

Lauri Ståhlberg
Risto Hotulainen
Juhani E. Lehto

Lukuseulan toimivuus ja luotettavuus peruskoululaisten lukutaidon arvioinnissa

Kohokohdat

- Lukuseula toimii luotettavasti peruskoululaisten lukutaidon arvioinnissa.
- Tytöt ovat jo alakoulun viimeisillä luokilla noin vuoden poikia edellä lukutaidossa.
- Alkuvuonna syntyminen näyttää tuovan etua lukutaitoon alaluokilla ja vielä siirtäessä yläluokille.

Tutkimuksen aiheena on digitaalisen Lukuseulan toimivuus ja luotettavuus lukutaidon arvioinnissa vuosiluokilla 2–7 (N = 2 815). Toimivuutta ja luotettavuutta arvioitiin analysoimalla Lukuseulan tehtävien psykometrisia ominaisuuksia ja peilaamalla testistä saatuja tuloksia aiempiin lukutaitotutkimuksiin. Analyysit osoittivat, että Lukuseulan tehtävien sisäinen johdonmukaisuus, korrelaatiot kriteerivalidoinnissa käytettyihin testeihin ja testi-uusintatesti-reliabiliteetti olivat hyväksyttäviä. Testitulokset olivat linjassa aiempien tutkimusten kanssa: tytöt menestyivät arvioinneissa poikia paremmin ja olivat todennäköisemmin erin-

omaisia lukijoita. Lisäksi alkuvuonna syntyneet olivat yliedustettuina erinomaisten lukijoiden joukossa. Tutkimus osoittaa, että Lukuseulalla voidaan arvioida peruskoululaisten teknisen lukutaidon ja tekstinymmärtämisen taitoja monipuolisesti, mikä mahdollistaa lukemisen vaikeuksien tarkan erittelyn. Lukuseulan avulla oppilaiden lukemisen vaikeudet voidaan tunnistaa aiempaa tehokkaammin, ja sen ansiosta myös oppimisen tukea osataan kohdentaa entistä tarkemmin. Lisäksi koulutuksen järjestäjät voivat hyödyntää Lukuseulaa arvioidessaan opetusta ja sen vaikuttavuutta. Näin ollen arviointiväline soveltuu lukutaidon osalta myös

peruskoulun tasa-arvon toteutumisen seuranta työkaluksi.

Asiasanat: lukuseula, digitaalinen lukutesti, lukutaito, tekninen lukutaito, tekstinymmärtäminen

JOHDANTO

Suomalaiset peruskoululaiset ovat menestyneet kansainvälisissä lukutaidon arvioinneissa erinomaisesti (Leino ym., 2017; 2019). Menestyksestä huolimatta 15-vuotiaiden heikkojen lukijoiden määrä kasvaa (Leino ym., 2019). Myös lukutaitoerot ovat aiempaa suurempia: uusimmassa PISA-tutkimuksessa arvioitiin, että tyttöjen ja poikien lukutaidon sukupuoliero vastaa reilun vuoden opintoja tyttöjen eduksi ja että ylipäätään sosioekonomiseen luokkaan kuuluvat oppilaat ovat noin kahden kouluvuoden opintojen verran edellä alimpaan sosioekonomiseen luokkaan kuuluvia (Leino ym., 2019).

Ilman riittävää lukutaitoa opetussuunnitelman (Opetushallitus, 2014) tärkeät tavoitteet, kuten arjesta selviäminen ja opintojen jatkaminen, vaarantuvat. Lukemisen vaikeuksia tulisikin seuloa kaikilla peruskoulun vuosiluokilla vähintään kerran vuodessa ja riskiryhmässä olevien lukutaitoa useammin. Näin menettelemällä oppilaiden lukutaidon kehityksestä saataisiin tarkkaa seuranta-aineistoa, heikosti lukevat oppilaat tunnistettaisiin tehokkaasti ja he pääsisivät varmemmin tehostettuun lukutaidon opetukseen.

Tässä artikkelissa esittelemme kehittämämme digitaalisen lukutaidon arviointivälineen Lukuseulan ja kuvailemme sen toimivuutta ja luotettavuutta lukutaidon arvioinnissa luokilla 2–7. Lukuseula mah-

dollistaa teknisen lukutaidon ja tekstinymmärtämisen kehityksen tarkan seurannan ajasta ja paikasta riippumatta. Lukuseula tarkistaa ja analysoi oppilaiden tulokset automaattisesti ja esittää tehtävistä saadut tulokset sekä luokan ja koulun keskiarvon helposti tulkittavina pylväsdiagrammeina. Tarve tällaiselle helppokäyttöiselle tiiviin ja yksilöllisen seurannan mahdollistavalle arviointivälineelle on ilmeinen (Aro, Eklund, Leppänen & Poikkeus, 2011).

Lukuseulan kehitystyö ja teoreettinen perusta

Lukuseulan kehitystyötä ohjasivat teoriasidonnaisuus, tutkimusperustaisuus ja perusopetuksen opetussuunnitelman perusteet (Opetushallitus, 2014). Tehtävien laadinnassa hyödynnettiin monipuolisesti sellaisia kynä-paperitestejä, joiden on osoitettu toimivan lukutaidon arviointivälineinä käytännön koulutyössä. Vaikka samankaltaisella logiikalla toimivia tehtäviä on käytetty aiemmin, Lukuseulan tehtävät poikkeavat niin sisällöiltään kuin toiminoiltaan aiemmista tehtävistä. Tehtävien tekstit ovat uusia, Lukuseulaa varten kirjoitettuja, eikä niitä ole käytetty aiemmin lukutesteissä. Pyrkimyksenä oli etsiä tekstien pohjaksi monipuolisia ja merkityksellisiä aiheita, jotta tehtävien tekeminen olisi oppilaille motivoivaa ja kiinnostavaa.

Lukuseulan teknisen lukutaidon tehtävien logiikat nojaavat kansallisiin ja kansainvälisiin kynä-paperitesteihin (Lindeman, 1998; Nation & Snowling, 1997; Nevala & Lyytinen, 2000; Takala & Kuusela, 2009). Tehtävillä arvioidaan neljää lukusujuvuuteen liittyvää osataitoa: lukemisen nopeus ja tarkkuus, sanantunnistus ja tavutustaito. Tutkimusten mukaan (esim.

Kuhn & Stahl, 2003) sujuva lukeminen edellyttää ainakin kolmen ensiksi mainitun osataidon hallintaa. Lisäksi muun muassa Aro, Huemer, Heikkilä ja Mönkkönen (2011) näkevät yhtenä osatekijänä siirtymän kirjainten lukemisesta tavujen lukemiseen. Artikkelissa emme käsittele laajasti teknisen lukutaidon teoriaa, koska siitä on kirjoitettu suomeksikin paljon (ks. esim. Ahvenainen & Holopainen, 2014; Aro ym., 2011; Huemer, Salmi & Aro, 2012).

Myös tekstinyymmärtämistä arvioidaan Lukuseulassa eri luokka-asteilla useammalla erilaisella tehtävyydellä. Näin toimitaan, koska ymmärtämisen prosessi on moniulotteinen eikä siitä ole mahdollista saada tarkkaa kuvaa yhdellä tehtävyydellä (ks. Keenan, 2016). Työmuistin (Cain, Oakhill & Bryant, 2004) ja tekstinyymmärtämisen perustaitojen, kuten kielien ymmärtämisen, dekodeustaidon (Hoover & Gough, 1990) ja sanavaraston (Verhoeven & Leeuwe, 2008) lisäksi tekstinyymmärtämiseen liittyy korkeamman tasoisia kielellisiä taitoja. Näitä ovat muun muassa päätelmien tekeminen tekstistä, taito arvioida omaa ymmärtämistä ja kyky hahmottaa erilaisia tekstirakenteita (Cain ym., 2004). Lisäksi tekstinyymmärtämiseen liittyvät ainakin eksekutiiviset toiminnot (Meixner, Warner, Lensing, Schiefele & Elsner, 2019), mielen teoria (Atkinson, Slade, Powell & Levy, 2017) ja semanttinen säilömuisti (Nouwens, Groen & Verhoeven, 2017).

Aiempien tutkimusten (esim. Keenan, Betjemann & Olson, 2008; Keenan & Meenan, 2014) mukaan monet tekstinyymmärtämisen testit arvioivat enemmän teknistä lukutaitoa kuin ymmärtämistä. Esimerkiksi tekstiin liittymättömien vastausvaihtoehtojen on havaittu

heikentävän testien validiteettia, koska testattava saattaa tietää oikean vastauksen kysymykseen yleistietonsa pohjalta lukematta tekstiä (Keenan, 2016). Myös lyhyet yhden tai kahden virkkeen pituiset tekstinyymmärtämisen testit ovat ongelmallisia, koska niissä yhdenkin sanan väärä dekodeaus saattaa johtaa väärään vastaukseen (ks. Keenan, 2016; Keenan ym., 2008). Tällaisissa testeissä haasteena on myös se, ettei lyhyt teksti tarjoa välttämättä kontekstia sanojen ymmärtämisen tueksi (Keenan ym., 2008). Tekstinyymmärtämisen arvioimiseen tarkoitettut testit saattavat siis arvioida enemmän jotain muuta taitoa kuin ymmärtämistä, ja oppilas voidaan luokitella heikoksi tekstinyymmärtäjäksi siksi, että testin teoriatausta on heikolla pohjalla (ks. Collins, Lindström & Compton, 2018; Keenan, 2016).

Kiinnitimme Lukuseulan tekstinyymmärtämisen tehtävien laadinnassa huomiota siihen, ettei niissä ole teksteihin liittymättömiä vastausvaihtoehtoja ja että niissä on pituutensa ja sisältönsä puolesta konteksti, jota voidaan hyödyntää ymmärtämisen tukena. Useammalla tehtävyydellä halusimme vähentää virheellisten arviointitulosten todennäköisyyttä ja varmistaa, että tekstinyymmärtämisestä saadaan tarkempi kuva.

Tekstinyymmärtämisen tehtävyyppiin teoriatausta kumpuaa Walter Kintschin ja Teun A. van Dijkin (1978) tekstinyymmärtämisen mallista (van Dijk & Kintsch, 1983). Mallin mukaan lukija rakentaa tekstistä monikerroksista hierarkkista muistiedustusta eli mentaalista representaatiota. Taustalla on teorettinen oletus, jonka mukaan tekstinyymmärtäminen on konstruktivinen prosessi, jossa yhdistyy vuorovaihtus luettavan tekstin, lukijan aiemman

tiedon sekä lukijan intention välillä (van Dijk & Kintsch, 1983).

Tois- ja kolmasluokkalaisille suunnattujen tekstinymmärtämisen tehtävien laadinnassa hyödynsimme koherenssitehtävää ja otsikkotehtävää (Aro, 2002; Cain ym., 2004). Näiden tehtävätyyppeihin pohjalta pystyi laatimaan lyhyitä tekstejä, joiden arvioimme mittaavan enemmän tekstinymmärtämistä kuin teknistä lukutaitoa. Oletimme myös, että nuoremmat oppilaat jaksavat keskittyä lyhyiden tekstien lukemiseen. Tehtävissä arvioidaan muun muassa tekstin otsikointia ja tekstin johdonmukaisuuden eli koherenttiuden hahmottamista, jotka mainitaan tavoitteina myös opetussuunnitelmassa (Opetushallitus, 2014).

Vuosiluokilla 4–7 hyödynsimme tehtävien laadinnassa hierarkiatestiä (Hierarchy-rating test; Lyytinen & Lehto, 1998) ja virkkeiden arviointitestiä (Sentence verification technique, SVT; Royer, Hastings & Hook, 1979). Vaikka testit ovat suhteellisen vanhoja, erilaisia muunnelmia niistä käytetään tutkimuksissa aktiivisesti (esim. Kyttälä & Lehto, 2008; Marcotte, Rick & Wells, 2018; Pichette, Béland, Serres & Lafontaine, 2014; Vainikainen, Hienonen & Hotulainen, 2017).

Tutkimuskysymykset

Esittelemme tutkimuksessamme Lukuseulan tehtävät ja tarkastelemme niiden toimivuutta ja luotettavuutta. Analysoimme Lukuseulan tehtävien reliabiliteettia ja validiteettia, ja vertaamme tehtävistä saatuja tuloksia aiemmin käytössä olleiden lukutaidon mittareiden tuloksiin. Tutkimme, menestyvätkö tytöt poikia paremmin Lukuseulan tehtävissä (esim. Leino

ym., 2019; Lindeman, 1998; Nevala & Lyytinen, 2000), ja analysoimme alku- ja loppuvuonna syntyneiden lukutaitoa (ks. McPhillips & Jordan-Black, 2009; Puhani & Weber, 2008). Tutkimuskysymyksiä muodostui lopulta kolme:

1. Mitkä ovat Lukuseulan tehtävien sisäiset johdonmukaisuudet, kriteerivaliditeetti ja testi-uusintatesti-reliabiliteetti?
2. Menestyvätkö tytöt tehtävissä paremmin kuin pojat, ja ovatko tytöt todennäköisemmin erinomaisia lukijoita ja pojat heikkoja lukijoita?
3. Ovatko alkuvuonna syntyneet yliedustettuina erinomaisissa lukijoissa ja loppuvuonna syntyneet heikoissa lukijoissa?

MENETELMÄ

Tutkittavat ja aineistonkeruu

Lukutaidon arviointiin osallistui syksyllä 2018 yhteensä 2 815 oppilasta 35 eri koulusta. Noin 60 prosenttia osallistujista (n = 1 672) kävi koulua samassa kunnassa. Loput vastaajat jakautuivat suhteellisen tasaisesti 14 kunnan kouluihin Keski- ja Etelä-Suomessa. Vuosiluokille 2–3 suunnattuun testiin osallistui 824 oppilasta (tyttöjä 415, poikia 391, ei tietoa 18) ja vuosiluokille 4–7 suunnattuun 1 991 oppilasta (tyttöjä 1 013, poikia 956, ei tietoa 22). Ennen aineiston keräystä Lukuseula pilotoitiin 150 oppilaalla keväällä 2018.

Opettajat teettivät arvioinnit oppilailleen osana normaalia koulutyötä lukuseula.fi-sivustolla. Laitteina käytettiin koulun omia tietokoneita tai tabletteja. Kun oppilaiden ikä kontrolloitiin, tulokset eivät

eronneet merkittävästi sen perusteella, mitä laitetta käytettiin (tabletti, tietokone sisäisellä hiirellä tai tietokone ulkoisella hiirellä). Tehtävien tekeminen oli mahdollista jakaa myös useammalle päivälle, mutta meillä ei ole tietoa siitä, kuinka moni opettaja toimi näin.

Arviointeihin anottiin lupa kuntien opetustoimen johtajilta sekä mittauksiin osallistuvien koulujen rehtoreilta. Testitilanteesta vastanneille opettajille toimitettiin huoltajille osoitettu kirje, ohjeet tehtävien tekemiseen ja oppilas- ja opettajatunnukset kirjautumista varten. Huoltajien kirjeessä annettiin mahdollisuus kieltäytyä tutkimukseen osallistumisesta. Toteutusta ohjasivat Tutkimuseettisen neuvottelukunnan (TENK, 2012) suositukset.

Lukuseulasovellus tallensi oppilaiden tulokset suojatulle palvelimelle, ja tutkijat latasivat oppilaiden tulokset palvelimelta CSV-muodossa. Taustatietoina kysyttiin konetyyppiä, sukupuolta, ikää, syntymäkuukautta ja osallistumista suomi toisena kielenä -opetukseen. Kaikkia tietoja ei saatu jokaiselta osallistujalta, koska taustatietoihin ei ollut pakko vastata. Aineiston keräyksessä ja säilytyksessä noudatettiin henkilötietojen käsittelyä sääntelevää tietosuojasetusta (2016/679), jota alettiin soveltaa EU-maissa keväällä 2018.

Lukuseulan teknisen lukutaidon tehtävät

Aloitimme Lukuseulan neljän teknisen lukutaidon osatehtävän kehitystyön vuonna 2016. Ensimmäiseen pilottitutkimukseen osallistui 113 oppilasta luokilta 4–6 (Ståhlberg, 2017). Pilotoinnin jälkeen muokkasimme tehtäviä niin, että testattavan tulee läpäistä harjoitustehtävä virheettömästi,

ennen kuin hän pääsee siirtymään tehtäviin. Tällä halusimme varmistaa, että testattavat ymmärtävät tehtävien idean ennen kuin he alkavat tehdä niitä.

Teknisen lukutaidon tehtävät eroavat vuosiluokilla 2–3 ja 4–7 siten, että alemmilla luokilla sanat ovat helpompia, tekstit tai sanalistat ovat lyhyempiä ja aikaa on enemmän. Kaikissa tehtävissä on aikaraja, jonka päätyttyä oppilas näkee tuloksensa näytöltä. Neljästä osatehtävästä on mahdollista saada alemmilla luokilla yhteensä 157 pistettä ja ylemmillä 205 pistettä. Oikeasta vastauksesta saa pisteen ja väärästä vastauksesta miinuspisteen. Poistimme miinus pisteet aineistosta, koska niitä oli suhteellisen vähän ja epäilimme pisteiden perusteella usean oppilaan pyrkineen saamaan tarkoituksella paljon miinus pisteitä.

Epäsanatehtävän pohjana on Tarzan-testi (Mynttinen & Lahti, 1999; Takala & Kuusela, 2009). Tehtävässä arvioidaan lukemisen nopeutta ja dekodauksen huolellisuutta. Muokkasimme tehtävässä käytetyn tekstin Zachris (Sakari) Topeliuksen (1874) *Adalminan helmi* -sadusta ja lisäsimme tekstin joukkoon 65 epäsanaa. Tekstin epäsanat (pseudosanat) noudattavat suomen kielen fonotaktisia sääntöjä: ne näyttävät suomen sanoilta, mutta eivät tarkoita mitään. Tehtävä simuloi neljästä teknisen lukutaidon osatehtävästä parhaiten tosielämän lukutilannetta, koska siinä oppilas lukee kirjoitettua tekstiä eteenpäin niin nopeasti kuin osaa. Tehtävässä oppilaan tulee etsiä tekstistä mahdollisimman monta epäsanaa 3,5 minuutissa ja merkitä ne painamalla niitä hiiren cursorilla tai sormella. Alemmilla luokilla tekstin pituus on 282 sanaa ja pisteitä voi saada 0–30.

Ylemmillä luokilla sanoja on 631 ja pisteitä voi saada 0–65.

Kirjoitusvirhetehtävän pohjana on Sanaketjutestin etsi kirjoitusvirheet -osatesti (Nevala & Lyytinen, 2000). Tehtävä mittaa lukunopeutta ja erityisesti lukemisen tarkkuutta. Tehtävässä etsitään kirjoitusvirheellisiä sanoja 62 sanan sanalistasta, joista puolet (31) on kirjoitettu väärin. Pisteitä voi saada 0–31. Sanat muodostettiin tutkimalla peruskoululaisten tyypillisiä kirjoitusvirheitä: niitä ovat kirjaimen puuttuminen kaksoiskonsonantista tai pitkästä vokaalista, ylimääräinen vokaali tai konsonantti sanassa sekä reversaali (kirjainten paikan vaihtuminen) ja rotaatio (kirjaimen kääntyminen peilikuvaksi). Oppilaan tehtävänä on merkitä sanalistasta mahdollisimman monta väärin kirjoitettua sanaa 1,5–2 minuutissa.

Sanajonotehtävän pohjana on ALLU-testin sanantunnistus-osatesti (Lindeman, 1998). Tehtävässä on 32–35 kirjaimen mittaisia sanajonoja, joissa on sanoja yhteen kirjoitettuna. Oppilaan pitää havaita mahdollisimman nopeasti, missä kohdin sanat vaihtuvat. Tehtävän laadinnassa pidimme tarkasti huolen siitä, ettei suomen kielen sanoja voi muodostua sellaisiin kohtiin, joihin niiden ei ole tarkoitus muodostua. Oppilaan tehtävänä on katkaista 1,5–2 minuutissa sanajonoja oikeilta kohdilta painamalla kirjainvälejä. Kirjainväliä painetaan siitä kohdasta, johon sana päättyy ja josta seuraava alkaa; sanajonon alkua ja loppua ei merkitä. Sanoja on kaikista sanaluokista, ja ne ovat perusmuodossa: niissä on 2–12 kirjainta ja ne ovat 1–5-tavuisia. Alemmilla luokilla pisteitä voi saada 0–38 ja ylemmillä luokilla 0–47.

Tavutustehtävän pohjana on Sanaketjutestin tavuta sanat -osatesti (Nevala

& Lyytinen, 2000). Tehtävässä arvioidaan oppilaan tavutustaitoa ja suomen kielen tavutussääntöjen ymmärrystä. Oppilas voi osata suomen kielen tavutussäännöt tunnistamatta sanaa tai tunnistaa sanan osaamatta tavuttaa eli tavutustaito ei ole välttämätön edellytys tekniselle lukutaidolle. Tehtävässä tavumerkki lisätään sanaan painamalla kirjainväliä; aikaa on 1,5–2 minuuttia. Alemmilla luokilla sanoja on 43 ja pisteitä voi saada 0–58, ylemmillä 50 ja pisteitä voi saada 0–62. Sanat ovat rakenteeltaan, pituudeltaan ja sisällöltään erilaisia nomineja, verbejä ja partikkeleita. Ne sisältävät 1–4 tavua ja niissä on 2–9 kirjainta. Sanat ovat perusmuodossa, ja niiden ensitavuuissa esiintyy kaikki suomen kielen kymmenen perustavutyyppiä.

Lukuseulan tekstinymmärtämisen tehtävät

Otsikko- ja koherenssitehtävä 2.–3-luokkalaisille. Vuosiluokkien 2–3 tekstinymmärtämisen tehtävät arvioivat kykyä tiivistää tekstin pääajatus otsikkoon tai virkkeeseen sekä kykyä ymmärtää tekstin johdonmukaista etenemistä (vrt. Opetushallitus, 2014). Tehtävät suoritetaan itsenäisesti, mutta tehtävien toimintalogiikka opetetaan oppilaille opettajajohtoisella harjoitustehtävällä. Harjoitustehtävät pitää saada oikein, ennen kuin voi siirtyä tekemään tehtäviä.

Kahdessa otsikkotehtävissä on kummassakin neljä vaihtoehtoa, joista oppilaan tulee valita ensimmäisessä osatehtävässä tekstiä tarkimmin kuvaava otsikko ja toisessa tekstiä tarkimmin kuvaava virke. Koherenssitehtävässä tavoitteena on löytää neljästä eri tekstistä kustakin virke, joka rikkoo tekstin johdonmukaisen etene-

misen. Tekstit ovat neljän virkkeen mittaisia, ja jokaisesta voi saada yhden pisteen. Kahdesta otsikkotehtävästä voi saada kaikkiaan 0–8 pistettä ja koherenssitehtävästä 0–4 pistettä. Ennen tehtävien teettämistä esitetasimme niitä kynä-paperiversioina pienellä aikuislukijoiden ryhmällä. Muokkasimme tekstejä ja vastausvaihtoehtoja niin kauan, kunnes kaikki esitestaajat läpäisivät ne virheettömästi.

Hierarkiatesti 4.–7.-luokkalaisille.

Hierarkiatestillä arvioidaan kykyä prosessoida tekstiä eli testattavan ymmärrystä siitä, mikä tekstissä on tärkeää ja mikä vähemmän tärkeää (Lyytinen & Lehto, 1998). Tehtävässä oppilas lukee 128–178 sanan mittaisia tietotekstejä, jotka ovat tehtävää tehdessä näkyvillä. Tehtävä alkaa lapsi-neroista kertovalla harjoitustekstillä (128 sanaa), ja sen tekemisen aikana sekä harjoitustehtävästä saatavalla palautteella testattavalle opetetaan, miten hänen tulee toimia tehtävää tehdessään. Ensimmäinen tehtävä kertoo elektroniikan uusista soveluksista (158 sanaa), toinen susilapsista (178 sanaa) ja kolmas eläinten hyödyntämisestä sodankäynnissä (163 sanaa). Tehtävänä on jakaa kahdeksan tekstiin liittyvää valmiiksi muodostettua virkettä hierarkkisesti kahteen ryhmään: niihin, jotka kuvaavat tekstiä kokonaisuutena (1 virke / teksti), ja niihin, joissa on jokin keskeinen tieto tekstistä (3 virkettä / teksti). Tavoitteena on, että virkkeistä poimitaan ryhmiin neljä ns. makrotason virkettä ja että loput neljä, jotka ovat epäoleellisempia mikrota-son virkkeitä, jäävät alkuperäiselle paikalleen.

Hierarkiatehtävän kolmesta osatehtävästä voi saada yhteensä 0–30 pistettä. Pääajatuksen tiivistävän virkkeen löytämisestä saa neljä pistettä, ja keskeistä

tietoa kuvaavasta virkkeestä saa kaksi pistettä (Lyytinen & Lehto, 1998). Mikäli kokonaisuutta kuvaavan virkkeen paikalle laittaa keskeistä tietoa kuvaavan virkkeen, saa kaksi pistettä, ja mikäli kokonaisuutta kuvaavan virkkeen laittaa keskeisen tiedon paikalle, saa yhden pisteen. Epäoleellisia yksityiskohtia kuvaavan virkkeen siirtämisestä ei saa pisteitä. Tällaisen painokerroin-pisteytyksen käyttö on perusteltua etenkin lyhyehkössä kolmen tekstin mittaisessa hierarkiatestissä (Rantanen, 2003). Ennen arviointia esitetasimme tehtävien toimivuutta kynä-paperiversioina 15–20 henkilön aikuisraadilla ja muokkasimme virkkeitä, kunnes saavutimme jokaisessa osatehtävässä yli 90-prosenttisen yksimielisyyden virkkeiden hierarkkisesta järjestyksestä.

Virkkeiden arviointitesti (SVT) 4.–7.-luokkalaisille. SVT:llä arvioidaan tekstistä muodostetun mentaalisen representaation tarkkuutta eli ”ymmärtämistä” (Royer ym., 1979). Tehtävässä oppilas lukee 12 virkkeen tietotekstejä, joista on muodostettu neljänlaisia väitetyyppejä (ks. Royer, 2001). Tehtävänä on pohtia, ovatko väitteet oikein vai väärin. Jokaisessa osatehtävässä testattavalle esitetään neljä väitetyppiä kolme kertaa (Royer, 2001; ks. myös Lehto & Anttila, 2003). Oikeat väitteet on kirjoitettu täsmälleen samalla tavalla kuin tekstin virkkeet (alkuperäiset, A) tai niiden sisältöä on muutettu, mutta virkkeen sisältämä informaatio on pysynyt samana (rinnakkaisilmaisut, R). Vääristä virkkeistä on muutettu joitain yksityiskohtia siten, että virkkeen sisältämä informaatio on tekstin sisällön kannalta virheellinen (merkitys muutettu, MM) tai ne sisältävät selkeästi tekstiin kuulumatonta tietoa (häiritsevä virke, H). Myös häiritsevät virkkeet

liittyvät tekstin teemaan.

Tehtävä alkaa harjoitustehtävällä (45 sanaa), jossa tehtävänä on pohtia, ovatko neljä väitettä oikein vai väärin. Jokainen väitetyyppi esitetään harjoitustehtävissä yhden kerran. Aina kun oppilas vastaa kysymykseen, Lukuseula ilmoittaa perustelut sille, miksi vastaus on oikein tai väärin. Harjoitustehtävän jälkeen oppilas lukee kolme 12 virkkeen tietotekstiä, jotka kertovat Afrikan eläimistä (123 sanaa), laskuvarjoa kehittäneestä räätälistä (122 sanaa) ja maailman vahvimmasta eläimestä (145 sanaa). Kunkin tekstin luettuaan oppilas vastaa 12 sitä koskevaan väittämään; oikeasta vastauksesta saa pisteen, eli kolmesta tekstistä (osatehtävästä) voi saada 0–36 pistettä. Väitevirkkeisiin siirryttyään testattava ei voi palata lukemaan tekstiä uudestaan, eikä kysymyksissä voi siirtyä taaksepäin vastaamisen jälkeen. Koska tekstiin ei voi palata, testiformaatti edellyttää myös muistia, joka on oleellinen tekstinymmärtämisen osa-alue (esim. Nouwens ym., 2017). Esitestasimme virkkeiden arviointitehtävää digitaalisessa muodossa pienellä aikuislukijoiden ryhmällä ja muokkasimme väitelauseita, kunnes kaikki testattavat saivat tehtävistä yli 90 prosenttia oikein.

Kriteerivaliditeetti ja testi-uusintatesti-mittaus

Käytimme Lukuseulan validoinnissa opettajien aiemmin teettämien teknisen lukutaidon ja tekstinymmärtämisen testien tuloksia, joita vertasimme Lukuseulan tehtävien tuloksiin. Lisäksi pyysimme seitsemäsluokkalaisilta vertailuun äidinkielen arvosanan, jonka he olivat saaneet testin jälkeisen joulun 2018 väliarvioinnissa. Vali-

doinnissa käytetyt testit vaihtelevat jonkin verran, koska eri kouluissa oli käytössä erilaisia seulatestejä. Testi-uusintatestimittaus tehtiin kahdessa koulussa 75 kuudesluokkalaiselle siten, että aikaa kului kahden arvioinnin välillä 4–6 viikkoa.

Saimme vuosiluokilla 2–4 teknisen lukutaidon validointiaineistoksi *Lukusujuvuustestin* tulokset (Salmi, Eklund, Järvisalo & Aro, 2011). Testissä oppilas lukee opettajalle ääneen tekstiä yhden minuutin ja opettaja merkitsee ajan loputtua omaan paperiinsa sen kohdan, johon oppilas ehti. Käytetyssä versiossa opettaja laski, kuinka monta tavua oppilas luki tekstistä. Vuosiluokilta 2 ja 3 sekä 5 ja 6 saimme teknisen lukutaidon tehtävien validointiin *Sanaketjutestin* (Nevala & Lyytinen, 2000), joka mittaa teknistä lukutaitoa ja koostuu neljästä osatehtävästä: epäsanana-, sanaketju-, kirjoitusvirhe- ja tavutus-tehtävä. Lisäksi saimme vuosiluokilta 4, 5 ja 6 hierarkiatehtävän ja virkkeiden arviointitehtävän (SVT) validointiin ALLU-testiin (Lindeman, 1998) kuuluvat kaksi tekstinymmärtämistä arvioivaa tietotekstiä.

Aineiston analyysi

Käytimme aineiston analyysissä SPSS-ohjelmaa (IBM SPSS Statistics 25). Teimme tilastolliset analyysit osatehtävistä laske- tuilla ratkaisuprosenttien keskiarvoilla eli jakamalla ensin oppilaan raakapistemäärän osatehtävän maksimipistemäärällä ja kertomalla sen sadalla (esim. epäsanatehtävän raakapistemäärä $45 / 65 = 0.69$; $0.69 \times 100 = 69\%$). Tämän jälkeen jaoimme osatehtävien ratkaisuprosenttien summan osatehtävien määrällä (esim. [epäsana-% + kirjoitusvirhe-% + sanajono-% + tavutus-%] / 4). Tällä tavalla kaikkien yksit-

täisten osatehtävien painoarvo oli sama. Toimimme näin, koska arvioimme ratkaisuprosentin keskiarvon tulkittamisen olevan selkeämpää kuin standardoidun Z-arvon. Näin on toimittu lukutaitotutkimuksessa myös aiemmin (esim. Vähäpassi, 1987).

Aloitimme analyysit tutkimalla Lukuseulan psykometrisia ominaisuuksia. Analysoimme kahden eri testiversion (2.–3.-lk. ja 4.–7.-lk.) osatehtävien reliabiliteettikertoimia (*Cronbachin alfa*). Seuraavaksi vertailimme Pearsonin korrelaatiokertoimen avulla osatehtävistä saatuja tuloksia luokka-asteittain kriteerivalidoinnissa käytettyjen testien tuloksiin. Tämän jälkeen vertasimme vielä Lukuseulan tuloksia seitsemäsluokkalaisten äidinkielen arvosanaan. Lisäksi analysoimme 75 kuudesluokkalaisten tuloksia Lukuseulan tehtävistä (TL, hierarkia, SVT) testi-uusintatesti-menetelmällä. Nämä analyysit teimme käyttämällä koko aineistoa.

Jatkoimme analyysija tarkastelemalla, kuinka suuren osan eri luokka-asteiden oppilaat saivat keskimäärin ratkaistua Lukuseulan tehtävistä. Analysoimme t-testin avulla, menestyivätkö tytöt poikia paremmin. Laskimme myös eri tehtävien keskiarvojen ja -hajontojen perusteella luokka-aste- ja sukupuolierojen efektikoot (*Cohenin d*; pieni efektikoko 0.2, keskikokoinen 0.5, suuri 0.8; Cohen, 1998; Lenhard & Lenhard, 2016).

Sukupuolierojen jälkeen siirryimme tutkimaan tehtävissä heikosti ja erinomaisesti menestyneiden oppilaiden sukupuoli- ja syntymäkuukausijakaumaa. Tämän tarkastelun teimme lukutaidon kokonaistuloksen perusteella ($[(TL\% + TY\%) / 2]$; toinen ja kolmas luokka $n = 737$ $[(TL\% + hierarkia\% + SVT\%) / 3]$; neljäs ja viides luokka $n = 925$; kuudes ja seitsemäs luok-

ka $n = 479$). Määritimme raja-arvot heikoille ja erinomaisille lukijoille jakamalla oppilaiden tulokset seitsemään osaan, jolloin kumpiakin, sekä heikkoja että erinomaisia lukijoita, oli noin 14 prosenttia jokaisen kolmen yhdistelmäluokan osallistujista. Toimimme näin, koska jako oli käytännöllinen toteuttaa ja uusimman PISA-tutkimuksen (Leino ym., 2019) mukaan lähes yhtä suurella osalla (13,5 %) 15-vuotiaista on heikko lukutaito. Lisäksi heikot tulokset (jakauman alin 14 %) asettuivat vähintään yhden keskihajonnan alle keskiarvon (toinen ja kolmas luokka 1,1–2,3; neljäs ja viides luokka 1,1–2,7; kuudes ja seitsemäs luokka 1,1–3,2) ja erinomaiset tulokset (jakauman ylin 14 %) vähintään yhden keskihajonnan yli keskiarvon (toinen ja kolmas luokka 1,1–3,4; neljäs ja viides luokka 1,1–2,7; kuudes ja seitsemäs luokka 1,1–2,0).

Analysoimme, oliko jokin ryhmä – tytöt tai pojat, alku- tai loppuvuodesta syntyneet – ylliedustettuna heikoissa tai erinomaisissa lukijoissa. Frekvenssijakaumien poikkeamia odotetuista satunnaisjakaumista tutkimme *Pearsonin* χ^2 -testillä. Suomi toisena kielenä -opetusta saavat oppilaat jätettiin näiden analyysien ulkopuolelle (keskiarvo ja -hajonta sekä heikosti ja erinomaisesti menestyneet; taulukot 4, 5 ja 6), koska arvioimme heidän tulostensa vääristävän analyysijä.

TULOKSET

Taulukossa 1 on esitetty Lukuseulan osa-tehtävien reliabiliteettikertoimet. Kertoimet ovat teknisen lukutaidon tehtävissä α

= 0.82–0.85. Vuosiluokkien 2–3 tekstinymmärtämisessä ja luokkien 4–7 virkkeiden arviointitehtävässä (SVT) kertoimet ovat alle $\alpha = 0.7$.

Taulukko 1. Lukuseulan tehtävien reliabiliteettikertoimet 2.–3. ja 4.–7. luokilla

| Luokat | Tehtävä | Osatehtäviä | <i>n</i> | α |
|--------|---------------------|-------------|----------|----------|
| 2.–3. | Tekninen lukutaito | 4 | 800 | 0.82 |
| 2.–3. | Tekstinymmärtäminen | 3 | 828 | 0.61 |
| 4.–7. | Tekninen lukutaito | 4 | 1 963 | 0.85 |
| 4.–7. | Hierarkia | 3 | 1 970 | 0.71 |
| 4.–7. | SVT | 3 | 1 931 | 0.66 |

Taulukossa 2 on esitetty Lukuseulan tehtävien yhteydet sekä Lukuseulan tehtävien ja validoinnissa käytettyjen testien yhteydet 2. ja 3. luokalla. Kriteerivalidointiin saatiin kaksi teknisen lukutaidon testiä (Sanaketju ja Lukusujuvuus). Lukuseulan teknisen lukutaidon ja tekstinymmärtämisen väliset yhteydet ovat 2. ja 3. luokalla kohtalaisia

($r = 0.36$ – 0.52 , $p < 0.001$). Lukuseulan tekninen lukutaito on kummallakin luokalla vahvassa yhteydessä validoinnissa käytettyihin teknisen lukutaidon testeihin ($r = 0.69$ – 0.93 , $p < 0.001$). Toisella luokalla tekstinymmärtämisen ja Lukusujuvuustestin yhteys ei ollut tilastollisesti merkitsevää.

Taulukko 2. Lukuseulan tehtävien, Sanaketjutestin ja Lukusujuvuustestin korrelaatiot 2. ja 3. luokalla

| | <i>n</i> | 1 | 2 | 3 | 4 |
|------------------------------------|----------|---------|---------|---------|---|
| 2. lk. | | | | | |
| 1. Tekninen lukutaito (Lukuseula) | 354 | – | | | |
| 2. Tekstinymmärtäminen (Lukuseula) | 342 | 0.36*** | – | | |
| 3. Sanaketjutesti | 21–22 | 0.93*** | 0.65** | – | |
| 4. Lukusujuvuustesti | 46–51 | 0.85*** | 0.26 | | – |
| 3. lk. | | | | | |
| 1. Tekninen lukutaito (Lukuseula) | 446 | – | | | |
| 2. Tekstinymmärtäminen (Lukuseula) | 439 | 0.52*** | – | | |
| 3. Sanaketjutesti | 38–41 | 0.72*** | 0.53*** | – | |
| 4. Lukusujuvuustesti | 40–43 | 0.69*** | 0.51*** | 0.71*** | – |

** $p < 0.01$, *** $p < 0.001$

Taulukossa 3 on esitetty korrelaatiot luokilla 4, 5 ja 6. Teknisen lukutaidon validointiin saatiin neljäsluokkalaisista Lukusujuvuustesti ja muista luokista Sanaketjutesti. Tekstinymmärtämisen validointiin saatiin ALLU-testin tietotekstit. Lukuseulan teknisen lukutaidon tehtävät ovat yhteydessä Lukuseulan tekstinymmärtämisen tehtä-

viin ($r = 0.36-0.53$, $p < 0.001$) ja ALLUn tekstinymmärtämiseen ($r = 0.55-0.66$, $p < 0.001$). Lukuseulan tekninen lukutaito on vahvasti yhteydessä validoinnissa käytettyihin teknisen lukutaidon testeihin ($r = 0.68-0.92$, $p < 0.001$). Hierarkian ja SVT:n korrelaatiot ALLU-testiin ovat $0.44-0.57$ ($p < 0.001$).

Taulukko 3. Lukuseulan tehtävien ja kriteerivalidoinnissa käytettyjen testien korrelaatiot 4., 5. ja 6. luokalla

| | <i>n</i> | 1 | 2 | 3 | 4 | 5 |
|-----------------------------------|----------|---------|---------|---------|---------|---|
| 4. lk. | | | | | | |
| 1. Tekninen lukutaito (Lukuseula) | 574 | – | | | | |
| 2. Hierarkia (Lukuseula) | 550 | 0.36*** | – | | | |
| 3. SVT (Lukuseula) | 544–555 | 0.53*** | 0.45*** | – | | |
| 4. ALLU-tietotekstit | 108–111 | 0.66*** | 0.52*** | 0.57*** | – | |
| 5. Lukusujuvuus | 62–64 | 0.85*** | 0.41** | 0.65*** | 0.65*** | – |
| 5. lk. | | | | | | |
| 1. Tekninen lukutaito (Lukuseula) | 696 | – | | | | |
| 2. Hierarkia (Lukuseula) | 673 | 0.36*** | – | | | |
| 3. SVT (Lukuseula) | 654–663 | 0.43*** | 0.46*** | – | | |
| 4. ALLU-tietotekstit | 127–129 | 0.55*** | 0.44*** | 0.46*** | – | |
| 5. Sanaketjutesti | 16 | 0.92*** | 0.32 | 0.63** | | – |
| 6. lk. | | | | | | |
| 1. Tekninen lukutaito (Lukuseula) | 449 | – | | | | |
| 2. Hierarkia (Lukuseula) | 444 | 0.41*** | – | | | |
| 3. SVT (Lukuseula) | 436–446 | 0.47*** | 0.47*** | – | | |
| 4. ALLU-tietotekstit | 163–167 | 0.61*** | 0.53*** | 0.57*** | – | |
| 5. Sanaketjutesti | 41 | 0.68*** | 0.21 | 0.18 | 0.65*** | – |

** $p < 0.01$, *** $p < 0.001$

Seuraavaksi analysoimme seitsemäsluokkalaisten äidinkielen arvosanan ja Lukuseulan tehtävien korrelaatioita. Havaitimme, että arvosana on kohtalaisessa yhteydessä Lukuseulan tehtäviin (TL, $n = 192$, $r = 0.56$; hierarkia, $n = 194$, $r = 0.47$; SVT, $n = 193$, $r = 0.42$, $p < 0.001$).

Tämän jälkeen siirryimme tarkastelemaan 4–6 viikon välein tehdyn testin ja uusintatestin yhteyksiä (kuudes luokka, $n = 75$). Havaitimme, että yhteys on lähes täydellinen teknisessä lukutaidossa ($r = 0.96$, $p < 0.001$), mutta heikompi tekstinymmärtämisen tehtävissä (hierarkia, $r = 0.66$; SVT, $r = 0.69$, $p < 0.001$).

Taulukossa 4 esitetään luokkien 2 ja 3 ratkaisuprosenttien keskiarvot ja -hajonnat eri tehtävistä sekä t-testin tulokset sukupuolieroista. Lisäksi tarkastelemme luokka-aste- ja sukupuolierojen efektikokoja. Taulukosta nähdään, että teknisen lukutaidon ja tekstinymmärtämisen tulokset ovat selvästi matalampia toisella kuin kolmannella luokalla. Tyttöjen tulokset ovat kaikissa tehtävissä korkeampia kuin poikien, ja erot ovat tilastollisesti merkitseviä yhtä poikkeusta lukuun ottamatta (toinen luokka, TL, $p = 0.06$). Effektikoko on kuitenkin selvästi suurempi luokka-asteiden kuin sukupuolten välillä.

Taulukko 4. Lukuseulan eri tehtävien keskiarvot ja -hajonnat sekä luokka-aste- ja sukupuolierot 2. ja 3. luokalla

| Tehtävät | Kaikki | | | Työt | | | Pojat | | | Sukupuoliero | | | Efektikoko* | |
|----------------------|----------|-------------|-------------|----------|-------------|-------------|----------|-------------|-------------|--------------|-----------|----------|-------------|-----------|
| | <i>n</i> | <i>ka</i> % | <i>kh</i> % | <i>n</i> | <i>ka</i> % | <i>kh</i> % | <i>n</i> | <i>ka</i> % | <i>kh</i> % | <i>t</i> | <i>df</i> | <i>p</i> | Luokka-aste | Sukupuoli |
| 2. lk. | | | | | | | | | | | | | | |
| Tekninen lukutaito | 334 | 46.1 | 19.4 | 169 | 48.1 | 19.5 | 165 | 44.1 | 19.0 | 1.9 | 332 | n.s. | | 0.2 |
| Tekstin ymmärtäminen | 352 | 52.3 | 19.9 | 176 | 56.8 | 19.9 | 176 | 47.9 | 18.8 | 4.3 | 350 | <0.001 | | 0.5 |
| 3. lk. | | | | | | | | | | | | | | |
| Tekninen lukutaito | 422 | 60.9 | 18.6 | 222 | 62.9 | 18.0 | 200 | 58.7 | 19.1 | 2.3 | 420 | <0.05 | 0.8 | 0.2 |
| Tekstin ymmärtäminen | 428 | 63.0 | 19.4 | 225 | 65.4 | 19.3 | 203 | 60.3 | 19.2 | 2.8 | 426 | <0.01 | 0.6 | 0.3 |

** $p < 0.01$, *** $p < 0.001$

Taulukossa 5 on tehty sama vertailu luokkien 4–7 tehtävistä. Effektikokojen perusteella lukutaito kehittyy ylemmillä luokilla hitaammin kuin alemmillä. Sukupuolten erot sen sijaan näyttäisivät korostuvan ylemmille luokka-asteille siirryttäessä. Tyttöjen tulokset ovat tilastollisesti merkitsevästi poikien vastaavia korkeampia kaikissa muissa tehtävissä paitsi SVT:ssä. Lisäksi tyttöjen ratkaisuprosentit ovat 5. luokalla teknisessä lukutaidossa ja hierarkiatehtävissä samantasoisia kuin

tyttöjen tulokset ovat tilastollisesti merkitsevästi poikien vastaavia korkeampia kaikissa muissa tehtävissä paitsi SVT:ssä. Lisäksi tyttöjen ratkaisuprosentit ovat 5. luokalla teknisessä lukutaidossa ja hierarkiatehtävissä samantasoisia kuin

6. luokan poikien ja 6. luokan tyttöjen kaikissa tehtävissä samantasoisia kuin 7. luokan poikien.

Taulukko 5. Lukuseulan eri tehtävien keskiarvot ja -hajonnat sekä luokka-aste- ja sukupuolierot 4., 5., 6. ja 7. luokalla

| Tehtävät | Kaikki | | | Tytöt | | | Pojat | | | Sukupuoliero | | | Efektikoko* | |
|--------------------|----------|-------------|-------------|----------|-------------|-------------|----------|-------------|-------------|--------------|-----------|----------|---------------|-----------|
| | <i>n</i> | <i>ka</i> % | <i>kh</i> % | <i>n</i> | <i>ka</i> % | <i>kh</i> % | <i>n</i> | <i>ka</i> % | <i>kh</i> % | <i>t</i> | <i>df</i> | <i>p</i> | Luokka-aste | Sukupuoli |
| 4. lk. | | | | | | | | | | | | | | |
| Tekninen lukutaito | 477 | 45.6 | 14.5 | 243 | 48.9 | 13.8 | 234 | 42.3 | 14.5 | 5.1 | 475 | <0.001 | | 0.5 |
| Hierarkia | 467 | 60.2 | 19.3 | 235 | 63.1 | 18.9 | 232 | 57.3 | 19.4 | 3.3 | 465 | <0.01 | | 0.3 |
| SVT | 462 | 71.7 | 11.2 | 234 | 72.5 | 10.7 | 228 | 70.9 | 11.7 | 1.6 | 460 | n.s. | | 0.1 |
| 5. lk. | | | | | | | | | | | | | | |
| Tekninen lukutaito | 530 | 54.8 | 14.4 | 280 | 57.9 | 14.4 | 250 | 51.3 | 13.5 | 5.4 | 528 | <0.001 | 4. vs. 5. lk. | 0.6 |
| Hierarkia | 519 | 70.5 | 19.5 | 278 | 73.0 | 19.1 | 241 | 67.6 | 19.6 | 3.2 | 517 | <0.01 | | 0.5 |
| SVT | 501 | 75.9 | 10.4 | 270 | 76.2 | 10.2 | 231 | 75,6 | 10.6 | 0.6 | 499 | n.s. | | 0.4 |
| 6. lk. | | | | | | | | | | | | | | |
| Tekninen lukutaito | 248 | 59.9 | 14.7 | 122 | 64.5 | 13.5 | 126 | 55.4 | 14.5 | 5.1 | 246 | <0.001 | 5. vs. 6. lk. | 0.4 |
| Hierarkia | 252 | 71.1 | 20.4 | 123 | 74.3 | 20.1 | 129 | 68.1 | 20.2 | 2.5 | 250 | <0.05 | | 0.0 |
| SVT | 251 | 77.9 | 10.7 | 123 | 78.3 | 10.4 | 128 | 77.5 | 11.1 | 0.7 | 249 | n.s. | | 0.2 |
| 7. lk. | | | | | | | | | | | | | | |
| Tekninen lukutaito | 234 | 64.6 | 12.7 | 121 | 66.5 | 12.8 | 113 | 62.6 | 12.3 | 2.4 | 232 | <0.05 | 6. vs. 7. lk. | 0.3 |
| Hierarkia | 235 | 77.7 | 18.8 | 121 | 82.0 | 16.7 | 114 | 73.2 | 19.8 | 3.7 | 233 | <0.001 | | 0.3 |
| SVT | 234 | 78.5 | 11.6 | 121 | 78.7 | 10.9 | 113 | 78.2 | 12.3 | 0.3 | 232 | n.s. | | 0.1 |

*Keskiarvojen ja -hajontojen perusteella laskettu luokka-aste- (kaikki vastaajat, 4. vs. 5. lk., 5. vs. 6. lk. ja 6. vs. 7. lk.) ja sukupuolieron (tytöt vs. pojat 4., 5., 6. ja 7. lk.) efektikoko eri tehtävissä.

*Efektikoko Cohenin *d*; pieni 0.2, keskikokoinen 0.5, suuri 0.8

Taulukossa 6 on esitetty kaikista lukutaidon tehtävistä lasketun ratkaisuprosentin keskiarvon perusteella testissä heikosti ja erinomaisesti menestyneiden oppilaiden määrät ja osuudet sukupuolen ja syntymäkuukauden mukaan jaoteltuina. Jokaisen ryhmän (esim. 2.–3. lk. heikosti menes-

tyneiden sukupuolijakauman) frekvenssien poikkeamista odotetusta satunnaisjakaumasta tutkittiin erikseen. Taulukosta havaitaan, että ylempien luokka-asteiden pojat (4.–5. lk., 6.–7. lk.) ovat yliedustettuina heikosti lukevissa ja tytöt ovat yliedustettuina kaikissa yhdistelmäluokissa

erinomaisesti lukevissa. Alku- ja loppuvuodesta syntyneitä tarkasteltaessa nähdään, että loppuvuodesta syntyneet ovat yliedustettuina heikosti lukevissa vain toi-

sella ja kolmannella luokalla. Sen sijaan alkuvuonna syntyneet ovat yliedustettuina erinomaisesti lukevissa alaluokilla ja siirryttäessä yläluokille.

Taulukko 6. Lukuseulassa heikosti ja erinomaisesti menestyneiden oppilaiden määrät ja osuudet sukupuolen ja syntymäkuukauden mukaan jaoteltuina 2.–3., 4.–5. ja 6.–7. luokilla

| Luokat (N)* | Sukupuoli | | | | | | Alku- ja loppuvuosi* | | | | | |
|-----------------|--------------------|-------|-------|--------------------|-------|-------|----------------------|------|------|--------------------|------|------|
| | Jakauman alin 14 % | | | Jakauman ylin 14 % | | | Jakauman alin 14 % | | | Jakauman ylin 14 % | | |
| | n* | tytöt | pojat | n* | tytöt | pojat | n* | 1 | 2 | n* | 1 | 2 |
| 2.–3. (N = 737) | 105 | 41 % | 59 % | 105** | 67 % | 33 % | 94** | 22 % | 46 % | 97** | 47 % | 21 % |
| 4.–5. (N = 925) | 131** | 37 % | 63 % | 132*** | 72 % | 28 % | 125 | 30 % | 37 % | 122* | 41 % | 26 % |
| 6.–7. (N = 479) | 68** | 31 % | 69 % | 68** | 71 % | 29 % | 67 | 28 % | 28 % | 68* | 49 % | 26 % |

N = Niiden oppilaiden kokonaismäärä, jotka tekivät kaikki lukutaidon tehtävät, ja joiden tiesimme opiskelevan suomen kieli ja kirjallisuus -opimäärää 2.–3., 4.–5. ja 6.–7. luokilla.

n = Niiden oppilaiden määrät, joiden lukutaidon kokonaistulos kuului jakauman alimpaan 14 prosenttiin tai ylimpään 14 prosenttiin 2.–3., 4.–5. tai 6.–7. luokilla. Prosenttiosuuksista nähdään, kuinka suuri osuus näistä oppilaista oli tyttöjä tai poikia tai syntynyt alku- (1. = tammi-huhtikuu) tai loppuvuonna (2. = syys-joulukuu). Alku- ja loppuvuositarastelussa prosenttiosuudet jäivät alle sadan, koska osa heikosti ja erinomaisesti menestyneistä oppilaista oli syntynyt touko–elokuussa.

Frekvenssien jakauma poikkeaa odotetusta satunnaisjakaumasta tilastollisesti merkitsevästi Pearsonin χ^2 -testillä., * $p < .05$ ** $p < .01$ *** $p < .001$

POHDINTA

Artikkelissa esittelimme digitaalisen Lukuseulan ja tutkimme sen toimivuutta ja luotettavuutta lukutaidon arvioinnissa luokilla 2–7. Tehtävien laadintaa ohjasivat teoriasedonaisuus, tutkimusperusteisuus ja opetussuunnitelman perusteet (Opetushallitus, 2014). Tutkimus osoitti, että Lukuseulalla voidaan arvioida luotettavasti peruskoululaisten lukutaitoa.

Ensimmäisessä tutkimuskysymyksessä analysoimme Lukuseulan tehtävien sisäistä johdonmukaisuutta, kriteerivaliditeettia ja testi-uusintatesti-reliabiliteettia. Tulokset osoittivat, että Lukuseulan tehtävien reliabiliteettikertoimet olivat hyväk-

syttävällä tasolla (esim. Taber, 2018). Varsin samanlaisia kertoimia on raportoitu aiemmin kynä-paperimenetelmällä tehtävissä teknisen lukutaidon testeissä (Kiuru ym., 2013), hierarkiatestissä (Lehto ym., 2001) ja virkkeiden arviointitestissä (Lehto & Anttila, 2003; Royer, 2001). Kertoimet jäivät kuitenkin luokkien 2–3 tekstinymmärtämisessä ja luokkien 4–7 virkkeiden arviointitehtävässä (SVT) alle yleisesti käytetyn raja-arvon ($\alpha = 0.7$) (esim. Nunnally, 1978). Nämä matalammat arvot saattavat selittyä suhteellisen lyhyillä tehtäväpatteristoilla: tehtävistä tulisi oletettavasti sisäisesti johdonmukaisempia, mikäli tehtäväpatteristot olisivat pidempiä (vrt. Cortina, 1993; Lehto ym., 2001; Marcotte ym.,

2018; Pichette ym., 2014).

Lukuseulan tehtävien korrelaatiot suhteessa kriteerivalidoinnissa käytettyihin testeihin ($r = 0.41\text{--}0.93$) olivat varsin samanlaisia kuin psykologisissa tutkimuksissa tavallisimmin ($r = 0.40\text{--}0.60$) (Lindeman, 1998). Myös Lukuseulan teknisen lukutaidon tehtävien ja validoinnissa käytettyjen teknisen lukutaidon testien yhteyksien ($r = 0.68\text{--}0.93$) voidaan katsoa olevan riittäviä, kun niitä verrataan esimerkiksi Sanaketjuestin (Nevala & Lyytinen, 2000) osatehtävien ja toisen teknisen lukutaidon tehtävän yhteyksiin ($r = 0.50\text{--}0.74$). Lisäksi teknisen lukutaidon ja SVT:n (Walczyk, 1990), teknisen lukutaidon ja hierarkiatestin (Takala, 2006) sekä hierarkiatestin ja seitsemäsluokkalaisten äidinkielen arvosanojen (Lyytinen & Lehto, 1998) yhteyksistä on raportoitu varsin samanlaisia tuloksia aiemmin.

Lukuseulan tekstinyymmärtämisen ja teknisen lukutaidon korrelaatiot ($r = 0.36\text{--}0.53$) viittaavat siihen, etteivät Lukuseulan tekstinyymmärtämisen tehtävät mittaa liikaa teknistä lukutaitoa (ks. Keenan & Meenan, 2014). Lisäksi Lukuseulan tekstinyymmärtämisen tehtävien (hierarkia ja SVT) väliset yhteydet ja näiden tehtävien yhteydet ALLU-testin (Lindeman, 1998) tekstinyymmärtämiseen olivat vain kohtalaisia. Havainto on samankaltainen aiempien tutkimusten kanssa, joissa on tarkasteltu erilaisten tekstinyymmärtämisen testien korrelaatioita (Keenan & Meenan, 2014; Lehto, Scheinin, Kupiainen & Hautamäki, 2001). Tulokset tukevat näkemystä, jonka mukaan tekstinyymmärtämistä on syytä arvioida useammalla kuin yhdellä tehtävyytyypillä (esim. Keenan, 2016).

Analysoidessamme 4–6 viikon välein tehdyn testin ja uusintatestin tulok-

sia havaitsimme, että teknisen lukutaidon tehtävissä reliabiliteetti oli vahva (vrt. Holopainen, Kairaluoma, Nevala, Ahonen & Aro, 2004; Nevala & Lyytinen, 2000) ja että tekstinyymmärtämisen tehtävissä reliabiliteettikertoimet olivat matalampia. Myös aiemmassa seulatestin validoinnissa samankaltaisesti eri aikoihin suoritettu tekstinyymmärtämisen rinnakkaistesti tuotti suurimmaksi osaksi alle 0.7:n reliabiliteettikertoimia (Lindeman, 1998). Onkin mahdollista, että tekstinyymmärtämisen moniulotteisuus tai patteristojen lyhyys heikentävät testi-uusintatesti-reliabiliteettia.

Toisessa tutkimuskysymyksessä tarkastelimme testituloksissa ilmenneitä sukupuolieroja. Havaitsimme, että tytöt menestyivät tehtävissä poikia paremmin ja olivat todennäköisemmin erinomaisia lukijoita ja pojat olivat todennäköisemmin heikkoja lukijoita (esim. Lindeman, 1998; Nevala & Lyytinen, 2000). Tutkimuksessa havaittiin myös PISA-tutkimusta (Leino ym., 2019) vastaava lukutaitoero tyttöjen ja poikien välillä viidennestä luokasta lähtien: 5. luokan tytöt menestyivät kahdessa Lukuseulan tehtävässä samantasoisesti kuin 6. luokan pojat ja 6. luokan tytöt kaikissa tehtävissä samantasoisesti kuin 7. luokan pojat. Tyttöjen ja poikien lukutaitoeron voi siis karkeasti arvioida vastaavan noin vuoden opintoja jo alakoulun viimeisillä luokilla. Edellisten lisäksi aiemmissa tutkimuksissa on raportoitu samankaltaisia eroja sukupuolten ja luokka-asteiden välillä hierarkiatestissä (Lehto ym., 2001) ja SVT:ssä (Diakidoy, Stylianou, Karefillidou & Papagergiou, 2005; Lehto & Anttila, 2003).

Kolmannessa tutkimuskysymyksessä analysoimme syntymäkuukauden yhteyttä lukutaitoon. Havaitsimme, että

nuoremmat (luokat 2–3) loppuvuodesta syntyneet olivat yliedustettuja heikoissa lukijoissa ja alkuvuonna syntyneet (luokat 2–3, 4–5 ja 6–7) yliedustettuina erinomaisissa lukijoissa (vrt. McPhillips & Jordan-Black, 2009). Tulosten perusteella näyttää siltä, että alkuvuonna syntyminen tuo jonkinlaista etua lukutaitoon alaluokilla ja vielä siirryttäessä yläluokille (vrt. Puhani & Weber, 2008).

Kaikella tutkimuksella on rajoitteensa. Tämän tutkimuksen tulosten luotettavuuteen saattaa vaikuttaa se, että aineisto kerättiin opettajan valvonnassa. Vaikka annoimme tarkat ohjeet tehtävien suorittamiseen, opettajien antama ohjeistus on voinut vaihdella. Ohjeistuksen vaihtelun mahdollista vaikutusta pyrimme minimoimaan pakollisilla harjoitustehtävillä. Emme myöskään saaneet kaikilta osallistujilta tietoa sukupuolesta, S2-taustasta tai syntymäkuukaudesta, koska taustatietoihin ei ollut pakko vastata. Osallistujamäärät olivat kuitenkin riittävän suuria luotettavien analyysien toteuttamiseksi. Puutteistaan huolimatta tutkimuksemme osoittaa, että Lukuseula toimii luotettavasti peruskoulu- ja lukutaidon arvioinnissa. Palautteen perusteella Lukuseula koetaan hyödylliseksi ja tarpeelliseksi työkaluksi peruskoulun oppilaiden lukemisen vaikeuksien seuloon ja lukutaidon arvioinnissa.

Kiitokset

Eriyiskiitos Mikko Asikaiselle tärkeästä avusta Lukuseulan teknisessä kehitystyössä. Kiitokset myös rehtori Marika Haapalaiselle ja erityisopettaja Riikka Qvintukselle sekä testauksissa mukana olleille oppilaille ja opettajille. Lisäksi kiitämme yliopistonlehtori Timo Jaakkolaa ja profes-

sori Pirjo Auniota sekä heidän tutkimusryhmäänsä Lukuseulan hyödyntämisestä tutkimuksessa. Tarkasta kommentoinnista kiitokset kasvatustieteiden tohtori Irene Rämälle ja kriittisistä näkökulmista vertaisarvioijille.

Kirjoittajatiedot:

Lauri Ståhlberg (KM) toimii Helsingin yliopiston Koulutuksen arviointikeskuksessa projektisuunnittelijana. Ståhlberg tutkii väitöskirjassaan lukutaitoa ja siihen yhteydessä olevia tekijöitä. Lisäksi Ståhlberg kehittää lukutaidon arviointia Lukuseula Oy:ssä.

Risto Hotulainen (professori) toimii Helsingin yliopiston Koulutuksen arviointikeskuksen johtajana.

Juhani E. Lehto (dosentti) toimi Helsingin yliopiston avoimessa yliopistossa erityispedagogiikan yliopistonlehtorina elokuuhun 2020 saakka.

LÄHTEET

- Ahvenainen, O. & Holopainen, E. (2014). Lukemis- ja kirjoittamisvaikeudet: Teoreettista taustaa ja opetuksen perusteita. Jyväskylä: Special Data.
- Aro, M., Eklund, K., Leppänen, P. & Poikkeus, A.-M. (2011). Lukivaikeusriskin arviointi ja lukivaikeuden tunnistaminen suomen kielessä. *Psykologia*, 46(2-3), 92-98.
- Aro, M., Huemer, S., Heikkilä, R. & Mönkkönen, V. (2011). Sujuva lukutaito suomalaislapsen haasteena. *Psykologia*, 46(2-3), 153-155.
- Aro, T. (2002). Luetun ymmärtämisen teoriaa ja harjoituksia. Jyväskylä: Niilo Mäki Instituutti.
- Cain, K., Oakhill, J. & Bryant, P. (2004). Children's reading comprehension ability: Concurrent prediction by working memory, verbal ability, and component skills. *Journal of Educational Psychology*, 96(1), 31-42.
- Cohen J. (1988). *Statistical power analysis for the behavioral sciences*. Toinen painos. Hillsdale (NJ): Lawrence Erlbaum.
- Collins, A. A., Lindström, E. R. & Compton, D. L. (2018). Comparing students with and without reading difficulties on reading comprehension assessments: A meta-analysis. *Journal of Learning Disabilities*, 51(2), 108-123.
- Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology*, 78(1), 98-104.
- Diakidoy, I.-A. N., Stylianou, P., Karefillidou, C. & Papagergiou, P. (2005). The relationship between listening and reading comprehension of different types of text at increasing grade levels. *Reading Psychology*, 26(1), 55-80.
- Duckworth, A. L. & Seligman, M. E. P. (2006). Self-discipline gives girls the edge: Gender in self-discipline, grades, and achievement test scores. *Journal of Educational Psychology*, 98(1), 198-208.
- Holopainen, L., Kairaluoma, L., Nevala, J., Ahonen, T. & Aro, M. (2004). Lukivaikeuksien seulontamenetelmä nuorille ja aikuisille. Jyväskylä: Niilo Mäki Instituutti.
- Hoover, W. A. & Gough, P. B. (1990). The simple view of reading. *Reading and Writing*, 2(2), 127-160.
- Huemer, S., Salmi, P. & Aro, M. (2012). Tavoitteena sujuva lukutaito. *NMI Bulletin*, 22(2), 18-24.
- Keenan, J. M. (2016). Assessing the assessments reading comprehension tests. *Perspectives on Language and Literacy*, 42(2), 17-21.
- Keenan, J. M., Betjemann, R. S. & Olson, R. K. (2008). Reading comprehension tests vary in the skills they assess: Differential dependence on decoding and oral comprehension. *Scientific Studies of Reading*, 12(3), 281-300.
- Keenan, J. M. & Meenan, C. E. (2014). Test differences in diagnosing reading comprehension deficits. *Journal of Learning Disabilities*, 47(2), 125-135.
- Kintsch, W. & van Dijk, T. A. (1978). Toward a model of text comprehension and production. *Psychological Review*, 85(5), 363-394.
- Kiuru, N., Lerkkanen, M.-K., Niemi, P., Poskiparta, E., Ahonen, T., Poikkeus, A.-M. & Nurmi, J.-E. (2013). The role of reading disability risk and environmental protective factors in students' reading fluency in grade 4. *Reading Research Quarterly*, 48(4), 349-368.
- Kuhn, M. & Stahl, S. (2003). Fluency: A review of developmental and remedial practices. *Journal of Educational Psychology*, 95(1), 3-21.
- Lehto, J. E. & Anttila, M. (2003). Listening comprehension in primary level grades two, four and six. *Scandinavian Journal of Educational Research*, 47(2), 133-143.
- Lehto, J. E., Scheinin, P., Kupiainen, S. & Hautamäki, J. (2001). National survey of reading comprehension in Finland. *Journal of Research in Reading*, 24(1), 99-110.
- Leino, K., Ahonen, A., Heinonen, N., Hiltunen, J., Lintuvuori, M., Lähteinen, S., ... & Vettenranta, J. (2019). PISA 18 ensituloksia. Opetus- ja kulttuuriministeriön julkaisu 40. Helsinki: Opetus- ja kulttuuriministeriö.
- Leino, K., Nissinen, K., Puhakka, E. & Rautapuro, J. (2017). Lukutaito luodaan yhdessä. Kansainvälinen lasten lukutaitotutkimus (PIRLS 2016). Jyväskylä: Koulutuksen tutkimuslaitos.
- Lenhard, W. & Lenhard, A. (2016). Calculation of effect sizes. Dettelbach (Saksa): Psychometrica. Haettu osoitteesta https://www.psychometrica.de/effect_size.html.
- Lindeman, J. (1998). *Ala-asteen lukutesti ALLU*. Turku: Turun yliopisto, Oppimistutkimuksen keskus.
- Lyytinen, S. & Lehto, J. E. (1998). Hierarchy rating as a measure of text macroprocessing: Relationship with working memory and school achievement. *Educational Psychology*, 18(2), 157-169.
- Marcotte, A. M., Rick, F. & Wells, C. S. (2018). Investigating the reliability of the Sentence Verification Technique. *International Journal of Testing*, 19(1), 74-95.
- McPhillips, M. & Jordan-Black, J.-A. (2009). The effect of month of birth on the attainments of primary and secondary school pupils. *British Journal of Educational Psychology*, 79(3), 419-438.
- Meixner, J. M., Warner, G. J., Lensing, N., Schiefele, U. & Elsner, B. (2019). The relation between executive functions and reading comprehension in primary-school students: A cross-lagged-panel analysis. *Early Childhood Research Quarterly*, 46(1), 62-74.

- Mynttinen, S. & Lahti, P. (1999). Hiljainen lukeminen. Proseminaaritutkielma Opettajankoulutuslaitos, erityispedagogiikka. Helsinki: Helsingin yliopisto.
- Nation, K. & Snowling, M. (1997). Assessing reading difficulties: the validity and utility of current measures of reading skill. *British Journal of Educational Psychology*, 67(3), 359–370.
- Nevala, J. & Lyytinen, H. (2000). Sanaketjutesti: käsikirja 2: tekninen opas. Jyväskylä: Jyväskylän yliopisto.
- Nouwens, S., Groen, M. A. & Verhoeven, L. (2017). How working memory relates to children's reading comprehension: The importance of domain-specificity in storage and processing. *Reading and Writing*, 30(1), 105–120.
- Nunnally, J. C. (1978). *Psychometric theory*. New York: McGraw-Hill.
- Opetushallitus (2014). Perusopetuksen opetussuunnitelman perusteet 2014. Helsinki: Opetushallitus.
- Pichette, F., Béland, S., de Serres, L. & Lafontaine, M. (2014). Psychometric validation of the Sentence Verification Technique to assess L2 reading comprehension ability. *The Quantitative Methods for Psychology*, 10(2), 95–106.
- Puhani, P. A. & Weber, A. M. (2007). Does the early bird catch the worm? *Empirical Economics*, 32(2), 359–386.
- Rantanen, P. (2003). Enemmän vähemmällä: monivalintatehtävien mittaustarkkuuden nostaminen. Turku: Suomen kasvatustieteellinen seura.
- Royer, J. M. (2001). Developing reading and listening comprehension tests based on the Sentence Verification Technique (SVT). *Journal of Adolescent & Adult Literacy*, 45(1), 30–41.
- Royer, J. M., Hastings, C. N. & Hook, C. (1979). A Sentence Verification Technique for measuring reading comprehension. *Journal of Reading Behavior*, 11(4), 355–363.
- Salmi, P., Eklund, K., Järvisalo, E. & Aro, M. (2011). LukiMat – Oppimisen arviointi: Lukemisen ja kirjoittamisen tuen tarpeen tunnistamisen välineet 2. luokalle. Käyttäjän opas. Jyväskylä: Niilo Mäki Instituutti.
- Ståhlberg, L. (2017). Uusi menetelmä lukutaidon pulmien seulomiseen – verkkopohjainen Lukuseula 4.–6.-luokkalaisten teknisen lukutaidon arvioimisessa. Proseminaaritutkielma. Erityispedagogiikka. Helsinki: Helsingin yliopiston avoin yliopisto.
- Taber, K. S. (2018). The use of Cronbach's Alpha when developing and reporting research instruments in science education. *Research in Science Education*, 48(6), 1273–1296.
- Takala, M. (2006). The effects of reciprocal teaching on reading comprehension in mainstream and special (SLI) education. *Scandinavian Journal of Educational Research*, 50(5), 559–576.
- Takala, M. & Kuusela, J. (2009). Tarzan helps to find dyslexics: A pseudo-word test. *Support for Learning*, 24(2), 81–85.
- TENK (2012). Hyvä tieteellinen käytäntö ja sen loukkausepäilyjen käsitteleminen Suomessa. Haettu osoitteesta https://www.tenk.fi/sites/tenk.fi/files/HTK_ohje_2012.pdf.
- Topelius, Z. (1874). *Lukemisia lapsille Osa 1*. Helsinki: A. Hougberg.
- Vainikainen, M.-P., Hienonen, N. & Hotulainen, R. (2017). Class size as a means of three-tiered support in Finnish primary schools. *Learning and Individual Differences*, 56, 96–104.
- van Dijk, T. A. & Kintsch, W. (1983). *Strategies of discourse comprehension*. New York: Academic Press.
- Verhoeven, L. & Leeuwe, J. (2008). Prediction of the development of reading comprehension: A longitudinal study. *Applied Cognitive Psychology*, 22(3), 407–423.
- Vähäpassi, A. (1987). Tekstinymmärtäminen: tekstinymmärtämisen tasosta suomalaisessa peruskoulussa. Kasvatustieteiden tutkimuslaitoksen julkaisusarja A. Tutkimuksia 10. Jyväskylä: Jyväskylän yliopisto.
- Walczyk, J. J. (1990). Relation among error detection, sentence verification, and low-level reading skills of fourth graders. *Journal of Educational Psychology*, 82(3), 491–497.