



UNIVERSITY OF TAMPERE

This document has been downloaded from
TamPub – The Institutional Repository of University of Tampere

 *Publisher's version*

The permanent address of the publication is
<http://urn.fi/URN:NBN:fi:uta-201601261130>

Author(s):	Sormunen, Eero
Title:	Kopioiva lähteistä kirjoittaminen lukion haasteena
Main work:	Identiteetistä informaatiolukutaitoon : tavoitteena itsenäinen ja yhteisöllinen oppija
Editor(s):	Ropo, Eero; Sormunen, Eero; Heinström, Jannica
Year:	2015
Pages:	200-219
ISBN:	978-951-44-9920-3
Publisher:	Tampere University Press
Discipline:	Computer and information sciences; Educational sciences; Media and communications
School /Other Unit:	School of Information Sciences
Item Type:	Article in Compiled Work
Language:	fi
URN:	URN:NBN:fi:uta-201601261130
Subject:	informaatiolukutaito; oppiminen; information literacy; learning

All material supplied via TamPub is protected by copyright and other intellectual property rights, and duplication or sale of all part of any of the repository collections is not permitted, except that material may be duplicated by you for your research use or educational purposes in electronic or print form. You must obtain permission for any other use. Electronic or print copies may not be offered, whether for sale or otherwise to anyone who is not an authorized user.

KOPIOIVA LÄHTEISTÄ KIRJOITTAMINEN LUKION HAASTEENA

Eero Sormunen

Johdanto

Oman tekstin kirjoittaminen on perinteinen menetelmä vahvistaa oppilaan oppimista. Lähteiden pohjalta kirjoittamisen tavoitteena on, että oppilas prosessoi lähteistä tulkitsemaansa informaatiota ja rakentaa sen pohjalta jäsenneilyn tiedollisen esityksen (Tynjälä, Mason, & Lonka, 2001). Informaatiolukutaidon opetuksessa kirjoittaminen perustuu pääasiassa oppilaiden itsensä hankkimaan kirjallisuuteen, jolloin tehtävän tavoitteena on oppia tiedonhakemista, lähteiden arviointia ja tulkintaa, argumentoivaa kirjoittamista ja viittaamiskäytäntöjä (Sundin, Francke, & Limberg, 2011). Kirjoittamistehtävät voivat saada hiukan eri muodon valitun tyylin mukaan. Vain opettajalle suunnattu kouluessee on luonteeltaan ja vaatimuksiltaan erilainen kuin esimerkiksi Wikipedia-artikkelina julkaistava teksti (vrt. Forte & Bruckman, 2009).

Netin kehityksen myötä *lähteiden kopiointi* on nostettu yleiseen keskusteluun koulunopetuksen ongelmana. Kirjoittamistehtävät eivät tuota toivottua oppimista, jos teksti ”googletetulta” nettisivulta siirretään mekaanisesti omaan tekstiin arvioimatta sen luotettavuutta

ja tulkitsematta sen merkitystä. Kopioivaa kirjoittamista on tutkittu lähinnä *lähteiden plagiointina* (esim. McGregor & Streitenberger, 2004; Williamson, McGregor, & Archibald, 2010), jolloin tarkastelu saa helposti moralisoivan sävyn. Oppimisen kannalta kopioivan kirjoittamisen olennaisin ongelma on kuitenkin se että oppija ei paneudu lähteiden kanssa työskentelyyn oppimista tuottavalla tavalla (ks. Alexandersson & Limberg, 2003).

Kopiointi ymmärretään tässä artikkelissa kirjoittamiseksi, jossa teksti siirretään lähteestä omaan tekstiin sanasta sanaan. Kopiointia pidetään hyväksyttävänä, kun teksti esitetään sitaattina lähde asianmukaisesti mainiten. Siteeraamisen tavoitteena on korostaa lähteen tapaa ilmaista jokin asia erityisen sattuvasti. Siteeraus on tehokeino jota yleensä käytetään harvakseltaan. Kopioinnista tulee plagiointia, kun käytetty lähde jätetään mainitsematta ja informaatio esitetään ”omissa nimissä”. Plagiointi ei aina ole kopiointia sillä käsitteellisemmällä tasolla tapahtuvat ideavarkaudet lasketaan myös plagioinniksi (McGregor & Williamson, 2005).

Tässä raportoitavan tutkimuksen tavoitteena oli kehittää systemaattinen menetelmä oppilaiden lähteiden käytön analysointiin tutkielma- ja esseetyyppisissä oppimistehtävissä. Menetelmää sovellettiin kahden lukioluokan ryhmätyönä tekemiin Wikipedia-artikkeleihin. Huomio kiinnitetään oppilaiden lähdekirjoittamisen tyyliin kuten kopiointiin, omasanaiseen kirjoittamiseen, tekstin tiivistämiseen ja lähteiden synteisiin.

Artikkeli on jäsennetty seuraavasti: Ensin esitellään aiempaa tutkimusta tavoitteena syventää käsitystä kopioinnista ja kehittyneemmistä kirjoittamisen muodoista. Toiseksi esitellään kehitetty lähteiden käytön analysointimenetelmä, menetelmän kokeilussa sovellettu tutkimusaineisto sekä tutkimuksen päätulokset. Artikkelin lopuksi keskustellaan tulosten pohjalta lukiolaisten lähdekirjoittamisen ja laajemminkin informaatiolukutaidon ohjaamisen haasteista.

Kopiointia vai plagiointia?

Aiempi tutkimus on löytänyt yhteyden kopioivan kirjoittamisen ja pintaoppimisen välillä. Limberg (1999) havaitsi että heikommin edistyvillä lukiolaisilla oli taipumusta faktasuuntautuneeseen tiedonhakuun, mikä johti helposti kopioivaan kirjoittamiseen. Nämä oppilaat eivät olleet kiinnostuneita aiheen tutkivasta tarkastelusta. He näkivät lähteistä kirjoittamisen hakuharjoituksena, jossa ”oikeat vastaukset” kerätään lähteistä ja siirretään omaan tekstiin (Alexandersson & Limberg, 2003). Limbergin tutkimusryhmä (2008) toteaa monien Ruotsin yläkouluissa ja lukioissa tekemiensä tutkimusten perusteella että lähteiden kanssa työskentely onnistuu vain, jos oppilaissa pystytään synnyttämään aito kiinnostus ymmärtää tarkasteltavaa asiaa (oppimistehtävän aihe). McGregorin ja Streitenbergerin (2004) tulokset tukevat tätä näkökulmaa: opettajan tiukempi lähdeviittausten kontrolli vähensi plagiointia mutta samalla lisäsi ”feikkiviitteitä”. Ulkokohtainen kontrollin lisääminen ei syventänyt lähteiden informaation prosessointia.

Heinströmin (2005; 2006) tulokset täydentävät kuvaa tiedonhaun näkökulmasta. Hän totesi oppilaiden vahvan sitoutumisen oppimistehtävän aiheeseen johtavan tyypillisesti strategiseen tai syväoppimiseen. Strategiset ja syväoppijat profiloituivat tiedonhaussa ”syväselkeltäjiksi”: he näkivät vaivaa löytääkseen hyviä lähteitä ja analysoidakseen niiden sisältöä. ”Surfarit” olivat tyypillisesti pintaoppijoita, jotka eivät panostaneet tiedonhakuun. Heinström (2006) rinnastaa heidät faktasuuntautuneisiin hakijoihin. Surfarit yksinkertaistavat informaatiolukutaidon hakutaidoksi eivätkä koe mielekkääksi harjoittaa lähteiden arvioinnin ja käytön valmiuksiaan.

Plagiarismin tutkijoiden mukaan kopiointia ja plagiointia esiintyy erityisesti oppilailla, jotka kiinnittävät oppimistehtävän suorituksessa paljon huomiota lopputuotteen muodollisiin piirteisiin ja vähemmän itse opiskeluprosessiin. Sen sijaan prosessisuuntautuneet oppilaat käsittelevät lähteiden informaatiota syvällisemmin ja ylläpitivät kiinnostusta aiheeseen (McGregor & Streitenberger, 2004). Plagiointi

vaikuttaa luonnollisesti oppimistuloksiin. Esimerkiksi (McGregor & Williamson, 2005) totesivat että vähemmän plagioineet oppilaat muistivat myöhemmin järjestelyssä testissä enemmän opiskellusta aiheesta kuin plagiointiin taipuvaiset oppilaat.

Lähteistä kirjoittamiseen haasteet

Kasvatustieteen piirissä on muotoutunut tutkimusyhteisö, jonka kiinnostus kohdistuu lähdekirjoittamiseen oppimistehtävänä (reading-to-write tasks). Lähinnä empiirisessä koeasetelmassa tehdyissä tutkimuksissa on selvitetty kuinka oppijat selviävät tehtävistä, joissa on syntetisoitava tietoa useasta annetusta lähteestä (ks. Segev-Miller, 2004; Spivey, 1997). Yleensä koehenkilöt eivät ole hakeneet lähteitään itse vaan ovat saaneet ne tutkijalta. Tutkijat ovat keränneet kontrolloidussa tilanteessa havaintoaineistoa luku- ja kirjoitusprosessista ja analysoineet millä tavalla koehenkilöt ovat valinneet, tulkinneet ja muokanneet informaatiota.

Spivey (1997, 136, 144–145) määrittelee lähdekirjoittamisen tehtäväksi, jossa henkilö on saman aikaisesti kahdessa roolissa: lukijana hän rakentaa merkityksiä lähteestä ja kirjoittajana pyrkii luomaan merkityksiä omaan tekstiinsä. Lukeminen (ymmärtäminen) ja kirjoittaminen (luominen) eivät ole lineaarisia, peräkkäisiä prosesseja vaan sekoittuvat toisiinsa. Kirjoittaja ei lue vain toisten tekstejä vaan lukee myös omaansa. Toisaalta kirjoittamisen kognitiivinen prosessi alkaa jo lähteitä lukiessa oman tekstin ideointina ja suunnitteluna. Erityisesti edistynyt kirjoittaja hyödyntää lähteiden lisäksi myös omaa aiempaa tietämystään aiheesta ja käsitystään tekstin ajatellun kohderyhmän diskursiivisista käytännöistä.

Lähdekirjoittamisen empiiriset tutkimukset ovat osoittaneet että synteessin tekeminen useasta tekstistä on kognitiivisesti haastavampaa kuin tiivistelmän kirjoittaminen yksittäisestä lähteestä (Mateos & Solé, 2009). Tiivistelmän kirjoittamisessa voidaan käyttää samaa kerronnallista rakennetta kuin alkuperäisessä lähteessä. Usean lähteen

syntetisointi edellyttää yleensä että kirjoittaja joutuu kehittämään oman kerronnallisen juonen, johon eri tavoin rakennettujen lähteiden antama informaatio voidaan muokata (Segev-Miller, 2004).

Edellä kuvattu lähdekirjoittamisen tutkimus perustuu vahvaa semantiikkaa edustavien tekstianalyysimenetelmien käyttöön. Sekä lähteiden että kirjoitettujen tekstien lauseet jäsennetään propositioesityksiksi jolloin niiden semanttisia sisältöjä voidaan vertailla. Näillä menetelmillä ei ole tutkimusekonomisesti mahdollista analysoida kuin pieniä määriä lähteitä ja niistä tuotettuja tekstejä (ks. Spivey, 1997, 149-163). Lähestymistapa ei ole käyttökelpoinen, jos kukin oppilas käyttää tekstinsä pohjana vaihtelevaa määrää itse hankkimiaan lähteitä.

Uusi menetelmä

Lähteiden käytön analyysiin kehitetyn menetelmän perusideana on se, että kirjoitettuja tekstejä tarkastellaan virkkeittäin. Jokaisen virkkeen osalta etsitään vastausta kysymykseen: Mihin lähteeseen virke perustuu? Aineiston kokoaminen jakaantuu viiteen osatehtävään (yksityiskohtainen kuvaus, ks. Sormunen, Heinström, Romu, & Turunen, 2012):

1. Artikkelit ja niissä viitatu lähteet kootaan sähköisessä muodossa, jos mahdollista.
2. Artikkelit pilkkotaan yksittäisiksi virkkeiksi. Virke toimii analyysiyksikkönä lähteiden käyttöä selvitettäessä.
3. Ensimmäiseksi selvitetään yksinkertaisilla sanahauilla löytyykö kirjoittajan viittaamista lähteistä sanallisesti ja sisällöllisesti vastaavaa virkettä, tekstikappaletta tai näiden yhdistelmää. Jos virkkeen yhteys lähteeseen ei löydy sanahauilla, viitattuja lähteitä luetaan valikoiden informaatioisisällöllisen yhteyden tunnistamiseksi. Omasanaisten virkkeiden lähteet eivät löydy aina sanahauilla vaan vaativat lähteiden valikoivaa lukemista.

4. Jos virkkeen tuottamiseen sopivaa tekstiä ei löydy viitatuista lähteistä, siirrytään plagiarismianalyysiin. Siihen voidaan soveltaa kirjoittajien toimintaympäristössä relevantteja hakupalveluja (esim. Google) tai varsinaisia plagiarismityökaluja (esim. Turn-It-In).
5. Jos yhteys lähteeseen on edelleen epäselvä, voidaan jäljitystä laajentaa kirjoittajien toimintaympäristön lähteisiin kuten oppikirjoihin ja sähköisiin opiskeluaineistoihin.

Virkkeen kirjoittamisessa potentiaalisesti käytettyjen lähteiden jäljittäminen perustuu siis tekstisisältöjen mekaaniseen vertailuun (sanahaku) että sisällölliseen tulkintaan (lukeminen). Lopullinen ratkaisu virkkeen linkittämisestä lähteeseen tai lähteisiin perustuu aina sisällölliseen tulkintaan ja ulkoisten tekijöiden arviointiin (onko lähteen käyttö kirjoittajalle mahdollista ja todennäköistä).

Menetelmän uskottava käyttö edellyttää että tekstien analysoijalla täytyy olla riittävä aihepiirin, sen lähteiden ja lähteiden käytön tuntemus. Mekaanisesti informaatioisisältöä siirtäville virkkeille on helppo löytää vastaava lähde. Sen sijaan omasanainen kirjoittaminen yhdistellen informaatiota useammista lähteistä tekee yhteyksien tunnistamisen haasteelliseksi. Toisaalta näissäkin tapauksissa hyvin lähdekirjoittamisen käytännöt hallitseva kirjoittaja antaa lukijalle yksiselitteiset viittaukset käyttämiinsä lähteisiin ja osaa ilmaista selkeästi milloin teksti ei perustu lähteisiin vaan omaan ajatteluun. Virkkeet, joiden lähde jää tunnistamatta, muodostavat oman kategoriansa jatkoanalyysissä.

Tekstimuunnostyytit

Aineistonkeruu tuottaa listan lähdekirjoitettuja virkkeitä ja kuhunkin virkkeeseen liittyen tiedon todennäköisesti käytetystä lähteistä ja käytetyn informaation paikkatiedot lähteissä. Menetelmän seuraavassa vaiheessa tarvitaan luokitusjärjestelmä, jonka mukaan lähteen erilaisten käyttötapojen yleisyyttä voidaan eri kirjoittajilla määrittää. Teksti-

muunnostyyppejä kehitettiin ja kokeiltiin kahdessa tutkimuksessa (ks. Sormunen et al., 2012; Sormunen & Lehtiö, 2011). Lähtökohtana oli soveltaa aiemman tutkimuksen perustelemia olennaisia lähteiden käytön ulottuvuuksia. Neljä ulottuvuutta liittyy varsinaiseen lähteiden käyttöön, yksi tuotettuihin virkkeisiin ja yksi lähdetyyppihin. Ulottuvuudet ovat seuraavat:

1. Omasanaisuus
2. Synteettisyys
3. Uskottavuus
4. Viittaustarkkuus
5. Virketyyppi
6. Lähdetyyppi

Omasanaisuus mittaa sitä kuinka paljon kirjoittaja käyttää omia tapojaan ilmaista lähteestä tulkittu informaatio. Omasanaisuuden asteesta tunnistettiin viisi kategoriaa: *kopio*, *lähes kopio*, *muokattu*, *oma teksti* ja *tuntematon*. Leikkaa-liimaa -tekniikalla tuotettu virke tulkitaan *kopioksi* ja pienin mekaanisin muutoksin lähteestä siirretty virke *lähes kopioksi* (esim. yksittäisten sanojen poisto tai lisäys). *Muokattu* viittaa siihen, että kirjoittaja on tehnyt virkkeessä merkityssisältöön liittyviä valintoja, joita ei voi pitää mekaanisina. *Oma teksti* viittaa kirjoittajan kommentteihin tai huomioihin, jotka eivät perustu mihinkään lähteeseen vaan omaan ajatteluun. Virkkeen omasanaisuus saa arvon *tuntematon*, jos lähde on tuntematon eikä virkettä voi perustellusti tulkita omaa pohdintaa edustavaksi.

Synteettisyys mittaa kirjoittajan kykyä koota ja tiivistää tekstiinsä informaatiota yhdestä tai useammasta lähteestä. Virkkeen synteettisyyden asteelle määriteltiin kuusi kategoriaa, jotka ilmaisevat käytetyn lähdepohjan laajuuden: *virke*, *kappale*, *lähde*, *useita lähteitä*, *omaan tietoon yhdistely* ja *tuntematon*. *Virke*, *kappale* ja *lähde* ilmaisevat että virkkeen informaation sisältö perustuu yhden lähteen virkkeeseen, tekstikappaleeseen tai laajempaan kokonaisuuteen lähteessä. *Useita lähteitä* viittaa siihen että virkkeen sisältö perustuu informaation yh-

distelyyn kahdesta tai useammasta lähteestä. *Omaan tietoon yhdistely* tarkoittaa että omaa tietämystä on yhdistelty lähteestä tai lähteistä saatuun informaatioon.

Uskottavuudella tarkoitetaan sitä kuinka vakaasti kirjoitetun virkkeen argumentit rakentuvat lähteen varaan (kategoriat: *uskottava, monitulkintainen, virheellinen, heikko lähde, EOS*). Tavoitteena ei ole arvioida onko virkkeen informaatio ”totta” vaan sitä miten pätevästi kirjoittaja käyttää lähdettä. *Uskottava* virke tuo selkeästi esille lähteen tarjoaman informaation ja tarvittaessa myös lähteeseen tai informaation liittyvät varaukset. *Monitulkintaisen* virkkeen sisältö on epäselvä ja lukijalla on virkkeen perusteella suuri riski tulkita lähteen sisältö virheellisesti (esim. perusteeton yleistys). *Virheellinen* virke sisältää tulkinnan joka ei vastaa lähteen informaatioisisältöä. *Heikko lähde* viittaa tilanteeseen, jossa kirjoittaja ei ole huomioinut lähteen luotettavuuteen liittyvää epävarmuutta. Kategoriaan *EOS* sijoitetaan virkkeet, joissa lähteen käytön uskottavuutta on vaikea arvioida.

Viittaustarkkuus mittaa virkkeen ja lähdeviittauksen etäisyyttä kategorioilla *virke, kappale, artikkeli, puuttuu, hämy, tuntematon*. Virke, kappale ja artikkeli viittaavat siihen että lähteeseen on viitattu virkkeen, saman tekstikappaleen tai vasta artikkelin lähdeluettelon tasolla. Puuttuu -kategoriassa virkkeen lähdettä ei ole mainittu ollenkaan. Hämy viittaa siihen että virkkeen tai kappaleen tasolla on viitattu vain sellaiseen lähteeseen, johon virke ei voi perustua. Puuttuu- ja hämy-virkkeiden yleinen esiintyminen vihjaa plagiointiongelmasta.

Virketyyppi luokittaa kirjoittaman tuottaman virkkeen argumentatiivisen luonteen perusteella. Tämä voidaan tehdä esimerkiksi Hartin (1998, 89-93) esittämän väittämä-typologian mukaan. Hartin typologia sisältää muun muassa seuraavat kategoriat: *käsitelmäritelmä, faktaväittäjä, tulkintaväittäjä ja arvoväittäjä*. *Käsitelmäritelmät* luonnehtivat tekstissä viitattujen käsitteiden ominaisuuksia. *Faktaväittäjät* ilmaisevat asioiden tiloja, jotka voidaan periaatteessa osoittaa tosiksi ja epätosiksi (esimerkiksi ”Useimpien suomalaisten suuryritysten pääkonttorit sijaitsevat pääkaupunkiseudulla”). *Tulkintaväittäjät* esittävät tulkintoja faktoista (esimerkiksi ”Pääkaupunkiseudulla on

keskeinen rooli Suomen taloudessa”). *Arvoväittämät* esittävät arvioita asioiden hyödyllisyydestä tai muusta arvosta.

Lähdetyyppi laajentaa näkökulman kirjoittajan käyttämien lähteiden ominaisuuksiin. Näiltä osin voidaan hyödyntää erityisesti tiedonhankintatutkimuksessa vakiintuneita lähdetyyppien kategorioita. Tärkeä pääjako on painetut ja nettilähteet, joille molemmille voidaan johtaa alakategorioita tarvitulla tarkkuustasolla. Sopiva lähdetyyppi-typologia riippuu kirjoittamisen kontekstista ja alakategoriat voidaan johtaa myös tutkimusaineistosta.

Tutkimuskysymykset, menetelmät ja aineistot

Tässä alaluvussa kuvataan esimerkkitutkimus, jossa edellä kuvattua menetelmää sovellettiin autenttiseen lukiosta kerättyyn aineistoon. Tutkimuksen pääkysymys muotoiltiin näin: Kuinka lukiolaiset käyttävät lähteitä kirjoittaessaan ryhmätyönä Wikipedia-artikkelin? Pääkysymyksestä muotoiltiin eri näkökulmia tarkentavat alakysymykset:

- TK1. Missä määrin lukiolaisten ryhmätyönä kirjoittamat Wikipedia-artikkelit sisälsivät
 - a. kopiointia ja omasanaista tekstiä?
 - b. tekstin tiivistämistä yhdestä lähteestä ja synteisiä useammasta lähteestä?
 - c. lähteiden plagiointia?
 - d. uskottavaa argumentointia lähteiden pohjalta?
 - e. erityyppisiä virkkeitä (fakta- ja tulkintaväittämiä)?

- TK2. Onko painettujen ja nettilähteiden käytön välillä havaittavissa eroja, kun tarkastellaan
 - a. kopiointia ja omasanaista tekstiä?
 - b. tekstin tiivistämistä yhdestä lähteestä ja synteisiä useammasta lähteestä?
 - c. lähteiden plagiointia?

- d. uskottavaa argumentointia lähteiden pohjalta?
- e. erityyppisiä virkkeitä (fakta- ja tulkintaväittämiä)?

Ensimmäinen tutkimuskysymys demonstroi kuinka menetelmää voidaan käyttää oppilaiden lähdekirjoitusten eri aspektien analysoinnissa. Toisessa tutkimuskysymyksessä tarkastelu laajennetaan lähdeyyppien vaikutusten tarkasteluun: käytetäänkö nettilähteitä eri tavalla kuin painettuja lähteitä?

Kurssikuvaukset

Tutkimusaineistona käytetyt tekstit kerättiin kahdelta kahdeksan viikkoa kestäneeltä lukion kurssilta Tampereella keväällä 2011. Kirjoitustehtävää varten äidinkielen opettaja organisoi kurssinsa 30 oppilasta kymmeneen kolmen hengen ryhmään ja historian opettaja kurssinsa 28 oppilasta seitsemään 3–5 hengen ryhmään. Molemmilla kursseilla ryhmien tehtävänä oli laatia Wikipedian kirjoitussääntöjen mukainen artikkeli mutta vain äidinkielen kurssilla artikkelit julkaistiin Wikipedian suomenkielisessä versiossa. Historian artikkelit tallennettiin luokan erilliseen wiki-sovellukseen. Molemmilla kursseilla opettaja nimisi ryhmät ja ryhmät valitsivat aiheensa opettajan valmistelemalta listalta.

Äidinkielen kurssilla tehtävänä oli laatia artikkeli suomenkielisestä klassikkoromaanista. Oppilaat lukivat romaanin ja laativat siitä henkilökohtaisen teosanalyysin ennen ryhmäkirjoitustehtävän aloittamista. Ryhmien piti kirjoittaa artikkelissa kuvaus teoksesta, kirjailijasta ja teoksen vastaanotosta käyttäen jäsentävänä mallina Wikipediassa jo olevia teosartikkeleita. Historian kurssilla oppilasryhmät kirjoittivat artikkelit Suomen sotien väliseen historiaan (1918–39) liittyvistä aiheista. Aiheet olivat melko laajoja. Jokaisesta aiheesta opettaja oli listannut alateemoja jotta oppilaat saisivat käsityksen, mitä teemoja artikkelien tuli käsitellä.

Kurssien toteutus on kuvattu tarkemmin toisaalla tässä kirjassa Alamettälän ja Sormusen artikkelissa.

Aineiston kokoaminen

Tutkimukseen palkattiin tutkimusapulaisiksi kaksi opintojen loppuvaiheessa olevaa maisteriopiskelijaa. Toisen pääaine oli kotimainen kirjallisuus ja toisen historia. Rekrytoinnissa pyrittiin varmistamaan että tutkimusapulaisilla oli riittävä tietämys artikkelien aihepiiristä ja aihepiirin kannalta relevanteista lähteistä lähtökohtana oppilaiden kirjoitelmien arviointiin.

Ensimmäisessä vaiheessa koottiin oppilasryhmien kirjoittamat artikkelit, pilkkottiin ne yksittäisiksi virkkeiksi ja tallennettiin tekstin-käsittelyohjelman taulukoiksi. Kaikki kirjoittajien mainitsevat lähteet kerättiin ja tallennettiin sähköisesti. Painetuista lähteistä skannattiin vain relevantit sivut. Virkkeisiin, joiden lähde ei löytynyt kirjoittajien viittauksista, sovellettiin plagiarisminanalyysiä, jossa käytettiin systemaattisesti tehtyjä Google- ja Wikipedia-hakuja. Jos tämän jälkeenkin merkittävä osa artikkelin virkkeistä (> 10 %) oli vailla lähde-ehdokasta, tutkimusapulaiset etsivät yhteyksiä luokassa käytössä olleisiin oppikirjoihin ja aineistoihin, jotka olivat tulleet esille oppilaiden haastatteluissa.

Aineiston analyysi

Tutkimusapulaiset tutustuivat menetelmään koodaamalla ensin yhden artikkelin omalta aihealueeltaan. Koodaajien kokemusten pohjalta tutkimustiimi tarkensi koodien määrittelyjä ja molemmat koodaajat käsittelivät itsenäisesti yhden artikkelin molemmilta aihealueilta. Koodaustuloksen erojen pohjalta tehtiin viimeiset tarkistukset koodien määrittelyihin ja suoritettiin koodaajien yhdenmukaisuustesti uudella artikkeliparilla. Yhdenmukaisuus kohosi 91–95 prosentin tasolle.

Yhdenmukaisuustestien jälkeen tutkimusapulaiset koodasivat yksin loput oman aihealueensa artikkeleista. Yhteenveto analysoitujen virkkeiden ja tunnistettujen lähteiden määristä on esitetty oppilasryhmittäin taulukossa 1. Taulukosta ilmenee myös niiden virkkeiden määrä, joissa käytettyjä lähteitä ei onnistuttu jäljittämään, lähteiden jakautuminen painettuihin ja nettilähteisiin sekä kirjoittajien viittaamat mutta käyttämättä jääneet lähteet. Äidinkielen artikkeleista analysoitiin vähemmän virkkeitä sillä suoraan klassikkoromaanin tulkintaan perustuvat osat kuten juoni- ja henkilökuvaukset jätettiin analyysin ulkopuolelle. Taulukon kolmas sarake kertoo kuinka monen virkkeen osalta lähde jäi tuntemattomaksi. Historian artikkeleissa lähde jäi epäselväksi 6 prosentissa ja äidinkielessä 11 prosentissa virkkeitä. Aineistoa voidaan pitää lähdekirjoittamisen tutkimisen kannalta edustavana.

Tulokset tiivistettiin 2 x 2 ja 2 x 3 taulukoiksi ja tilastollinen merkitsevyys tarkistettiin χ^2 (khii-toiseen) -testillä (ks. Siegel & Castellan, 1988, 123-124). Merkitsevyyden kriteerinä käytettiin arvoa $p < 0,05$. Riippuvuuden vahvuuden mittana käytettiin Cramérin kerrointa (V), joka sopii käytettäväksi erikokoisilla tulostaulukoilla (Siegel & Castellan, 1988, 232). Riippuvuuden vahvuuden tulkinnassa sovellettiin Rean ja Parkerin (1997) kriteereitä, joiden mukaan arvot $V < 0,10$ edustavat käytännössä merkityksetöntä eroa. Tässä tutkimuksessa esiintyi enintään kohtalaisen vahvoja riippuvuuksia (moderate associations, $V = 0,20-0,40$).

Taulukko 1. Yhteenveto tutkimusaineistosta: artikkelien virkkeet ja niiden kirjoittamisessa käytettyjen lähteiden jakautumat.

Ryhmä	Virkkeet		Viitatus lähteet			Plagioituiden lähteet			Kaikki käytetyt lähteet			Viitatus käyttämättömät lähteet
	Analysoidut	Lähde tunte-maton	Netti	Painetut	Yhteensä	Netti	Painetut	Yhteensä	Netti	Painetut	Yhteensä	
Historian kurssi												
H1	116	8	11	4	15	4	1	5	15	5	20	1
H2	125	11	10	4	14		0	1	11	4	15	0
H3	68	3	12	3	15	10	0	10	22	3	25	0
H4	74	2	7	1	8	1	0	1	8	1	9	0
H5	143	12	12	4	16	3	1	4	15	5	20	0
H6	59	1	5	1	6	3	1	4	8	2	10	11
H7	116	6	10	5	15	1	1	2	11	6	17	2
Keskiano	100	6	10	3	13	3	1	4	13	4	17	2
Keskilähj.	33	4	3	2	4	3	1	3	5	2	6	4
Äidinkielen kurssi												
L1	46	0	6	3	9	0	0	0	6	3	9	2
L2	17	1	0	4	4	1	1	2	1	5	6	0
L3	18	1	1	3	4	0	0	0	1	3	4	0
L4	69	17	0	6	6	2	1	3	2	7	9	0
L5	36	0	4	6	10	1	0	1	5	6	11	1
L6	41	9	0	5	5	0	0	0	0	5	5	0
L7	48	5	0	4	4	2	0	2	2	4	6	1
L8	30	0	2	4	6	6	1	7	8	5	13	0
L9	23	3	0	3	3	0	0	0	0	3	3	0
L10	24	1	3	2	5	2	0	2	5	2	7	2
Keskiano	35	4	2	4	6	1	0	2	3	4	7	1
Keskilähj.	16	6	2	1	2	2	1	2	3	2	3	1

Tulokset

Tutkimuskysymys 1: Miten oppilaat kirjoittivat lähteiden pohjalta?

Ensimmäinen tutkimuskysymys hakee vastausta siihen, missä laajuudessa lukiolaiset muokkasivat lähteiden informaatiota, syntetisoivat sitä, viittasivat käyttämiinsä lähteisiin ja tekivät uskottavia tulkintoja kirjoittaessaan Wikipedia-artikkeleita. Samalla selvitetään oliko oppilasryhmien välillä havaittavissa eroja historian ja äidinkielen kurseilla. Empiirisen aineiston antamat vastaukset on tiivistetty taulukon 2 sarakkeisiin 3-5. Tilastollisesti merkitsevät erot luokkien välillä on esitetty **vahvennettuna**.

Omasanaisuus. Hiukan yli puolet (54%) oppilaiden lähteiden pohjalta kirjoittamista virkkeistä edusti omasanaista esitystapaa. Äidinkielen kurssin oppilaat olivat omasanaisempia (65%) kuin historian kurssin oppilaat (49%). Käytännössä ero luokkien välillä jäi vähäiseksi, jos sitä arvioidaan Cramérin kerroimella ($V=0,16$). Keskimäärin äidinkielen artikkeleissa 65 virkettä sadasta oli omasanaisia, kun vastaava luku historian artikkeleissa oli 49.

Toisin päin tulkittuna tulos kertoo että lähes puolet (46%) oppilaiden kirjoittamista virkkeistä oli lähteiden virkkeiden kopiota tai lähes kopioita. Molemmissa luokissa suoraan kopioitujen virkkeiden osuus oli 9 prosenttia, mutta historian ryhmissä harrastettiin selvästi enemmän virkkeiden teknistä ”tuunaamista”. 42 virkettä sadasta oli lähes kopioita (äidinkielessä 26 sadasta).

Synteettisyys. Useimmat kirjoitetut virkkeet (58%) perustuivat lähteen virkkeeseen tai tekstikappaleeseen. Noin kolmannes (35%) tiivistyi informaatiota useammasta lähteen tekstikappaleesta ja vähäinen osa virkkeistä (7%) syntetisoi sisältöä kahdesta tai useammasta lähteestä. Äidinkielen ryhmissä kirjoitettiin hiukan enemmän tiivistäen ja syntetisoiden kuin historian ryhmissä (47% vs. 39%). Ero on tilastollisesti merkitsevä mutta ero käytännössä vähäinen ($V=0,11$).

Viittaustarkkuus. Kolmessa virkkeessä viidestä (60%) käytettyyn lähteeseen viitattiin virkkeen tai tekstikappaleen yhteydessä. Kummankin kurssin osalta 16 prosentissa virkkeitä oikea viittaus puuttui kokonaan lähdeluettelosta eli kirjoittamista voidaan tulkita plagioinniksi. Historian oppilaat käyttivät jonkin verran yleisemmin tarkkoja viittauksia (65% vs. 49%). Ero luokkien välillä tarkoissa viittauksissa on tilastollisesti merkitsevä mutta käytännössä vähäinen ($V=0,17$).

Uskottavuus. Oppilaat muotoilivat virkkeet lähteiden pohjalta pääosin uskottavasti (85%) eikä kurssien välillä ollut tilastollisesti merkitsevää eroa. Lukuja tulkitessa kannattaa huomioida että uskottavuus säilyy korkealla tasolla, jos lähteistä kirjoittaminen on kopioivaa.

Virketyypit. Kolme viidestä virkkeestä (60%) edusti lähteistä poimittuja faktoja ja lopuista lähes kaikki (37%) edusti tulkintaväittämiä. Määritelmät ja arvoväittämät olivat harvinaisia. Historiassa artikkelit olivat huomattavasti enemmän faktapainotteisia (67% vs. 31%) ja äidinkielessä tulkintaväittämät olivat vastaavasti yleisempiä (52% vs. 31%). Ero luokkien välillä oli tilastollisesti merkitsevä ja käytännössäkin ero kohtalaisen suuri ($V=0,21$).

Tutkimuskysymys 2: Onko painettujen ja nettilähteiden käytössä eroja?

Vastaukset kysymykseen on tiivistetty taulukon 2 sarakkeisiin 6–8. Tästä tarkastelusta jätettiin pois 29 virkettä, jotka perustuivat sekä painettujen että nettilähteiden käyttöön ($n=916$ virkettä).

Omasanaisuus. Tulokset paljastavat tendenssin kopioida lauseita nettilähteistä (14% vs. 5%) ja omasanaisemman ilmaisutavan painettuja käytettäessä (57% vs. 48%). Lähdetyyppien välinen ero oli tilastollisesti merkitsevä. Käytännössä ero ei ole kuitenkaan liene silmiinpistävä ($V=0,15$).

Synteettisyys. Oppilaat tiivistivät ja syntetisoivat lähteiden informaatiota samaan tapaan lähdetyyppistä riippumatta. Kolme viidestä kirjoitetusta virkkeestä perustui yhteen virkkeeseen tai tekstikappaleeseen niin painettuja kuin nettilähteitäkin käytettäessä.

Viittaustarkkuus. Tulokset vahvistavat yleisen käsityksen plagioinnin liittymisestä erityisesti nettilähteiden käyttöön. Kirjoittajat

eivät olleet liittäneet viittausta 29 prosenttiin nettilähteiden pohjalta kirjoitetuista virkkeistä. Painettu lähde oli unohtunut mainita vain 5 prosentissa virkkeistä. Näin suuri ero on jo käytännössäkin helppo tunnistaa ($V=0.33$).

Lähdekirjoittamisen ominaisuudet		Luokka		Yhteensä (n=945)	Lähdetyyppi		Yhteensä (n=916)
		Historia (n= 653)	Äidinkieli (n= 292)		Painetut lähteet (n=486)	Netti- lähteet (n=430)	
Oma-	kopio	9 %	9 %	9 %	5 %	14 %	9 %
sanaisuus	lähes kopio	42 %	26 %	37 %	38 %	38 %	38 %
	muokattu	49 %	65 %	54 %	57 %	48 %	53 %
Yhteensä		100 %	100 %	100 %	100 %	100 %	100 %
Synteettisyys	virke-kappale	61 %	53 %	58 %	60 %	60 %	60 %
	tiivistelmä	34 %	37 %	35 %	36 %	37 %	36 %
	synteesi	5 %	10 %	7 %	4 %	3 %	4 %
Yhteensä		100,0 %	100,0 %	100,0 %	100 %	100 %	100 %
Viittaus- tarkkuus	virke-kappale	65 %	49 %	60 %	67 %	51 %	60 %
	artikkeli	19 %	34 %	24 %	28 %	20 %	24 %
	puuttuu-hämy	16 %	16 %	16 %	5 %	29 %	16 %
Yhteensä		100 %	100 %	100 %	100 %	100 %	100 %
Uskottavuus	uskottava	87 %	82 %	85 %	86 %	84 %	85 %
	monitulk.-virh.	13 %	18 %	15 %	14 %	16 %	15 %
Yhteensä		100 %	100 %	100 %	100 %	100 %	100 %
Virketyyppi	faktaväittäjä	67 %	45 %	60 %	55 %	65 %	60 %
	tulkintaväittäjä	31 %	52 %	37 %	42 %	33 %	38 %
	muu	2 %	3 %	3 %	3 %	2 %	3 %
Yhteensä		100 %	100 %	100 %	100 %	100 %	100 %

Uskottavuus. Tuotettujen virkkeiden uskottavuus oli samalla tasolla lähdeyydestä riippumatta.

Virketyyppi. Nettilähteiden pohjalta kirjoitettiin enemmän faktaväittämiä (65% vs. 55%) ja painettujen pohjalta vastaavasti enemmän tulkintaväittämiä (42% vs. 33%). Ero on käytännön kannalta pienehkö ($V=0.10$).

Keskustelua

Edellä esitellyt tulokset demonstroivat sitä, että kehitetty menetelmä mahdollistaa lähteiden käytön yksityiskohtaisen ja systemaattisen analyysin. Menetelmää on mahdollista kehittää eteenpäin esimerkiksi käyttämällä syvällisempiä kategorisointeja kirjoitettujen virkkeiden uskottavuuden ja virketyypin määrittelyssä. Tarkastelu on myös helppo nostaa virkkeistä tekstikappaleisiin jolloin päästään ehkä luontevammalle tasolle esimerkiksi tekstin synteettisyyden arvioinnissa.

Menetelmä on tarkoitettua ensisijaisesti tutkijan välineeksi. Se on liian työläs toimiakseen opettajan työkaluna lähdekirjoittamisen tuotosten arvioinnissa. Toki opettajakin voi soveltaa menetelmää kertaluonteisesti perehtyäkseen oppilaiden tapaan kirjoittaa lähteistä. Myöskin perinteisessä tutkielmien arvioinnissa voi tarkkailla lähteiden käyttöä menetelmän yhteydessä kehitetyillä kriteereillä (omasanaisuus, synteettisyys, uskottavuus ja viittaustarkkuus). Analyysi voi antaa vihjeitä oppilaiden kirjoitusprosessien ohjauksen kehittämiseen.

Esimerkkitutkimuksen tulokset kertovat lukiolaisten tavoista kirjoittaa lähteiden pohjalta ja miltä osin nämä tavat poikkesivat kahden eri oppiaineen ryhmätehtävissä. Äidinkielen ryhmissä oppilaat olivat omasanaisempia, tiivistivät ja syntetisoivat enemmän lähteiden informaatiota sekä kirjoittivat enemmän tulkintatyyppisiä virkkeitä kuin historian oppilaat. Vastaavasti historian artikkeleissa lähdeviittaukset olivat tarkempia eli enemmän virke- ja kappaletasolla kuin äidinkielen. Kysymystä, mistä luokkien väliset erot mahdollisesti johtuvat, käsitellään Alamettälän ja Sormusen artikkelissa toisaalla tässä kirjassa.

Kaksi lähdekirjoittamisen piirrettä näyttäytyi täsmälleen samanalaisena molemmilla kurseilla: sekä äidinkielen että historian kurseilla artikkelien virkkeistä oli 9 prosenttia lähteen tekstin täsmällisiä kopioita ja toisaalta 16 prosenttia virkkeistä oli jäänyt vaille lähdemerkintää eli täyttivät plagiaatin tunnusmerkit. Tulos voi johtua sattumasta, mutta toinenkin tulkinta on uskottava. 'Sopivassa laajuudessa' tehty lähdetekstin kopiointi ja viitteiden kirjaamatta jättäminen voi muodostua jaetuksi käytännöksi koulutyön arjessa. Molempia ”paheita” esiintyi oppiaineesta, oppimistehtävän erityispiirteistä tai annetusta ohjauksesta riippumatta. Muun muassa Limberg on korostanut koulun käytäntöjen (ns. koulukulttuurin) merkitystä oppilaiden työskentelytapojen selittäjänä (vrt. Limberg et al., 2008).

Toinen tärkeä havainto oli kopioimisen ja plagioinnin liittyminen erityisesti nettilähteiden käyttöön. Oppiainekurseihin liittyvät tekijät eivät voi selittää eroja kopioinnin ja plagioinnin määrässä, koska ne olivat täsmälleen samalla tasolla. Aineiston lähempi tarkastelu paljasti että suurin osa plagioiduista virkkeistä oli lähtöisin Wikipediasta. Oppilaat käyttivät Wikipediaa lähteenä, mutta eivät voineet kirjjata sitä lähteeksi, sillä se oli vastoin opettajan ohjeita. Kun Wikipediassa mahdollisesti mainittuja alkuperäisiä lähteitä on usein vaikea saada käyttöön tai se koetaan tarpeettomaksi lisätyöksi, plagioinnista tulee houkutteleva vaihtoehto. Jos kiinnijäämisen riski arvioidaan pieneksi yksittäisten virkkeiden osalta, ratkaisu on oppilaan kannalta perusteltavissa. Tulos on linjassa aiempien plagiointitutkimusten kanssa (vrt. McGregor & Williamson, 2005; Purdy, 2009)

Kopioinnin yleisyyttä nettilähteistä selittää varmaankin leikkaa-liimaa -kopioinnin tekninen helppous, jolloin kiusaus kopiointiin on suuri. Painettuja lähteitä käytettäessä tekstiä on luettava huolellisemmin jotta sen voisi edes kopioida. Positiivisesti katsoen, painetun lähteen huolellisempi lukeminen voi antaa myös ideoita kirjoittaa virke omasanaisesti. Äidinkielen artikkeleissa käytettiin suhteessa enemmän painettuja lähteitä, mikä saattaa selittää myös omasanaisten virkkeiden suuremman osuuden. Historian artikkeleissa käytettiin puolestaan

enemmän nettilähteitä, mikä saattaa selittää lähes kopioitujen (tek-
nisesti tuunattujen) virkkeiden suurempaa osuutta.

Oppilaat tiivistivät tai syntetisoivat lähteistä poimimiaan sisältöjä
yhtä vähän painettuja ja nettilähteitä käyttäessään. Toisin sanoen
informaation korkeamman tason prosessointi tapahtuu suunnilleen
samalla tavalla lähdetyyppistä riippumatta. Tämän vuoksi ei kannata
liikaa korostaa nettilähteisiin liittyviä lieveilmiöitä. Suurin opetuk-
sellinen haaste lienee se, että lukiolaisen lähdekirjoittamista dominoi
melko suoraviivainen tekstin siirtoa lähteestä omaan tekstiin. Kolme
viidestä kirjoitetusta virkkeestä perustui yhteen tai enintään 2-3 läh-
teen virkkeeseen.

Lähteet

- Alexandersson, M., & Limberg, L. (2003). Constructing meaning through
information artefacts. *New Review of Information Behaviour Research*,
4(1), 17–30.
- Forte, A., & Bruckman, A. (2009). Writing, citing, and participatory media:
wikis as learning environments in the high school classroom. *International
Journal of Learning and Media*, 1(4), 23–44.
- Hart, C. (1998). *Doing a literature review. Releasing the social science research
imagination*. London: Sage Publications.
- Heinström, J. (2005). Fast surfing, broad scanning and deep diving: The
influence of personality and study approach on students' informati-
on-seeking behavior. *Journal of Documentation*, 61(2), 228–247.
- Heinström, J. (2006). Fast surfing for availability or deep diving into quality
– motivation and information seeking among middle and high school
students. *Information Research*, 11(4), paper 433. Saatavissa osoitteessa:
<http://informationr.net/ir/11-4/paper265.html> (käytetty 3.12.2014).
- Limberg, L. (1999). Experiencing information seeking and learning: a study of
interaction between two phenomena. *Information research*, 5(1). <[http://
informationr.net/ir/5-1/paper68.html](http://informationr.net/ir/5-1/paper68.html)>. Luettu 3.12.2014.
- Limberg, L., Alexandersson, M., Lantz-Andersson, A., & Folkesson, L.
(2008). What matters? Shaping meaningful learning through teaching
information literacy. *Libri*, 58(2), 82–91.

- Mateos, M., & Solé, I. (2009). Synthesising information from various texts: A study of procedures and products at different educational levels. *European journal of psychology of education*, XXIV, 435–451. <<http://www.springerlink.com/index/P215371452L14N07.pdf>>. Luettu 3.12.2014.
- McGregor, J., & Streitenberger, D. (2004). Do scribes learn? Copying and information use. In M. K. Chelton & C. Cool (Eds.), *Youth information-seeking behavior: theories, models and issues* (pp. 95–118). Lanham: Scarecrow Press.
- McGregor, J., & Williamson, K. (2005). Appropriate use information at the secondary school level: Understanding and avoiding plagiarism. *Library and Information Science Research*, 27(4), 496–512.
- Purdy, J. (2009). When the tenets of composition go public: A study of writing in Wikipedia. *College Composition & Communication*, (December 2009), 351–373. <<http://datacenter2.aucegypt.edu/bgironda/rhet343/wikiwrite.pdf>>. Luettu 3.12.2014.
- Rea, L. M., & Parker, R. A. (1997). *Designing and conducting survey research: a comprehensive guide*. San Francisco: Jossey-Bass.
- Segev-Miller, R. (2004). Writing from sources: The effect of explicit instruction on college students' processes and products. *L1-Educational Studies in Language and Literature*, 4, 5–33. <<http://www.springerlink.com/index/WK6T602H2474RL1K.pdf>>. Luettu 3.12.2014.
- Siegel, S., & Castellan, N. J. (1988). *Nonparametric statistics for the behavioral sciences*. New York: McGrawHill.
- Sormunen, E., Heinström, J., Romu, L., & Turunen, R. (2012). A method for the analysis of information use in source-based writing. *Information Research*, 17(4), paper 535. <<http://InformationR.net/ir/17-4/paper535.html>>. Luettu 3.12.2014.
- Sormunen, E., & Lehtiö, L. (2011). Authoring Wikipedia articles as an information literacy assignment – copy-pasting or expressing new understanding in one's own words? *Information Research*, 16(4). <<http://informationr.net/ir/16-4/paper503.html>>. Luettu 3.12.2014.
- Spivey, N. N. (1997). *The constructivist metaphor. Reading, writing, and the making of meaning*. San Diego: Academic Press.
- Sundin, O., Francke, H., & Limberg, L. (2011). Practicing information literacy in the classroom. *Dansk Biblioteksforskning*, 7(2/3), 7–17.
- Tynjälä, P., Mason, L., & Lonka, K. (2001). *Writing as a learning tool: Integrating theory and practice*. Dordrecht: Kluwer Academic Publishers.
- Williamson, K., McGregor, J., & Archibald, A. (2010). Assisting students to avoid plagiarism. *Access*, (June), 21–25.