman nykyisillä arviointiohjeilla toteutettuun arviointiin verrattuna.

On mielenkiintoista pohtia, mistä arviointimenetelmien väliset piste-erot mahdollisesti johtuivat. Arviointimenetelmät painottivat erilaisia asioita, joten eroavat pistejakaumat olivat ennakoitavissa. Nykyiset arviointiohjeet korostivat tarkastellussa tehtävässä huomioiden määrää, kun taas matriisiarvioinnissa huomioitiin sisältöjen lukumäärän lisäksi muun muassa käsitteiden käyttö, maantieteen näkökulmien hallinta sekä asiatekstin laatu. Aiemmassa tutkimuskirjallisuudessa on todettu, että arviointimatriisit voivat esimerkiksi laajentaa arvioinnin kohteita tai vahvistaa tietynlaisen osaamisen, kuten kriittisen ajattelun sekä tiedon prosessoinnin, arviointia (Hidayati ym. 2024; Reynders ym. 2020), ja myös tässä tutkimuksessa havaitut arviointimenetelmien väliset piste-erot voivat viitata siihen, että arviointitavat tekivät näkyväksi erilaista vastauksissa esitettyä osaamista. Arviointimatriisilla toteutetulla arvioinnilla arvioijat pystyivät arvioimaan useampia vastauksen ominaisuuksia kuin nykyisillä arviointiohjeilla, mikä vaikutti alentavasti tutkielmassa tarkasteltujen vastausten pisteisiin. Vastauksen oli oltava arviointimatriisiin jokaisen kriteerin osalta laadukas saadakseen korkeat pisteet matriisiarvioinnista, kun taas nykyisillä arviointiohjeilla pisteitettynä vastaus saattoi yltyä melko korkeisiin pisteisiin pelkästään asianmukaisia huomioita luettelemalla. Arviointimenetelmien erilaisten painotusalueiden lisäksi piste-erojen syitä voidaan selvittää myös vastausten laadullisella tarkastelulla, joten arviointimenetelmien välisten piste-erojen syiden käsittelyä jatketaan kolmannen tutkimuskysymyksen tulosten tarkastelussa (ks. alaluku 5.3).

Tuloksia tarkasteltaessa on kuitenkin muistettava arvioinnin kontrolloiva funktio ja huomioitava, että arviointi on voinut vaikuttaa vastaajien opiskeluun takaistusvaikutuksen kautta

(Alqahtani 2021; Nieminen 2019). On siis mahdollista, että erilaiset arviointiohjeet voivat vaikuttaa eri tavoin opiskelijoiden osaamiseen sekä ylioppilaskokeissa käytettävään vastaustapaan. Norton (2004) toteaa, että opiskelijat voivat käyttää arviointikriteerejä strategisesti saadakseen hyvän arvosanan, jolloin arviointikriteerit ohjaavat oppimista ja suorituksia kriteerien osoittamaan suuntaan. Tutkimuksen tuloksia tarkasteltaessa on siis huomioitava, että sekä opiskelijat että heidän opettajansa ovat tienneet, millainen rakenne ylioppilaskokeen arviointiohjeissa tulee todennäköisesti olemaan. Mikäli tarkastellun ylioppilaskokeen arvioinnissa olisi käytetty jo etukäteen esiteltyä arviointimatriisia, on mahdollista, että ylioppilaskokelaat olisivat pyrkineet laatimaan vastauksensa enemmän matriisin kriteerejä vastaaviksi. Siksi tutkimustulos ei todennukaisesti kerro vastaajien osaamistasosta matriisin kuvaamien taitojen osalta, vaan tulos kertoo ainoastaan siitä, miten matriisi vaikuttaisi nykymuotoisten vastausten pisteisiin.

Piste-erojen syiden lisäksi on tärkeää pohtia, mitä seurauksia arviointikriteerien rakenteen muuttamisella voisi olla maantieteen opetukselle tai oppiaineen ylioppilaskokeiden arvioinnille. Mikäli matriisiarvioinnilla tuotetut pisteet olisivat tämän tutkimuksen tulosten tavoin keskimäärin matalampia kuin nykyisillä arviointiohjeilla laaditut pisteet, on mahdollista, että maantieteen ylioppilaskoearvosanojen pisterajat laskisivat arvosanojen perustuessa suhteelliseen arviointiin (ks. Ouakrim-Soivio 2015). Toisaalta on mahdollista, että arviointiohjeiden muuttuminen saisi aikaan muutoksen myös maantieteen opiskelussa sekä ylioppilaskoetehtävien vastauksissa takaistusvaikutuksen myötä. Kun arviointi vaikuttaa oppimiseen ja opettamiseen hyödyllisesti, puhutaan positiivisesta takaistusvaikutuksesta (Alqahtani 2021). Johnson ja Shaw (2018) toteavat, että mitä paremmin koe edustaa osaamistavoitteita ja tiedollista osaamista, sitä hyödyllisempi takaistusvaikutus kokeella on. Jos siis maantieteen ylioppilaskokeessa arvioitaisiin opetussuunnitelman mukaisia osaamistavoitteita aikaisempaa kattavammin, vaikutukset opiskelijoiden osaamiseen voisivat takaistusvaikutuksen myötä olla positiiviset. Arviointikriteerien muuttamisella voitaisiin näin arvioinnin kontrollointifunktion mukaisesti ohjata oppimisen kohteita ja tapoja. Arviointimenetelmän muuttumisen seurauksia ei kuitenkaan voida tietää etukäteen, joten mikäli maantieteen arvioinnissa siirryttäisiin käyttämään arviointimatriisia, olisi mielenkiintoista selvittää muutoksen vaikutuksia maantieteen opiskeluun ja opettamiseen sekä lukiolaisten osaamistasoon.

## 5.2 Arviointimatriisilla toteutetun arvioinnin yhdenmukaisuus vaikuttaa lupavalta

Tutkimuksen tulosten mukaan arviointimatriisin avulla suoritettu arviointi oli varsin yhdenmukaista sekä arvioijien 1 ja 2 että arvioijien 1 ja 3 välillä. Sisäkorrelaatiokertoimen mukaan arvioinnin arvioitsijareliabiliteetti oli molempien arvioijaparien välillä hyväksyttävällä ja pääasiassa korkealla tasolla, mutta täydellistä yhdenmukaisuutta kuvaavat prosenttiluvut sekä annetuista pisteistä luodut hajontakaaviot paljastivat myös pisteiden välisiä eroavaisuuksia. Keskeinen tuloksista tehty havainto oli, että arvioinnin yhdenmukaisuus ei eronnut eri arviointimenetelmien välillä huomattavasti, ja arvioinnin yhdenmukaisuus kahden arvioijan välillä oli paikoitellen jopa hieman korkeampaa matriisiarvioinnissa kuin nykyisillä arviointiohjeilla toteutetussa arvioinnissa. Tämän tutkimustuloksen perusteella siis vaikuttaa siltä, että maantieteen ylioppilaskoevastaukset olisi mahdollista pisteittää kriteeriperustaisella arviointimatriisilla ilman, että arvioinnin yhdenmukaisuus heikkenisi huomattavasti nykyisten arviointiohjeiden yhdenmukaisuuteen verrattuna. Arvioinnin reliabelius on laadukkaan arvioinnin keskeinen ominaisuus (Ouakrim-Soivio 2015), minkä vuoksi tutkimustulos on huomionarvoinen, kun pohditaan arviointimatriisin soveltuvuutta maantieteen ylioppilaskokeiden arviointiin. Tutkimustulosta tarkasteltaessa on kuitenkin huomioitava, että matriisiarvioinnin toteuttivat samat henkilöt, jotka myös loivat matriisin: siksi on mahdollista, että heidän käsityksensä matriisin käytöstä on ollut parempi kuin vastausten arvioijilla normaalissa arviointitilanteessa olisi. Myös arvioijien taustat on syytä huomioida reliabiliteettikeskustelussa, sillä arvioijat 2 ja 3 ovat kokeneita ylioppilaskokeiden arvioijia. Arvioija 1 ei puolestaan ollut ennen tutkimusta arvioinut ylioppilaskoevastauksia, minkä vuoksi hän edusti tässä tutkimuksessa uutta opettajaa tai sensoria.

Kummallakin menetelmällä tuotetuissa pisteissä ilmeni yksittäisissä vastauksissa melko suuria eroja eri arvioijien välillä, vaikka molempien arviointimenetelmien arvioitsijareliabiliteetit olivat tämän tutkimuksen mukaan keskimäärin hyvät. Arvioijien väliset piste-erot selittyvät todennäköisesti laadulliseen kriteeriperustaiseen arviointiin liittyvällä tulkinnallisuudella (ks. Keurulainen 2013). Nykyiset arviointiohjeet eli Ylioppilastutkintolautakunnan laatimat hyvän vastauksen piirteet muistuttavat pitkälti Brookhartin (2013) kuvaamaa listamaista arviointitapaa, jossa arviointiohjeet sisältävät listan suoritukselta vaadittavista asioista. Brookhartin (2013) mukaan listamaiset arviointiohjeet ovat usein arviointimatriiseja helppokäyttöisempiä arviointivälineitä, sillä ne vaativat arvioijalta matriiseja vähemmän tulkintaa. Nykyiset listamaiset arviointiohjeet asettuvat laadullisen ja määrällisen kriteeriperustaisen arvioinnin väliin (ks. Keurulainen 2013). Toisaalta arvioitava vastaus kerryttää pisteitä yksittäisillä huomioilla,

joiden esiintymisen havaitseminen ei vaadi arvioijalta subjektiivista tulkintaa, mutta toisaalta osa pisteistä kertyy arvioijan tulkinnasta siitä, ovatko huomiot esimerkiksi kuvattu pintapuolisesti tai onko vastaus kypsä ja loogisesti jäsennelty (Hyvän vastauksen piirteet: FI – Maantiede 2023). Matriisiarviointi puolestaan perustuu täysin arvioijan tulkintaan siitä, mitkä laadulliset suoritustasojen kuvaukset vastaavat arvioitavaa suoritusta, jolloin se lukeutuu puhtaasti laadulliseen kriteeriperustaiseen arviointiin (ks. Keurulainen 2013). Laadulliseen kriteeriperustaiseen arviointiin liittyvä tulkinnallisuuden periaate kuvaa siis molempia tutkimuksessa tarkasteltuja arviointimenetelmiä ja samalla johtaa mahdollisiin piste-eroihin arvioijien välillä. Arvioijilla voi olla esimerkiksi erilaisia tulkintoja arvioitavien suoritusten laadusta tai eroavia odotuksia arvioitavien osaamistasosta, mikä voi heikentää arvioijien välistä yhdenmukaisuutta (Rinne 2024). Siksi tutkimuksen tuloksissa havaitut erot pisteiden välillä eivät ole yllättäviä.

Erityisesti viime vuosien arviointikeskustelussa on alettu korostaa myös ajattelun taitoja pelkän asiaosaamisen sijaan (Al-Salmani & Thacker 2021), ja Lukion opetussuunnitelman perusteis-sakin (2019) maantieteen arvioinnin keskeisiksi kohteiksi mainitaan muun muassa maantieteel-linen ajattelu sekä maantieteellisten taitojen osaaminen. Mikäli siis arvioinnissa on tarkoituk-sena huomioida tämänkaltaiset osa-alueet, on arviointiväline suunniteltava korkean validiteetin saavuttamiseksi niin, että maantieteellisten taitojen ja ajattelun hallintaa vaaditaan hyvän arvo-sanana saavuttamiseksi (ks. Brookhart 2010). Arviointimatriisit toimivat hyvänä esimerkkinä arviointivälineestä, jonka avulla ajattelun ja taitojen arvioiminen voidaan tehdä näkyväksi. Myös tässä tutkimuksessa luodussa arviointimatriisissa on pyritty laadullisten suoritustasojen kuvausten avulla huomioimaan monipuolisesti tietojen, taitojen ja ajattelun arviointi opetus-suunnitelman mukaisesti. Vaikka siis laadulliset arviointimenetelmät voivat tulkinnallisuutensa vuoksi johtaa arvioijien välisiin eroavaisuuksiin annetuissa pisteissä, niissä arvioinnin yhden-mukaisuuden sijaan korostuu validiteetin merkitys. Keurulaisten (2013) mukaan laadulliseen kriteeriperustaiseen arviointiin liittykin juuri validiteettityyppinen luotettavuus sen sijaan, että arvioinnissa painotettaisiin määrällisen arvioinnin tavoin reliabiliteettityyppistä luotettavuutta. Maantieteen ylioppilaskokeiden arvioiminen arviointimatriisin avulla voisikin mahdollistaa entistä monipuolisemman ja laaja-alaisemman osaamisen arvioinnin, vaikka se korostaisikin arviointin laadullisuutta aiempia arviointiohjeita enemmän.

Brookhart ja Chen (2015) toteavat katsausartikkelissaan, että arviointimatriisien avulla toteute-tun arvioinnin on mahdollista olla yhdenmukaista ja reliabiliteetiltaan luotettavaa, mutta arvi-oinnin yhdenmukaisuus ei aina toteudu matriisiarvioinnissakaan. Tässäkin tutkimuksessa voi-

daan vastaavasti todeta, että vaikka tutkimuksen tulokset matriisiarvioinnin arvioitsijareliabiliteetista vaikuttivat lupaavilta ylioppilaskokeiden arvioinnin kannalta, ne eivät yksiselitteisesti tarkoita, että matriisin avulla toteutettu arviointi olisi aina yhdenmukaisuudeltaan erinomaista. Tulos viittaa kuitenkin siihen, että eri arvioijien on mahdollista pisteittää ylioppilaskoevastauksia luodun arviointimatriisin avulla riittävän yhdenmukaisesti. Ylioppilaskokeiden kaltaiseen *high stakes* -arviointiin liitetään vaatimus korkeasta reliabiliteetista ja validiteetista sen merkittävien seurausten, kuten arvioitavan akateemiseen etenemiseen liittyvien vaikutusten, vuoksi (French ym. 2023). Mikäli siis maantieteen ylioppilaskokeiden arvioinnissa siirryttäisiin muiden oppiaineiden tavoin arviointimatriisin käyttöön, olisi tärkeää tarkastella, miten matriisiarvioinnin arvioitsijareliabiliteettia voitaisiin vahvistaa edelleen. Lähtökohta yhdenmukaiselle matriisiarvioinnille on laadukkaasti laadittu arviointimatriisi, mutta esimerkiksi arvioijien kouluttaminen on tyypillinen keino lisätä arvioinnin reliabiliteettia (Graham ym. 2012; Moskal & Leydens 2000). Lisäksi arviointimatriisissa kuvattujen suoritustasojen välisiä eroja voidaan selventää liittämällä arviointiohjeisiin matriisilla pisteitettyjä mallivastauksia, joita voidaan käyttää vertailukohtina arviointipäätöstä tehtäessä (Jonsson & Svingby 2007).

### **5.3 Matriisiarviointi laski erityisesti listamaisten ja arkikielisten vastausten pisteitä**

Vastausten laadullisessa analyysissä havaittiin joitain ominaisuuksia, jotka olivat tyypillisiä niille vastauksille, joiden pisteet erosivat arviointimenetelmien välillä toisistaan eniten. Yksi yhdistävä piirre lähes kaikissa niissä vastauksissa, jotka saivat matriisiarvioinnissa heikkomat pisteet kuin nykyisillä arviointiohjeilla toteutetussa arvioinnissa, oli vastauksissa esitettyjen huomioiden perustelujen puuttuminen tai niukkuus. Suurimmassa osassa näistä vastauksista huomiot esitettiin täysin ilman perusteluja tai vain puutteellisten perustelujen kera. Tämä teki vastauksista paikoitellen listamaisia, sillä ilman perusteluja vastaukset koostuivat lähinnä yksittäisten huomioiden luettelemisesta. Vastauksia yhdisti perustelujen puuttumisen lisäksi myös arkikielisyys sekä maantieteellisen tarkastelutavan puuttuminen: tämä havaittiin vastausten sijoittumisena heikkoihin laatuluokkiin käsitteet-kriteerin molemmissa osa-alueissa sekä maantieteen näkökulmien hallinta -osa-alueessa. Maantieteellisen tarkastelutavan puuttuminen näkyi vastauksissa esimerkiksi niin, että vastaukset saattoivat perustua ainoastaan yhden näkökulman käsittelemiseen, kun maantieteellinen ajattelu- ja tarkastelutapa tarkoittaa yleensä useiden monipuolisten näkökulmien tarkastelua (Bendl ym. 2024). Esimerkiksi joissain osatehtävän 7.1 vastauksissa keskityttiin vertailemaan liikennemuotoja ainoastaan ihmisen arkielämän näkö-

kulmasta liittämättä huomioita lainkaan maantieteelliseen teoriaan. Vastaavasti joissain osatehtävän 7.2 vastauksissa esitetyt huomiot liittyivät ainoastaan esimerkiksi siihen, mikä reitti olisi tietyistä paikoista lähteville ihmisille edullisin vaihtoehto, vaikka tehtävänannossa ohjattiin vertaamaan reittivaihtoehtoja paljon laajemmin. Näiden tulosten mukaan vaikuttaa siis siltä, että matriisiarviointi mahdollisti nykyistä arviointitapaa paremmin ja monipuolisemmin vastauksen heikkouksien huomioimisen.

Tutkimuksen aineistoon kuului vain seitsemän osatehtävän vastausta, joiden pisteet nousivat kolmella tai neljällä pisteellä arviointimatriisilla pisteitettynä, minkä vuoksi laatuluokkien jakaumien tarkastelu ei ole kovin tarkoituksenmukaista tai luotettavaa. Voidaan kuitenkin todeta, että vastaukset painoutuivat hyvin eri tavoin analyysirungon laatuluokkiin kuin ne vastaukset, joiden pisteet laskivat arviointimatriisilla arvioituina. Vastaukset sijoittuivat kolmeen parhaaseen laatuluokkaan kaikkien kriteerien jokaisessa osa-alueessa, minkä vuoksi ne vastaukset, joiden pisteet nousivat eniten arviointimatriisilla arvioituina, olivat laadukkaita kaikkien tarkasteltujen ominaisuuksien kannalta. Näin vastaukset olivat keskimäärin hyvin erilaisia kuin ne vastaukset, joiden pisteet laskivat eniten matriisiarvioinnissa. Tässä tutkimuksessa matriisiarviointi mahdollisti siis joissain tapauksissa tietyyntyyppisten vastausten palkitsemisen vielä nykyistä arviointitapaa paremmin.

Laadullisesti analysoitujen vastausten lukumäärä oli melko pieni eikä niiden jakautuminen laatuluokkiin ollut yksiselitteistä, minkä vuoksi tuloksista ei voida tehdä yleistettäviä johtopäätöksiä siitä, millaisten vastausten pisteisiin matriisiarviointi vaikuttaisi eniten. Voidaan kuitenkin todeta, että matriisiarviointi estää vastausta yltämästä tehtävän enimmäispistemäärään, mikäli yksikin sen osa-alueista on muita heikompi, sekä että arviointimatriisin avulla vastausten pisteityksessä on mahdollista huomioida nykyisiä arviointiohjeita monipuolisempia asioita. Tämän tutkimuksen tulosten perusteella vaikuttaa, että niissä maantieteen ylioppilaskoevastauksissa, joiden pisteitä matriisiarviointi laskee nykyiseen arviointimenetelmään verrattuna, esiintyy puutteita erityisesti huomioiden perusteluissa, käsitteiden käytössä, maantieteen näkökulmien hallinnassa sekä havaintojen kytkemisessä maantieteellisiin ilmiöihin ja prosesseihin. Nämä osa-alueet voivat myös liittyä toisiinsa, sillä esimerkiksi maantieteellisten näkökulmien asianmukainen käsittely huomioita perustelematta voi olla haastavaa.

Tutkimustulosten perusteella vaikuttaa siltä, että eri arviointimenetelmät palkitsivat arvioinnissa erilaisia asioita, mistä myös arviointimenetelmien väliset piste-erot todennäköisesti joh-

tuivat. Mikäli menetelmät arvioivat eri asioita, asiaa voidaan tarkastella validiteetin näkökulmasta eli pohtia, mittasiko arviointi sitä, mitä sen oli tarkoituskin mitata (Ouakrim-Soivio 2015). Reaaliaineiden kokeiden määräyksien ja ohjeiden (2024) mukaan arvioinnissa on kiinnitettävä huomiota esimerkiksi esitettyjen väitteiden perusteluun sekä eri näkökulmien esittämiseen. Koska matriisiarviointi laski tutkimustulosten mukaan erityisesti niiden vastausten pisteitä, joissa esiintyi puutteita juuri näissä osa-alueissa eli huomioiden perusteluissa sekä maantieteellisten näkökulmien huomioinnissa, vaikuttaa siltä, että ainakin näiden osa-alueiden arvioinnissa matriisiarviointi onnistui nykyisiä arviointiohjeita paremmin. Arvioinnin validiteettia on kuitenkin tärkeää tarkastella useista näkökulmista, eikä yhtäkään näkökulmaa voida sivuuttaa (Jonsson & Svingby 2007), minkä vuoksi tutkimustulosten kaltaisista yksittäisistä huomioista ei voida tehdä suoria päätelmiä arvioinnin validiteetista.

#### **5.4 Tutkimuksen luotettavuus ja jatkotutkimusmahdollisuudet**

Tutkimus oli luonteeltaan tapaustutkimus (Saaranen-Kauppinen & Puusniekka 2006). Tutkimuksessa tutkittiin rajattua kokonaisuutta monipuolisten menetelmien avulla ja sen tavoitteena oli lisätä ymmärrystä tutkittavasta ilmiöstä ilman, että sillä oltaisiin pyritty tuottamaan yleistettävää tietoa. Tutkimuksen tarkoituksena oli siis tuottaa kuvailevaa tietoa arviointimatriisin käytöstä maantieteen ylioppilaskokeiden arvioinnissa sekä matriisiarvioinnin eroista nykyiseen arviointimenetelmään verrattuna sen sijaan, että tutkimuksen tarkoituksena olisi ollut tuottaa esimerkiksi yleistettävää kuvausta eri arviointimenetelmien arvioitsijareliabiliteeteista. Maantieteen ylioppilaskokeiden arvioinnista arviointimatriisilla ei olla tehty aiempaa tutkimusta, joten tapaustutkimuksella oli suuri ymmärrystä syventävä arvo. Tapaustutkimuksen kuvailevasta luonteesta huolimatta on tärkeää tarkastella tekijöitä, jotka ovat voineet vaikuttaa tutkimuksen tulosten luotettavuuteen.

Tutkimusasetelma pyrittiin luomaan niin, että tulokset olisivat mahdollisimman luotettavia. Arvioijat pisteittivät eri arviointivaiheissa eri vastaukset, jotta aiemmassa arviointivaiheessa tehdyt tulkinnat vaikuttaisivat seuraavaan arviointivaiheeseen mahdollisimman vähän. Arvioijat eivät keskustelleet arvioinnista toistensa kanssa kummankaan arviointivaiheen aikana, jotta muiden tekemät tulkinnat tai havainnot eivät vaikuttaisi arviointiprosessiin. Arviointivaiheen lisäksi myös analyysivaiheessa tehtiin tulosten luotettavuutta parantavia päätöksiä: esimerkiksi arviointimenetelmien reliabiliteetteja kuvattiin useilla tilastollisilla tunnusluvulla ja kuvaajilla, sillä reliabiliteetin kuvaaminen vain yhdellä tunnusluvulla voi olla harhaanjohtavaa (Stemler

2004). Tutkimuksen luotettavuutta ja pätevyyttä voidaan kehittää kriittisellä ja arvioivalla tarkastelulla (Saaranen-Kauppinen & Puusniekka 2009), joten myös tässä alaluvussa esitetty pohdinta tukee tutkimuksen luotettavuutta.

Vaikka tutkimusasetelma pyrittiin luomaan mahdollisimman luotettavaksi, sitä kuitenkin heikentää useampi tekijä, joista yksi on tutkimusasetelmaan liittyvä subjektiivisuus. Molemmat tutkimuksessa tarkastellut arviointimenetelmät edustavat laadullista kriteeriperustaista arviointia, jolloin kumpaankin menetelmään liittyy tulkinnallisuuden periaate (Keurulainen 2013). Näin kaikki tutkimuksessa tilastollisesti tarkastellut pisteet perustuvat arvioijien tekemään tulkintaan eivätkä siten kuvaa vastauksissa esitettyä osaamista täysin objektiivisesti: tähän teemaan tutkimuksessa toteutettu reliabiliteetin tarkastelu toisaalta myös perustuu. Subjektiivisuus epävarmuustekijänä liittyy erityisesti kolmatta tutkimuskysymystä varten toteutettuun vastausten laadulliseen analysointiin, sillä laadullisessa tutkimuksessa tutkijan tekemät tulkinnat ja valinnat vaikuttavat merkittävästi tutkimuksen tuloksiin (Tuomi & Sarajärvi 2018). Toinen tutkija saattaisi siksi tehdä tässä tutkimuksessa tehdyn vastausten luokittelun täysin eri tavalla tai löytää vastausaineistosta esimerkiksi sellaisia asioita, joita tässä tutkimuksessa ei havaittu lainkaan. Vastaavasti tässä tutkimuksessa luotu arviointimatriisi tehtiin tutkimuksen tutkimusryhmän tulkintojen ja päätösten perusteella, jolloin toinen tutkimusryhmä voisi luoda samasta aineistosta hyvin erilaisen matriisin, joka puolestaan voisi vaikuttaa arviointiin hyvin eri tavalla. Lisäksi arviointimatriisista on huomioitava, että vaikka tutkimuksessa luotiin yleinen arviointimatriisi, jonka tarkoituksena on toimia erilaisten ylioppilaskoetehtävien vastausten arvioinnissa, sen luomisessa hyödynnettiin ainoastaan yhden tehtävän vastauksia, mikä saattoi vaikuttaa matriisin luomiseen epätoivotulla tavalla. Täten on mahdollista, että luotu arviointimatriisi soveltuu paremmin tutkimuksessa käytetyn tehtävän vastausten arviointiin kuin muiden maantieteen ylioppilaskoetehtävien vastausten arviointiin.

Tutkimustulosten luotettavuuteen vaikuttaa myös tutkimukseen osallistuneiden arvioijien määrä. Tutkimuksessa tarkasteltiin vastausten alkuperäisten sensoripisteiden lisäksi ainoastaan kolmen arvioijan tuottamia pisteitä, ja valtaosa vastauksista arvioitiin valituilla menetelmillä vain kaksi kertaa. Kattavamman käsityksen menetelmien arvioitsijareliabiliteeteista olisi voinut saada, jos arvioijia olisi ollut enemmän ja jokainen vastaus olisi arvioitu molemmilla menetelmillä useampia kertoja. Arvioijien määrän lisäksi myös tutkimusasetelma rajasi tuloksia. Tutkimusasetelman mukaisesti yksikään arvioija ei pisteittänyt samaa vastausta kahta kertaa, minkä vuoksi tutkimuksessa ei voitu verrata tietyn arvioijan eri arviointimenetelmillä antamia

rinnakkaisia pisteitä. Siksi ensimmäisessä tutkimuskysymyksessä selvitettyihin arviointimenetelmien välisiin piste-eroihin voivat vaikuttaa myös esimerkiksi satunnaiset ja arvioijaan liittyvät tekijät.

Arvioijien määrän lisäksi on tärkeää pohtia myös tutkimuksessa tarkastellun aineiston kokoa. Tutkimuksen vastausaineisto koostui 70 vastauksesta, mikä voisi olla tulosten yleistämistarkoitukseen liian pieni otos. Otos oli kuitenkin harkinnanvaraisesti luotu ja sen avulla pyrittiin selvittämään vastaus kuvailevaan tutkimukseen tulosten yleistämisen sijaan, jolloin otos vastaa hyvin tutkimuksen tarkoituksiin (ks. Vilka 2007). Koska arvioijat pisteittivät eri vastaukset eri arviointimenetelmillä, tilastollisia testauksia ei voitu tehdä 70 vastauksen pisteitä vertaillen. Nykyisten arviointiohjeiden reliabiliteetin selvittämiseksi testeissä verrattiin 40 vastauksen pisteitä toisiinsa ja matriisiarvioinnin reliabiliteetin selvittämiseksi verrattiin 30 vastauksen pisteitä toisiinsa. Tämänkokoiset otokset olivat riittävän suuret tutkimuksessa käytettyjen tilastollisten testien tekemiseen, mutta suurempi otos olisi parantanut tulosten luotettavuutta sekä yleistettävyyttä (Vilka 2007).

Kolmanteen tutkimuskysymykseen liittyvään laadulliseen tarkasteluun valittiin vain ne vastaukset, joissa eri arviointimenetelmillä tuotetut pisteet erosivat toisistaan eniten eli kolmella tai neljällä pisteellä. Tämän vuoksi laadullisesti tarkasteltujen vastausten määrä oli melko pieni suhteessa vastausten todelliseen määrään. Laadullisessa tutkimuksessa aineiston riittävyys voi määrittää saturaatio eli aineiston kylläntyminen, mikä tarkoittaa, että uudet tapaukset eivät tuo tutkimukseen enää uutta tietoa (Saaranen-Kauppinen & Puusniekka 2009). Tutkimuksessa tarkastelluista vastauksista tehtiin paljon samoja havaintoja ja vastaukset painottuivat selvästi tietyissä kriteereissä tiettyihin laatuluokkiin, mikä viittasi aineiston kylläntymiseen, mutta erityisesti pienen aineiston vuoksi on hyvin mahdollista, että suuremmasta aineistosta löytyisi uusia näkökulmia ja vastauksia tutkimuskysymykseen. Kylläntymisen voidaan kuitenkin ajatella liittyvän tulosten yleistettävyyteen (Saaranen-Kauppinen & Puusniekka 2009), ja koska tutkimuksen aineisto toisti itseään ja samat asiat toistuivat useassa vastauksessa, on odotettavissa, että aineistosta tehdyt havainnot vaikuttaisivat vastausten pisteisiin myös laajemmassa aineistossa ja erityyppisissä tehtävissä.

Tutkimuksessa tarkasteltujen vastausten lukumäärän lisäksi on tärkeää huomioida myös tehtävätyyppi, sillä tutkimukseen valitun koetehtävän jokaiseen osatehtävään kuului vastata essee-  
muotoisella vastauksella. Maantieteen ylioppilaskoe voi kuitenkin sisältää myös tehtäviä, joissa vaaditaan monipuolista tiedon prosessointia, kuten piirtämistä, kaavioiden laatimista tai kuvien

muokkaamista (Hyvän vastauksen piirteet: FI – Maantiede 2023), jolloin koko kokeen vastausaineisto ei välttämättä koostu ainoastaan esseemuotoisista vastauksista. Jotta arviointimatriisiin käytettävyyttä maantieteen ylioppilaskoevastausten arvioinnissa sekä matriisin vaikutusta pisteisiin voitaisiin arvioida kattavammin, olisi tärkeää tarkastella matriisin käyttöä myös erityyppisten vastausten arvioinnissa.

Tutkimuksia toistamalla ja esimerkiksi tutkimusasetelmaa kehittämällä voidaan nähdä, tarjoavatko uudet tutkimukset samansuuntaisia tuloksia, jolloin tutkimuksen luotettavuus ja informaatioarvo kasvavat (Saaranen-Kauppinen & Puusniekka 2009). Siksi tätä tutkimusta olisi mielenkiintoista jatkaa ja laajentaa niin, että edellä mainittuja tutkimuksen luotettavuutta heikentäviä seikkoja kehitettäisiin. Olisi mielekäästä tarkastella esimerkiksi laajemman arvioijajoukon, kuten kaikkien maantieteen ylioppilaskokeen sensoreiden, antamia pisteitä arviointimenetelmien todenmukaisempien arvioitsijareliabiliteettien selvittämiseksi. Lisäksi arviointimatriisin käytöstä saataisiin kattavampi näkemys, jos arvioitavia vastauksia olisi enemmän ja ne edustaisivat monipuolisesti erilaisia maantieteen ylioppilaskokeen tehtävätyyppejä. Tässä tutkimuksessa tarkasteltiin reliabiliteetin osa-alueista ainoastaan arvioijien välistä yhdenmukaisuutta, ja olisikin mielenkiintoista ja tärkeää tarkastella myös muita reliabiliteetin osa-alueita, kuten yhden arvioijan toteuttaman arvioinnin johdonmukaisuutta (ks. Stemler 2004). Laajemmassa tutkimuksessa myös vastausta kolmanteen tutkimuskysymykseen voitaisiin pyrkiä selvittämään esimerkiksi monipuolisemmilla tekstianalyysin menetelmillä sekä useamman tutkijan toteuttamana niin, että tutkijatriangulaatio vahvistaisi tulosten luotettavuutta (Saaranen-Kauppinen & Puusniekka 2009).

Aiheeseen liittyvässä jatkotutkimuksessa olisi mielenkiintoista tarkastella niiden arviointimatriisien reliabiliteettia, joita jo käytetään tiettyjen oppiaineiden ylioppilaskokeiden arvioinnissa. Mikäli maantieteen ylioppilaskokeiden arvioinnissa siirryttäisiin käyttämään arviointimatriisia, olisi kiinnostavaa selvittää, vaikuttaisiko arviointiohjeiden muuttuminen esimerkiksi opettajien tai opiskelijoiden toimintaan takaistusvaikutuksen kautta. Olisi kiinnostavaa tutkia, miten erilaisia asioita painottavat arviointiohjeet vaikuttaisivat ylioppilaskoevastausten rakenteeseen, sisältöön ja laatuun sekä opiskelijoiden osaamiseen, vai voitaisiinko arviointiohjeiden rakenteen muuttumista havaita lainkaan.

## 6 Johtopäätökset

Tutkimuksessa luotuun arviointimatriisiin perustuvat pisteet olivat keskimäärin nykyisiin arviointiohjeisiin perustuvia pisteitä matalampia. Erot eri arviointimenetelmillä annettujen pisteiden välillä olivat tilastollisesti merkitseviä.

Tutkimuksen tulosten mukaan eri arvioijat voivat pisteittää maantieteen ylioppilaskoevastauksia melko yhdenmukaisesti tutkimuksessa luodun kriteeriperustaisen arviointimatriisin avulla. Arviointimatriisilla toteutetun arvioinnin arvioitsijareliabiliteetti eli arvioijien välinen yhdenmukaisuus oli pääasiassa hyvällä tasolla, eikä matriisiarvioinnin yhdenmukaisuus eronnut huomattavasti nykyisiin arviointiohjeisiin perustuvan arvioinnin yhdenmukaisuudesta. Molemmissa tarkastelluissa arviointimenetelmissä huomattiin kuitenkin yksittäisiä melko suuriakin eroja eri arvioijien antamien pisteiden välillä.

Tutkimuksessa luodun arviointimatriisin käyttö laski eniten arkikielisten ja listamaisten vastausten pisteitä. Näissä vastauksissa esitetyt huomiot perusteltiin vain harvoin, mikä teki vastauksista pääosin listamaisia. Vastausten arkikielisyyteen vaikutti maantieteellisten käsitteiden ja näkökulmien puuttuminen sekä se, että vastauksissa esitettyjen huomioiden kytkeminen maantieteellisiin ilmiöihin ja prosesseihin oli keskimäärin hyvin heikkoa. Ne vastaukset, joiden pisteitä matriisiarviointi nosti eniten, olivat keskimäärin tasaisen laadukkaita jokaisen tarkastellun ominaisuuden osalta. Tutkimustulos voi viitata siihen, että nykyiset arviointiohjeet ja tutkimuksessa luotu arviointimatriisi palkitsivat arvioinnissa erilaisia asioita, vaikka ne perustuivatkin samoihin arvioinnin lähtökohtiin.

Tutkimus lisäsi ymmärrystä arviointimatriisin käytöstä maantieteen ylioppilaskoevastausten arvioinnissa, vaikka tapaustutkimuksen luonne ja rajaus rajoittavatkin tulosten yleistettävyyttä. Aihetta olisi mielekäästä tutkia monipuolisemmilla menetelmillä sekä laajemmalla aineistolla jatkotutkimusten parissa.

## Kiitokset

Kiitän tutkielmaani osallistuneita henkilöitä mielenkiintoisen tutkimusaiheen ehdottamisesta sekä asiantuntevasta ja innostavasta yhteistyöstä. Ilman teitä tätä tutkimusta ei olisi ollut mahdollista toteuttaa.

Kiitän myös Ylioppilastutkintolautakuntaa tutkimusluvan myöntämisestä sekä tutkimusaineiston toimittamisesta.

## Lähteet

- Abdallah, M., Taiwo, A. & Ukamaka, A. (2022) Influence of high-stakes on teaching styles of science in senior secondary schools in Nigeria. *African Educational Research Journal* 10(3) 291—301. DOI: 10.30918/AERJ.103.22.041
- Acosta, S., Garza, T., Hsu, H-Y., Goodson, P., Padrón, Y., Goltz, H. & Johnston, A. (2020) The accountability culture: a systematic review of high-stakes testing and English learners in the United States during No Child Left Behind. *Educational Psychology Review* 32(2) 327—352. DOI: 0.1007/s10648-019-09511-2
- Aijjawi, R., Dawson, P. & Boud, D. (2018). Conceptualizing evaluative judgement for sustainable assessment in higher education. Teoksessa Boud, D., Aijjawi, R., Dawson, P. & Tai, J (toim.) *Developing evaluative judgement in higher education: assessment for knowing and producing quality work*. Routledge, Lontoo.
- Al-Salmani, F. & Thacker, B. (2021) Rubric for assessing thinking skills in free-response exam problems. *Physical Review Physics Education Research* 17(1). DOI: 10.1103/PhysRevPhysEducRes.17.010135
- Alqahtani, F. (2021) The impact of language testing washback in promoting teaching and learning processes: a theoretical review. *English Language Testing* 14(7) 21—26. DOI: 10.5539/elt.v14n7p21
- Atjonen, P. (2007) *Hyvä, paha arviointi*. Kustannusosakeyhtiö Tammi, Helsinki.
- Bendl, T., Krajňáková, L, Marada, M. & Řezníčková, d. (2024) Geographical thinking in geography education: a systematic review. *International Research in Geographical and Environmental Education*. DOI: 10.1080/10382046.2024.2354097
- Biggs, J. B. & Collis, K. F. (1982) Evaluating the quality of learning: the SOLO taxonomy (Structure of the observed learning outcome). Academic Press, New York.
- Biggs, J. & Tang, C. (2011) *Teaching for quality learning at university*. 4 p. Open University Press, Maidenhead.
- Brookhart, S. M. (2018) Appropriate criteria: key to effective rubrics. *Frontiers in Education* 3(22). DOI: 10.3389/feduc.2018.00022
- Brookhart, S. M. (2010) How to assess high-order thinking skills in your classroom. ASDC, Alexandria.
- Brookhart, S. M. (2013) *How to create and use rubrics for formative assessment and grading*. The Association for Supervision and Curriculum Development, Washington D. C.

- Brookhart, S. M. & Chen, F. (2017) The quality and effectiveness of descriptive rubrics. *Educational Review* 67(3) 434—368. DOI: 10.1080/00131911.2014.929565
- Cantera, M. A., Arevalo, M-J., García-Marina, V. & Alves-Castro, M. (2021) A rubric to assess and improve technical writing in undergraduate engineering courses. *Education Sciences* 11(4) 146—166. DOI: 10.3390/educsci11040146
- Dawson, P. (2017) Assessment rubrics: towards clearer and more replicable design, research and practice. *Assessment & Evaluation in Higher Education* 42(3) 347—360. DOI: 10.1080/02602938.2015.1111294
- Dirkx, K., Joosten-ten Brinke, D., Arts, J. & van Diggelen, M. (2019) In-text and rubric-referenced feedback: differences in focus, level, and function. *Active Learning in Higher Education* 22(3) 1—13. DOI: 10.1177/1469787419855208
- Ellonen, N. & Kaakinen, M. (s.a.) Monitasomallit. Tietoarkisto. <https://www.fsd.tuni.fi/fi/palvelut/menetelmaopetus/kvanti/regressio/monitasomallit/> 29.1.2025
- French, S., Dickerson, A. & Mulder, R. A. (2023) A review of the benefits and drawbacks of high-stakes final examinations in higher education. *Higher Education* 88(3) 1—26. DOI: 10.1007/s10734-023-01148-z
- Gerritsen-van Leeuwenkamp, K., Joosten-ten Brinke, D. & Kester, L. (2017) Assessment quality in tertiary education: an integrative literature review. *Studies in Educational Evaluation* 55(6) 94—116. DOI: 10.1016/j.stueduc.2017.08.001
- Graham, M., Milanowski, A. & Miller, J. (2012) *Measuring and promoting inter-rater agreement of teacher and principal performance ratings*. Center for Educator Compensation Reform.
- Hadibarata, T., Hidayat, T. & Kwabena, J. (2024) The use of scoring rubrics in university. *Review* 3(1) 1—12. DOI: 10.53623/apga.v3i1.312
- Hafner, J. C. & P. M. Hafner (2003) Quantitative analysis of the rubric as an assessment tool: an empirical study of student peer-group rating. *International Journal of Science Education* 25(12) 1509—1528. DOI: 10.1080/0950069022000038268
- Hidayati, N., Suryanti, S., Rahmayumita, R. & Aisya, S. (2024) Development of critical thinking skills instruments: cases for essay tests. *Jurnal Kependidikan: Jurnal Hasil Penelitian dan Kajian Kepustakaan di Bidang Pendidikan, Pengajaran dan Pembelajaran* 10(1) 77-88.
- Hypoteesien testaus – SPSS-harjoitus 1 (2014) KvantiMOTV. 8.4.2014. <https://www.fsd.tuni.fi/menetelmaopetus/hypoteesi/harjoitus1.html> 17.2.2025

- Hyvän vastauksen piirteet: FI – Filosofia (2024) Ylioppilastutkintolautakunta. 27.3.2024.  
[https://tiedostot.ylioppilastutkinto.fi/kokeet/2024-03-27\\_FF\\_fi/grading-instructions.html](https://tiedostot.ylioppilastutkinto.fi/kokeet/2024-03-27_FF_fi/grading-instructions.html) 7.6.2024
- Hyvän vastauksen piirteet: FI – Maantiede (2023) Ylioppilastutkintolautakunta. 21.9.2023  
[https://tiedostot.ylioppilastutkinto.fi/kokeet/2023-09-21\\_GE\\_fi/grading-instructions.html](https://tiedostot.ylioppilastutkinto.fi/kokeet/2023-09-21_GE_fi/grading-instructions.html) 4.2.2025
- Hyvän vastauksen piirteet: FI – Psykologia (2023) Ylioppilastutkintolautakunta. 27.3.2024.  
[https://tiedostot.ylioppilastutkinto.fi/kokeet/2024-03-27\\_PS\\_fi/grading-instructions.html](https://tiedostot.ylioppilastutkinto.fi/kokeet/2024-03-27_PS_fi/grading-instructions.html) 7.6.2024
- Jakku-Sihvonen, R. (2013) Oppimistulosten arviointijärjestelmistä ja niiden kehittämishaasteista. Teoksessa Räisänen, A. (toim.) *Oppimisen arvioinnin kontekstit ja käytännöt*. Raportit ja selvitykset 2013:3, Opetushallitus.
- Johnson, M. & Shaw, S. (2019) What is computer-based testing washback, how can it be evaluated and how can this support practitioner research? *Journal of Further and Higher Education* 43(9) 1255—1206. DOI: 10.1080/0309877X.2018.1471127
- Jonsson, A. & Svingby, G. (2007) The use of scoring rubrics: reliability, validity and educational consequences. *Educational Research Review* 2(2) 130—144. DOI: 10.1016/j.edurev.2007.05.002
- Kaakinen, M. & Ellonen, N. (s.a.) Regressioanalyysi. Tietoarkisto.  
<https://www.fsd.tuni.fi/fi/palvelut/menetelmaopetus/kvanti/regressio/analyysi/>  
 18.2.2025
- Kestilä-Kekkonen, E. (s.a.) Kovarianssi ja korrelaatio. Tietoarkisto.  
<https://www.fsd.tuni.fi/fi/palvelut/menetelmaopetus/kvanti/korrelaatio/korrelaatio/>  
 18.2.2025
- Keurulainen, H. (2013) Pelisääntöjä arviointipäätösten tekemistä varten. Teoksessa Räisänen, A. (toim.) *Oppimisen arvioinnin kontekstit ja käytännöt*. Raportit ja selvitykset 2013:3, Opetushallitus.
- Kolaree, B. & Ninčević, M. (2022) Comparison of two exam evaluation methods for objectivity. *International Conference on Education in Mathematics, Science and Technology* 158—168.
- Laki ylioppilastutkinnosta 502/2019. Annettu Helsingissä 12.4.2019.
- Lok, B., McNaught, C. & Young, K. (2016) Criterion-referenced and norm-referenced assessments: compatibility and complementarity. *Assessment & Evaluation in Higher Education* 41(3) 450—465. DOI: 10.1080/02602938.2015.1022136

- Lukion opetussuunnitelman perusteet 2019*. Määräykset ja ohjeet 2019:2a. Opetushallitus, Helsinki. [https://www.oph.fi/sites/default/files/documents/lukion\\_opetussuunnitelman\\_perusteet\\_2019.pdf](https://www.oph.fi/sites/default/files/documents/lukion_opetussuunnitelman_perusteet_2019.pdf) 31.5.2024
- Luostarinen, A. & Nieminen, J. H. (2019) *Arvioinnin käsikirja*. PS-kustannus, Jyväskylä.
- Luostarinen, A. & Ouakrim-Soivio, N. (2019) Arvioinnin erilaiset tehtävät. Teoksessa Luostarinen, A. & Nieminen, J. H. (toim.) *Arvioinnin käsikirja*. PS-kustannus, Jyväskylä.
- Martin-Kniep, G. O. (2000) *Becoming a better teacher: eight innovations that work*. The Association for Supervision and Curriculum Development, Washington D. C.
- Metsämuuronen, J. (2009) *Tutkimuksen tekemisen perusteet ihmistieteissä*. Gummerus Kirjapaino Oy, Jyväskylä.
- Moskal, B. M. & Leydens, J. A. (2000) Scoring rubric development: validity and reliability. *Practical Assessment, Research, and Evaluation* 7(10). DOI:10.7275/q7rm-gg74
- Neil, D. T., Wadley, D. A. & Phinn, S. R. (2010) A generic framework for criterion-referenced assessment of undergraduate essays. *Journal of Geography in Higher Education* 23(3) 303—325. DOI: 10.1080/03098269985263
- Nieminen, J. H. (2019) Arviointikulttuuri. Teoksessa Luostarinen, A. & Nieminen, J. H. (toim.) *Arvioinnin käsikirja*. PS-kustannus, Jyväskylä.
- Nitko, A. J. & Brookhart, S. (2011) *Educational assessment of students*. 6. p. Pearson, Boston.
- Norton, L. (2004) Using assessment criteria as learning criteria: a case study in psychology. *Assessment & Evaluation in Higher Education* 29(6) 687—702. DOI: 10.1080/0260293042000227236
- Ositettu otanta (s.a.) Tilastokeskus. [https://stat.fi/meta/kas/ositettu\\_otanta.html](https://stat.fi/meta/kas/ositettu_otanta.html) 3.6.2024
- Otos ja otantamenetelmät (2003) KvantiMOTV. <https://www.fsd.tuni.fi/menetelmaopetus/otos/otantamenetelmat.html> 2.6.2024
- Ouakrim-Soivio, N. (2015) *Oppimisen ja osaamisen arviointi*. Helsinki, Kustannusosakeyhtiö Otava.
- Panadero, E. & Jonsson, A. (2013) The use of scoring rubrics for formative assessment purposes revisited: a review. *Educational Research Review* 9 129—144. DOI: 10.1016/j.edurev.2013.01.002
- Panadero, E. & Jonsson, A. (2020) A critical review of the arguments against the use of rubrics. *Educational Research Review* 30(1). DOI: 10.1016/j.edurev.2020.100329

- Panadero, E., Jonsson, A., Pinedo, L. & Fernández-Castilla, B. (2023) Effects of rubrics on academic performance, self-regulated learning, and self-efficacy: a meta-analytic review. *Educational Psychology Review* 35(113). DOI: 10.1007/s10648-023-09823-4
- Pisterajat (s.a.) Ylioppilastutkintolautakunta. <https://www.ylioppilastutkinto.fi/fi/tutkinnon-suorittaminen/pisterajat> 13.8.2024
- Popham, W. J. (1997) What's wrong – and what's right – with rubrics. *Educational Leadership* 55(2) 72—75.
- Rahman, A., Seraj, P. M. I., Hasan, M. K., Namaziandost, E. & Tilwani, S. A. (2021) Wash-back of assessment on English teaching-learning practice at secondary schools. *Language Testing in Asia* 11(12) 1—23. DOI: 10.1186/s40468-021-00129-2
- Reaaliaineiden kokeiden määräykset ja ohjeet (2024) Ylioppilastutkintolautakunta 23.1.2024 <https://www.ylioppilastutkinto.fi/fi/tutkinnon-toimeenpano/maaraykset-ja-ohjeet/koe-kohtaiset-maaraykset-ja-ohjeet/reaaliaineiden> 6.6.2024
- Reddy, Y. M. & Andrade, H. (2010) A review of rubric use in higher education. *Assessment & Evaluation in Higher Education* 35(4) 435—448.  
DOI: 10.1080/02602930902862859
- Rind, I. A. & Mari, M. A. (2019) Analysing the impact of external examination on teaching and learning of English at the secondary level education. *Cogent Education* 6(1) 1—14. DOI: 10.1080/2331186X.2019.1574947
- Rinne, I. (2024) Same grade for different reasons, different grades for the same reason? *Assessment & Evaluation in Higher Education* 49(2) 220—232. DOI: 10.1080/02602938.2023.2203883
- Reynders, G., Lantz, J., Ruder, S. M., Stanford, C. L. & Cole, R. S. (2020) Rubrics to assess critical thinking and information processing in undergraduate STEM courses. *International Journal of STEM Education* 7(9). DOI: 10.1186/s40594-020-00208-5
- Räkköläinen, M. (2013) Kontrollista luottamukseen. Teoksessa Räisänen, A. (toim.) *Oppimisen arvioinnin kontekstit ja käytännöt*. Raportit ja selvitykset 2013:3, Opetushallitus.
- Saaranen-Kauppinen, A. & Puusniekka, A. (2009) Menetelmäopetuksen tietovaranto Kvali-MOTV: Kvalitatiivisten menetelmien verkko-oppikirja. Yhteiskuntatieteellisen tietoariston julkaisuja 2009. <https://www.fsd.tuni.fi/fi/tietoarkisto/julkaisut/kvalimotv.pdf> 26.2.2025
- Salamanca, S. L. C., Parra-Martínez, A., Chang, A., Maeda, Y. & Traynor, A. (2024) The effect of scoring rubrics use on self-efficacy and self-regulation. *Educational Psychology Review* 36(3). DOI: 10.1007/s10648-024-09906-w

- Sawaki, Y. (2016) Norm-referenced vs. criterion-referenced approach to assessment. Teoksessa Tsagari, D. & Banerjee, J. (toim.) *Handbook of second language assessment*, 45—60. De Gruyter, Berliini.
- Stemler, S. E. (2004) A comparison of consensus, consistency, and measurement approaches to estimating interrater reliability. *Practical assessment, research & evaluation* 9(4).
- Taanila, A. (2020) Korrelaatio. Akin menetelmäblogi. 23.5.2022 <https://tilastoapu.wordpress.com/korrelaatio/> 18.2.2025
- Ten Cate, O. & Regehr, G. (2019) The power of subjectivity in the assessment of medical trainees. *Academic Medicine* 94(3) 333—337. DOI: 10.1097/ACM.0000000000002495
- Tilastollinen päättely (s.a.) Tietoarkisto. <https://www.fsd.tuni.fi/fi/palvelut/menetelmaopetus/kvanti/paattely/paattely/> 13.1.2025
- Tuomi, J. & Sarajärvi, A. (2018) Laadullinen tutkimus ja sisällönanalyysi. Helsinki, Kustannusosakeyhtiö Tammi.
- Valtioneuvoston asetus lukiokoulutuksesta 810/2018. Annettu Helsingissä 4.10.2018.
- Valtioneuvoston asetus ylioppilastutkinnosta 612/2019. Annettu Helsingissä 9.5.2019.
- Vilkka, H. (2007) *Tutki ja mittaa: määrällisen tutkimuksen perusteet*. Kustannusosakeyhtiö Tammi, Helsinki.
- Wallace, M. P. & Ng, J. S. W. (2023) Fairness of classroom assessment approach: perceptions from EFL students and teachers. *English Teaching & Learning* 47(4) 529—548. DOI: 10.1007/s42321-022-00127-4
- Yleiset määräykset ja ohjeet (2024) Ylioppilastutkintolautakunta 2.7.2024 <https://www.ylioppilastutkinto.fi/fi/tutkinnon-toimeenpano/maaraykset-ja-ohjeet/yleiset-maaraykset-ja-ohjeet#chapter-1-2-1> 24.7.2024

## Liitteet

### Liite 1. Tehtävän 7 nykyiset arviointiohjeet eli hyvän vastauksen piirteet

7.5.2025 klo 12.03

Hyvän vastauksen piirteet: FI – Maantiede

#### 7. Kuusamon alueen ratahanke 30 p.

Tehtävässä arvioidaan kokelaan perustietoja liikennemuodoista sekä monipuolista yleis- ja teemakarttojen lukutaitoa. Lisäksi arvioidaan kykyä tarkastella infrastruktuurihankkeen vaikutuksia monipuolisesti ja maantieteellistä kirjoitustaitoa. Tehtävä edellyttää kokelaalta monitieteistä ja ympäristöön liittyvää laaja-alaista osaamista.

##### 7.1 Kuvaile raideliikenteen etuja ja haittoja verrattuna tieliikenteeseen. 6 p.

Hyvin perustellusta edusta tai haitasta tieliikenteeseen verraten 2 p., pintapuolisesta kuvauksesta tai vertailun puuttuessa 1 p. Täysiin pisteisiin vaaditaan sekä etujen että haittojen käsittelyä. Vertailuasetelman puuttuessa voi saada vain puolet saavutetusta pistemäärästä.

Esimerkkisisältöjä:

- Raideliikennettä voidaan pitää turvallisempänä kuin tieliikennettä, sillä onnettomuudet ovat harvinaisempia.
- Raideliikenne on energiatehokkaampaa ja saastuttaa vähemmän kuin tieliikenne niin henkilö- kuin tavaraliikenteessäkin.
- Raideliikenne on tehokkaampi liikennemuoto kuin tieliikenne, sillä se kattaa myös metro- ja raitiotieliikenteen, joka sopii lyhyillä matkoilla esimerkiksi kaupunkien lähiliikenteessä suurten ihmismäärien kuljettamiseen.
- Raideliikenne vähentää ruuhkia kaupungeissa ja on siten tieliikennettä nopeampi keskipitkillä matkoilla ja kaupunkiseutujen liikenteessä.
- Raideliikenne ei ole niin joustavaa kuin tieliikenne, sillä raideliikenteen aikataulut ja asemien sijainti rajoittavat matkustamista ja asemilta matkaa on jatkettava toisella tavalla.
- Raideliikenne on etenkin tavaraliikenteen osalta kustannustehokkaampi, sillä sen kuljetuskapasiteetti suhteessa kuluihin on alhaisempi kuin tieliikenteen.
- Tavaraliikenteessä junien lastaus ja purku onnistuvat vain osalla asemista, mikä hankaloittaa ja hidastaa tavaroiden liikkumista verrattuna tieliikenteeseen.
- Raideliikenne sopii tieliikennettä paremmin sellaisten raskaiden ja suurten lastien kuljettamiseen, joiden toimitusaika on pitkä.
- Raiteiden rakennuskustannukset ovat suuret verrattuna maanteihin, minkä vuoksi raideverkko on yleensä harva ja siten saavutettavuus paikoin heikko.

##### 7.2 Rautatieyhteyttä Kuusamoon on suunniteltu jo 1800-luvun lopulta lähtien.

Vuosikymmenien aikana hanke on saanut sekä kannatusta että vastustusta ja radasta on tehty erilaisia kannattavuuslaskelmia. Uuden radan rakentamiskustannukset ovat eräiden arvioiden mukaan noin viisi miljoonaa euroa kilometriä kohden. Valitse kartan 7.A reittivaihtoehdoista A–C mielestäsi järkevin.

Arvioi valitsemasi reittivaihtoehdon etuja ja haittoja karttojen 7.A–7.E perusteella vertaamalla sitä kahteen muuhun vaihtoehtoon. **12 p.**

Tehtävässä ei ole yhtä oikeaa reittivaihtoehtoa, vaan kaikki vaihtoehdot (A–C) hyväksytään.

Kypsästä ja loogisesti jäsennellystä vastauksesta 2 pistettä, valitun reittivaihtoehdon nimeämisestä 1 piste ja vertailien kirjoitetuista perusteluista 9 pistettä. Hyvin kuvatussa perustelusta muihin vaihtoehtoihin verraten 2 p., pintapuolisesta kuvauksesta 1 p. Täysiin pisteisiin vaaditaan vähintään kolmen kartan hyödyntämistä vastauksessa. Jotta tehtävästä saisi täydet pisteet, tulee kuvata valittua ratavaihtoehtoa sekä etujen että haittojen näkökulmasta. Vertailuasetelman puuttuessa voi saada vain puolet saavutetusta pistemäärästä.

Esimerkkisisältöjä:

- Reitin pituus ja rakentamiskustannukset: Reittivaihtoehto A Kemijärven kautta olisi lyhin ja pelkän linnuntie-etäisyyden perusteella kustannuksiltaan edullisin. Reittivaihtoehto B Oulusta olisi kallein, sillä se edellyttäisi pisimmän radan rakentamista. Reittivaihtoehto C Kontiomäeltä olisi lähes yhtä pitkä kuin Oulusta rakennettava rata, ja siten kustannukset olisivat lähes yhtä suuret. (Kartta 7.A.)
- Vesistöjen sijainti: Reitillä A Kemijärveltä Kuusamoon on eniten suuria järviolueita, mikä hankaloittaisi radan suunnittelua ja rakentamista. Muilla ratavaihtoehdoilla suuria järviolueita ei ole tai ne olisi helpompi kiertää (kartta 7.A).
- Korkeuserot: Reitillä B korkeuseroja on vähemmän kuin reitillä C. Sen perusteella reitin B rakentamiskustannukset voivat olla pienemmät kuin reitin C.
- Tavaraliikenteen määrä: Reittivaihtoehto C Kajaanin ja Kontiomäen kautta Kuusamoon olisi tavaraliikenteen kannalta kannattavin, sillä verrattuna muihin vaihtoehtoihin tavaraliikennettä kulkee eniten tätä kautta (kartta 7.B).
- Henkilöliikenteen määrä: Ouluun saapuvat kaikkein suurimmat henkilöliikenteen matkustajavirrat, minkä vuoksi reittivaihtoehdon B rakentaminen Oulun kautta palvelisi junamatkustajia parhaiten ja olisi sitenärkevin vaihtoehto. Reittivaihtoehdot A ja C eivät palvelisi kovin suuria matkustajamääriä eivätkä siten olisi järkeviä. (Kartta 7.B.)
- Luonnonsuojelualueet: Kaikki reittivaihtoehdot A–C sivuavat paikoin nykyisiä luonnonsuojelualueita ja olisivat siten haitallisia luonnolle ja ympäristölle. Kontiomäen kautta kulkevan vaihtoehdon C reitille osuu useita pieniä luonnonsuojelualueita, joiden huomioiminen radan suunnittelu- ja rakennusvaiheessa olisi haastavaa. Reittivaihtoehtojen A ja B kohdalle osuu muutamia suurempia luonnonsuojelualueita, joiden kiertäminen voisi olla helpompaa toteuttaa. (Kartta 7.C.)
- Väestötiheys: Väestötiheys on suurempi Oulun alueella kuin Itä-Lapissa ja Kainuussa, ja siten Oulun seudulla on enemmän mahdollisia matkustajia.

Tämä puoltaisi reittivaihtoehtoa B Oulun kautta. (Kartta 7.D.)

- Aluepolitiikka: Yhteydet joko Kemijärven (A) tai Kajaanin ja Kontiomäen (C) kautta Kuusamoon voisivat olla aluepoliittisesti kannatettavia vaihtoehtoja, koska ne voisivat lisätä Itä-Lapin ja Kainuun alueen houkuttelevuutta ja vähentää niiden muuttotappiota. Uusi rata voisi myös lisätä painetta rakentaa Itä-Lapin, Koillismaan ja Kainuun kattava rengasrata ja siten kehittää aluetta entisestään. (Kartta 7.D.)

Myös muihin tietoihin perustuvat olennaiset perustelut hyväksytään.

### 7.3 Pohdi uuden ratayhteyden vaikutuksia ihmisen toimintaan ja luontoon Kuusamon alueella pitkällä aikavälillä. 12 p.

Kypsästä ja loogisesti jäsennellystä vastauksesta 2 p. ja vaikutuksista yhteensä enintään 10 p.: vaikutuksista ihmisen toimintaan 4–6 p. ja vaikutuksista luontoon 4–6 p. Hyvin kuvatusta vaikutuksesta 2 p., pintapuolisesta kuvauksesta 1 p. Esimerkkisisältöjä:

Vaikutukset ihmisen toimintaan (4–6 p.):

- Liikkumisen mahdollisuudet monipuolistuvat, jolloin Kuusamon saavutettavuus muun Suomen näkökulmasta paranisi.
- Liikkumismahdollisuuksien monipuolistuminen voi vähentää yksityisautoilua ja siten pienentää porokolareiden ja muiden liikenneonnettomuuksien määrää.
- Rakennusvaiheessa ratahanke parantaisi alueen työllisyystilannetta, sillä radan rakentaminen vaatii paljon työvoimaa.
- Liikennemuotojen monipuolistuminen voi kasvattaa matkustajien ja turistien määrää (entisestään), mikä puolestaan voi parantaa matkailuelinkeinon mahdollisuuksia ja lisätä matkailutuloja alueella.
- Kuusamon-rata mahdollistaisi laajemman työssäkäyntialueen. Tämä lisäisi paikallisen väestön työskentelymahdollisuuksia, työvoiman liikkuvuutta sekä alueen vetovoimaa työssäkäynti- ja matkailualueena.
- Maantielikenteen kuljetukset vähenisivät, mikä voi vaikuttaa ammattikuljettajien työllisyyteen negatiivisesti.
- Lisääntynyt matkailijoiden määrä voidaan kokea häiritsevänä ja alueen elämysarvoa heikentävänä.
- Rautatien välittömässä läheisyydessä melu ja värinä voidaan kokea häiritsevänä ja alueen arvoa vähentävänä seikkana.

Vaikutukset luontoon (4–6 p.):

- Junaliikenteen lisääntymisen myötä yksityisautoilu ja lentomatkustaminen voivat vähentyä, ja sitä kautta liikkumisen hiilidioksidipäästöt pienenevät. Tämä auttaa hillitsemään ilmastonmuutosta.

7.5.2025 klo 12.03

Hyvän vastauksen piirteet: FI – Maantiede

- Rautatien välittömässä läheisyydessä melu ja värinä häiritsevät luonnonympäristöä ja eläimistöä.
- Rautatien rakentaminen voi vaikuttaa kielteisesti eläinten luontaiseen liikkumiseen alueella.
- Rautatien rakennusvaiheessa luonnonympäristö voi vahingoittua rata-alueen läheisyydessä.
- Rautatie saattaa edellyttää siltojen rakentamista ja maanpinnan muokkausta laajalti, mikä vaikuttaa merkittävästi luonnonmaisemaan, kallio- ja maaperään, vesistöihin ja siten geodiversiteettiin.
- Mahdollisten junaonnettomuuksien seuraukset luonnonympäristölle voivat olla suuren kuljetuskapasiteetin takia vakavat, etenkin jos tavarajunissa kuljetetaan vaarallisia aineita.
- Rautatien rakentaminen voi lisätä matkailua ja rakentamista alueella, mikä osaltaan saattaa vaikuttaa kielteisesti alueen ympäristöön.

## Liite 2. Vastausten sisällönanalyysin analyysirunko

Matriisin kriteeri	Kriteerin osa-alueet	Osa-alueen laatuluokkien kuvaukset (0, 5, 10, 15, 20)	Lukumäärä
Tehtävänannon noudattaminen	Sisältöjen osuvuus	ei vastaa tehtävänantoon; tehtävä ymmärretty väärin	
		osin tehtävänannon mukainen; heikosti rajattu, epäselvä tai harhaileva	
		pääosin tehtävänannon mukainen; pienissä määrin epäolennaisuuksia	
		pääosa vastauksen sisällöistä on olennaisia	
		käsittelee kattavasti ja harkitusti olennaiset asiat; ei tehtävänannon kannalta epäoleellisia asioita	
	Tehtävänannon mukainen tarkastelutapa	annettua tarkastelutapaa ei noudateta	
		ei täysin tehtävänannon mukainen/vastaus hyvin suppea	
		pääosin tehtävänannon mukainen	
		selkeä ymmärrys tehtävänannosta	
	Vastauksen esitystapa	erinomainen ymmärrys tehtävänannosta	
		annettua esitystapaa ei noudateta	
		ei täysin tehtävänannon mukainen	
pääosin tehtävänannon mukainen			
Sisällöt	Huomioiden määrä	selkeä ymmärrys esitystavasta	
		erinomainen ymmärrys esitystavasta	
		- (pääasiassa vain virheellistä sisältöä)	
		yksittäisiä oikein kuvattuja huomioita	
		muutamia laajemmin käsiteltyjä huomioita/useita suppeammin käsiteltyjä huomioita	
	Huomioiden oikeellisuus	muutamia syvällisesti käsiteltyjä huomioita/useita pinnallisemmin käsiteltyjä huomioita	
		tehtävänannon mukaisia huomioita kattavasti	
		pääasiassa vain virheellistä sisältöä	
		yksittäisiä oikein kuvattuja sisältöjä; jos useita huomioita, voi olla muutamia asiavirheitä	
		yksi asiavirhe tai jonkin verran epätasällisyyttä	
	Huomioiden perustelut	lievää epätasällisyyttä; ei varsinaisia virheitä	
		ei asiavirheitä tai epätasällisyyttä	
huomioita ei perusteltu			
osalle huomioista voidaan esittää niukat perustelut			
muutamalla huomiolla tyydyttävät perustelut/usealla huomiolla hatarat perustelut			
Maantieteellinen tarkastelu- ja jäsentelytapa	Maantieteen näkökulmien hallinta	muutamia syvällisesti ja virheettömästi perusteltuja/useita pinnallisesti perusteltuja huomioita	
		huomiot perusteltu osuvasti, virheettömästi ja tarkoituksenmukaisesti	
		ei olennaisia näkökulmia	
		joitain olennaisia näkökulmia; epätasapainoinen vastaus	
		jäsennellytään näkökulmien mukaisesti	
	jäsennellytään tehtävänantoon kuuluvien monipuolisten näkökulmien mukaisesti; näkökulmia liitetty toisiinsa		
	monipuoliset näkökulmat ja niiden väliset vuorovaikutussuhteet kuvattu; johdonmukainen jäsentely		
rakenne hajanainen ja sekava/vastaus hyvin suppea			
puutteellinen rakenne, eteneminen epäloogista			

	Vastauksen looginen eteneminen ja rakenne	pääosin selkeä rakenne; joitain pieniä rakenteellisia puutteita tai ei etene täysin loogisesti toimiva rakenne, etenee loogisesti ehea ja tasapainoinen rakenne	
Käsitteet	Täsmällinen käsitteiden käyttö	ei käytetä käsitteitä; arkikielinen vastaus	
		käsitteidenkäytössä selviä puutteita tai virheitä	
		käsitteitä käytetään; käytössä lieviä puutteita tai yksittäinen virhe	
		käytetään keskeisiä käsitteitä; ei varsinaisia virheitä	
		käsitteitä käytetään monipuolisesti, täsmällisesti ja huolellisesti	
	Havaintojen kytkeminen maantieteellisiin ilmiöihin tai prosesseihin käsitteellistämällä	ei kytketä lainkaan ilmiöihin tai prosesseihin saatetaan mainita jokin ilmiö tai prosessi viitataan pintapuolisesti ilmiöön tai prosessiin hyödynnetään maantieteellistä teoriaa; kytkeminen ilmiöihin ja prosesseihin kevyttä asiat kytketty hyvin ilmiöihin ja prosesseihin	
Viestinnän keinot	Suomenkielinen asiateksti	ilmaisu vaikeasti ymmärrettävää	
		osittain puhekielinen/kömpelö/paljon kielioppivirheitä	
		pääosin selkeää asiatekstiä; voi olla paikoitellen kirjoitusvirheitä	
		selkeää ja lähes virheetöntä asiatekstiä kirjoitusasu täysin virheetön	
	Paikannimistö (n=16)	paikannimet kirjoitettu kokonaan tai pääosin väärin	
		paikannimissä paikoin asia- tai kirjoitusvirheitä	
		paikannimissä pieniä kirjoitusvirheitä	
		paikannimet kirjoitettu pääosin oikein paikannimet kirjoitettu virheettömästi	