



**TURUN
YLIOPISTO**

Ohjelmoinnillisen ajattelun mittarit

Systemaattinen kirjallisuuskatsaus

Kasvatustiede/ Opettajankoulutuslaitos
pro gradu –tutkielma

Laatijat:
Kim Erola & Jim Mirel

12.4.2023
Turku

Turun yliopiston laatu järjestelmän mukaisesti tämän julkaisun alkuperäisyys on tarkastettu
Turnitin OriginalityCheck -järjestelmällä.

Pro gradu -tutkielma

Oppiaine: Kasvatustiede / Opettajankoulutuslaitos

Tekijät: Kim Erola & Jim Mirel

Otsikko: Ohjelmoinnillisen ajattelun mittarit

Ohjaaja(t): professori Marjaana Veermans

Sivumäärä: 41 sivua

Päivämäärä: 12.4.2023

Kouluissa pohditaan jatkuvasti millä tavoin voidaan parhaalla mahdollisella tavalla kasvattaa lapsia tulevaisuutta varten. Ohjelmoinnillinen ajattelu nähdään yhtenä tärkeänä taitona hallita nyt ja tulevaisuudessa. Ohjelmoinnillinen ajattelu on monisäkeinen kokonaisuus ja sen vuoksi vaikeasti arvioitavissa, sen arviointi on välttämätöntä, jotta se voidaan laaja-alaisesti yhdistää osaksi opetussuunnitelmia.

Tämän pro gradu –tutkielman tarkoituksena oli selvittää, millaisia mittareita kasvatustieteellisessä tutkimuksessa on käytetty arvioimaan alakouluikäisten oppilaiden ohjelmoinnillisen ajattelun taitoja. Toisena tarkoituksena oli tutkia näiden mittareiden raportoituja psykometrisiä ominaisuuksia, sisällön validiteettia ja rakennevaliditeettia sekä reliabiliteettia. Näiden asioiden avulla voidaan paremmin määritellä mitä ohjelmoinnillisen ajattelun arvioinnista jo tiedetään ja mitä tulisi kehittää.

Tutkielmassa menetelmänä käytettiin systemaattista kirjallisuuskatsausta. Aineistonkeruu suoritettiin Finkin mallin mukaisesti, aloittaen artikkeleiden hausta EBSCOHost-tietokantajärjestelmässä, jonka pohjalta artikkeleita löytyi 185 kappaletta. Artikkeleiden haun jälkeen suoritimme itsenäisen artikkelien tiivistelmien arvioinnin. Lopulta tutkimukseen valikoitui 185 artikkelia, joissa oli 11 ohjelmoinnillisen ajattelun mittaria, joiden psykometrisiä ominaisuuksia oli mitattu ja raportoitu.

Tutkimuksen päätuloksia olivat vähäinen psykometrisiä ominaisuuksia raportoivien mittarien määrä etenkin reliabiliteetin osalta. Sisällön validiteettia oli pyritty vahvistamaan ja vastaamaan näin aikaisemmissa tutkimuksissa esiin nousseisiin haasteisiin, mutta sisällön validiteetinkin osalta tulosten raportointi jäi vähäiseksi. Rakennevaliditeettia rinnakkaisvaliditeetin ja kriteerivaliditeetin osalta esiintyi vain muutamassa mittarissa. Reliabiliteetin mittausta ja raportointia jäi lähes jokaisen mittarin kohdalla Cronbachin alfan selvittämiseen.

Tutkielma vahvisti käsitystä heikosta psykometristen ominaisuuksien testaamisen tasosta ohjelmoinnillisen ajattelun mittareiden tutkimuksessa ja raportoinnissa. Jatkotutkimuksena tulisikin kehittää ennen kaikkea jo olemassa olevien mittarien luotettavuutta ja reliabiliteettia, validoimalla näitä laajemmin ja tuottamalla lisää reliabiliteettia parantavia testejä sisältäviä tutkimuksia mittareilla.

Avainsanat: ohjelmoinnillinen ajattelu, arviointi, mittarit, validiteetti, reliabiliteetti

Sisällys

1. Johdanto	1
2 Ohjelmoinnillinen ajattelu	3
2.1 Ohjelmoinnillisen ajattelun määrittely	3
2.1 Ohjelmoinnillisen ajattelun arviointi osana opetusta	5
3 Ohjelmoinnillisen ajattelun arviointi	7
3.1 Arvioinnin luotettavuus	7
3.2 COSMIN -taksonomia validiteetin ja reliabiliteetin tarkastelussa	8
4 Tutkimusongelmat	11
5. Tutkimuksen toteuttaminen	13
5.1 Finkin malli systemaattisen kirjallisuuskatsauksen kehikkona.....	13
5.2 Tutkimustehtävä ja tutkimusaineiston hankinta.....	16
5.3 Tutkimusaineiston valintakriteerit ja -prosessi.....	17
5.4 Tutkimusaineiston analyysi	20
6 Tutkimustulokset	23
6.1 Tutkimuksissa käytetyt mittarit.....	23
6.2. Mittareiden validiteetti	25
6.3. Mittareiden reliabiliteetti	29
7 Pohdinta	32
7.1 Tutkimuksen luotettavuus ja tutkimusetiikka	35
Lähdeluettelo	37

1. Johdanto

Nykypäivänä koulutusalan keskusteluissa on yhä kasvavassa määrin huomioitu tulevaisuuden taitojen merkitys sekä digitalisaation integroiminen opetukseen. (Tanhua-Piironen, Viteli, Syvänen, Vuorio, Hintikka & Sairanen, 2016, 68–69). Teknologian rooli ja merkitys kasvavat yhteiskunnassa, kuten myös koulumaailmassa hurjaa vauhtia, mikä vaatii niin oppilailta kuin opettajilta kykyä käyttää teknologiaa sekä ratkaista sen synnyttämiä haasteita (Tanhua-Piironen ym., 2016, 64). Ohjelmoinnillinen ajattelu nähdään yhtenä ratkaisuna näihin haasteisiin. Sen avulla pystytään ymmärtämään teknologian toimintaa sekä sen tuomia mahdollisuuksia paremmin. Lisäksi se kehittää erilaisia taitoja, kuten ongelmanratkaisua sekä järjestelmällistä ajattelua (CSTA & ISTE, 2011).

Koulujen yhtenä tarkoituksena on kasvattaa oppilaista tulevaisuudessa tehokkaita työssä käyviä yksilöitä yhteiskunnalle (Ahtola, 2016, 15). Tulevaisuudessa myös yhä useammat työpaikat liittyvät teknologian ja ohjelmoinnin pariin ja näissä teknologian ja ohjelmoinnin taitaminen antaa työtä hakeville etulyöntiaseman työmarkkinoiden kilpailussa (Manyika, 2017). Lisäksi erilaisten teknologisten laitteiden käyttö yleistyy ja määrä kasvaa, mikä tarkoittaa, että teknologisilla ratkaisuilla on yhä suurempi rooli ihmisten elämässä. Tämä tarjoaa mahdollisuuksia uusien ja innovatiivisten ratkaisujen luomiseen yhteiskunnan haasteissa. Ohjelmoinnillinen ajattelu voi tukea näiden ratkaisujen kehitystä.

Ohjelmoinnillinen ajattelu on lähtöisin 1980-luvulta, jolloin se nähtiin ohjelmoinnin alakäsitteenä ja tukena ohjelmoinnin oppimiseen (Papert, 1980). Ohjelmointi tarkoittaa tietokoneelle tai samantapaiselle ohjelmoidulle laitteelle jollakin tavalla, tyypillisesti kirjoittamalla, annettavia komentoja formaalilla ohjelmointikielellä. (Balanskat & Engelhardt, 2015.) Papertin (1980; 1991) alkuperäisiä ajatuksia on nykypäivänä edelleen kehitetty vastaamaan 2000-luvun tarpeisiin erityisesti internetin ja globaalin informaation hallinnan alueilla. Tämän kehityksen myötä ohjelmoinnillisen ajattelun käsite on vakiinnuttanut asemansa koulutuksen kontekstissa. (Voogt, Fisser & Good, 2015).

Ohjelmoinnillinen ajattelu on kasvattanut asemaansa koulutuksesta keskusteltaessa Jeanette Wingin (2006) nostettua se esille. Hänen mukaansa ohjelmoinnillinen ajattelu tulee olemaan yksi 2000-luvun tärkeimmistä taidoista, jota käyttää jokainen vuoteen 2050 mennessä. Ohjelmoinnillisen ajattelun tärkeyttä on korostettu aiemmin varsinkin ohjelmoinnin opetuksen apuna (Ioannidou, Bennett, Repenning, Koh & Basawapatna, 2011), mutta Wingin ja

Resnickin mukaan “Ohjelmoinnillinen ajattelu on enemmän kuin ohjelmointia, samalla tavalla kuin kirjallisuus on enemmän kuin vain kirjoittamista.” (National Research Council, 2010, 13; Wing, 2017). Ohjelmoinnillinen ajattelu onkin myöhemmin noussut keskeiseksi teemaksi ei pelkästään tietojenkäsittelytieteen alalla vaan laajalti koulutuksen ja perusopetuksen opetussuunnitelmien sisällä (Voogt ym., 2015). Wing (2017) on muovannut itsekin näkemystään myöhemmin ja määrittää ohjelmoinnillisen ajattelun ajatusprosesseiksi, jotka vaaditaan ongelman muotoilemiseen sellaiseksi, jonka ihminen tai kone voi tehokkaasti ratkaista.

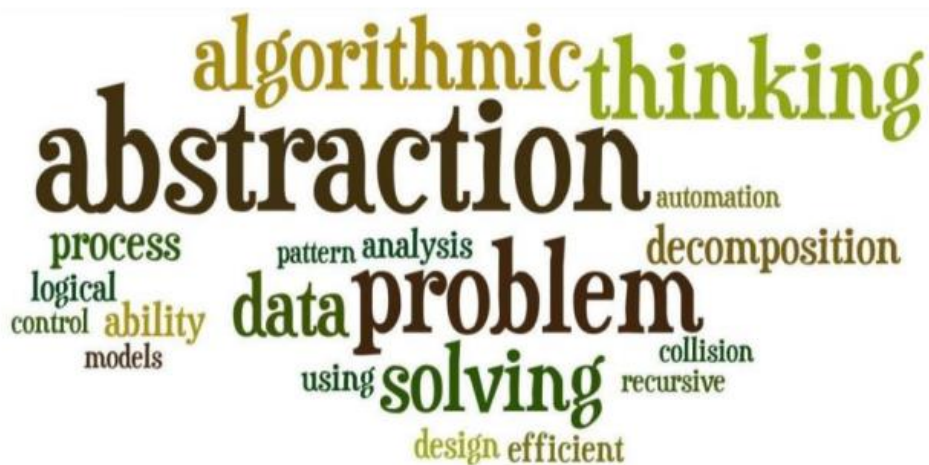
Ohjelmoinnillinen ajattelu nähdään nykyään tärkeänä monipuolisena taitona, joka jokaisen tulisi hallita, samalla tavalla kuin kirjoittaminen, lukeminen ja laskeminen, ja täten sen liittäminen koulujen opetussuunnitelmiin on ensiarvoisen tärkeää. (Wing, 2006, 33; Ehsan & Cardella, 2021.) Tangin, Tsoun ja Chain (2020) katsaustutkimuksen perusteella 2006–2018 välillä, 52 maassa ohjelmoinnillinen ajattelu oli jollain tavalla sisällytetty maan opetussuunnitelmaan. Ohjelmoinnillisen ajattelun liittäminen sellaisenaan opetussuunnitelmaan vaatisi, että se olisi arvioitavissa oleva taito ja jotta sitä voitaisiin arvioida, tulisi olla yksiselitteisempi käsitys sen määritelmästä (Voogt ym., 2015).

Tämän tutkimuksen tarkoituksena oli kartoittaa, systemaattisen kirjallisuuskatsauksen avulla, minkälaisia mittareita on käytetty ohjelmoinnillisen ajattelun taitojen arviointiin alakoulussa sekä tutkia näiden olemassa olevien ohjelmoinnillisen ajattelun mittareiden psykometrisiä ominaisuuksia.

2 Ohjelmoinnillinen ajattelu

2.1 Ohjelmoinnillisen ajattelun määrittely

Ohjelmoinnillisen ajattelun määritelmästä opetuksessa ei ole vielä yhtä yhtenäistä konsensusta tutkijoiden kesken, mutta siitä ollaan yhtä mieltä, että ohjelmoinnillinen ajattelu tarkoittaa eri asiaa kuin ohjelmointi (Tikva & Tambouris, 2021; Passey, 2017). Monet tutkijat ovat pyrkineet löytämään mahdollisimman yksikäsitteisen ja selkeän määritelmän käsitteelle. (Curzon, Dorling, Selby, Woollard & Ng, 2014; National Research Council, 2011; Shute, Sun, & Asbell-Clarke, 2017). Kalelioglu, Gulbahar ja Kukul (2016) selvittivät määritelmää pohtiessaan yleisimmin ohjelmoinnillisen ajattelun kanssa samoissa asiayhteyksissä esiintyvien sanojen lukumäärää ja tulivat siihen tulokseen, että useimmin esiintyvät sanat abstraktio, ongelma, ratkaisu, algoritminen ja ajattelu. (Kuvio 1.)



Kuvio 1. Yleisimmät sanat, jotka esiintyvät ohjelmoinnillisen ajattelun kanssa samoissa asiayhteyksissä (Kalelioglu, Gulbahar & Kukul, 2016, 586)

The International Society for Technology in Education (ISTE) ja Computer Science Teacher Association (CSTA) julkaisivat oman määritelmänsä ohjelmoinnillisesta ajattelusta seuraavasti “ohjelmoinnillinen ajattelu on ongelmanratkaisuprosessi, joka sisältää ainakin seuraavanlaisia piirteitä:

Ongelmien muotoileminen ratkaistavaan muotoon tietokoneiden tai muiden välineiden avulla,

Looginen organisointi ja tiedon analysointi,

Tiedon esitleminen abstraktioiden kuten mallien ja simulaatioiden avulla,

Ratkaisujen automatisointi algoritmisen ajattelun kautta,

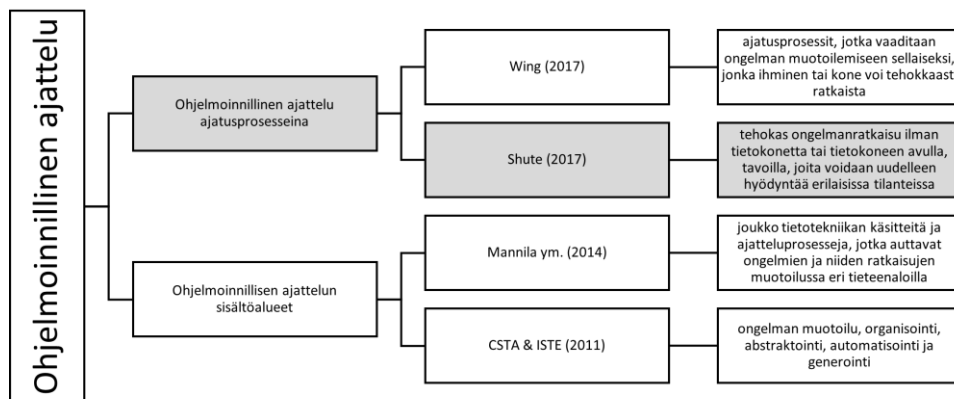
Mahdollisimman tehokkaan ratkaisukeinon löytäminen,

Ratkaisun generointi käytettäväksi muiden ongelmien ratkaisuun.”

(CSTA & ISTE, 2011)

Tämän määritelmän lisäksi Mannila ym. (2014) määrittivät, että ohjelmoinnillinen ajattelu on termi, joka kattaa joukon tietotekniikan käsitteitä ja ajatteluprosesseja, jotka auttavat ongelmien ja niiden ratkaisujen muotoilussa eri tieteenaloilla. Aho (2012) sen sijaan määrittelee ohjelmoinnillisen ajattelun "ajatusprosesseiksi, jotka liittyvät ongelmien muotoilemiseen siten, että niiden ratkaisut voidaan esittää loogisten vaiheiden avulla ja algoritmeina". Ohjelmoinnilliselle ajattelulle on siis useita eri määritelmiä.

Ohjelmoinnillisen ajattelun määrittely voidaan kuitenkin jakaa karkeasti kahteen pääkategoriaan: yleisiin määritelmiin, jotka käsittävät ohjelmoinnillisen ajattelun ajatusprosessina tai -prosesseina, kuten Wing ja Shute, ja määrittelymalleihin, jotka esittävät mitä ohjelmoinnillinen ajattelu sisältää, kuten Mannila ym. sekä CSTA ja ISTE. (Kuvio 2.)



Kuvio 2. Ohjelmoinnillisen ajattelun määrittely

Tässä katsauksessa käsittelemme ohjelmoinnillista ajattelua Shuten (2017) määritelmän mukaan, tehokkaana ongelmanratkaisuna ilman tietokonetta tai tietokoneen avulla, tavoilla, joita voidaan uudelleen hyödyntää erilaisissa tilanteissa (Shute ym., 2017, 142) sekä jaamme sen kolmeen yleisesti todettuun alakäsitteeseen (Tikva & Tambouris, 2021; Zhang & Nouri, 2019): näkökulmiin eli ohjelmoinnillisiin tapoihin jäsentää maailmaa ympärillä; sisältöihin, jotka ovat hyödynnettävissä ohjelmoitaessa sekä käytänteisiin, joiden avulla ohjelmoinnin sisältöjä pyritään hyödyntämään mahdollisimman tehokkaasti. Shuten määritelmä on ytimekäs, mutta kattava ja pohjautuu yleisesti todettuihin alakäsitteisiin, joten näimme sen parhaana määritelmänä meidän tutkimukseemme.

2.1 Ohjelmoinnillisen ajattelun arviointi osana opetusta

Useat asiantuntevat tahot ja organisaatiot ovat kannattaneet ohjelmoinnillisen ajattelun sisällyttämistä opetussuunnitelmiin. (CSTA, 2017; OECD, 2019; UNESCO, 2011) European Schoolnet (2017) on julkaissut useita raportteja, joissa suositellaan ohjelmoinnillisen ajattelun sisällyttämistä opetussuunnitelmiin. Tämän pedagogiseksi perusteen tueksi on löydettävä tapoja arvioida sen osaamisen tasoa. (Zapata-Caceres, Martin & Roman-Gonzalez, 2021) Ohjelmoinnillisen ajattelun määritelmän tavoin sen arvioinnin tavasta perusopetuksessa ei ole yhteistä näkemystä. Ohjelmoinnillisen ajattelun opetuksen yhtenä suurena ongelmana pidetään opettajien heikkoa ohjelmoinnillisen ajattelun opetuksen tasoa. (Tang, Yin, Lin, Hadad & Zhai, 2020) Ohjelmoinnillisen ajattelun lisääminen opetussuunnitelmiin johtaisi

ennen pitkään myös sen liittämiseen opettajankoulutuslaitosten koulutussisältöihin Suomessa ja maailmalla, mikä parantaisi opettajien kykyä toteuttaa laadukasta ohjelmoinnillisen ajattelun opetusta.

Tangin ym. (2020) mukaan ohjelmoinnillisen ajattelun arviointiin on hyödynnetty neljää arviointitapaa: monivalintakysymyksiä ja avoimia kysymyksiä sisältäviä perinteisiä testejä, portfolioarviointia, haastatteluja ja kyselyitä. Tutkimuksemme keskittyy arviointityökalujen validiteetin ja reliabiliteetin selvittämiseen, joten perinteisten testien osuus tämän tutkimuksen kannalta on korostuneessa asemassa.

Tikva & Tambouris (2021) kehittivät katsauksensa perusteella käsitteellisen mallin ohjelmoinnillisen ajattelun opetuksesta ohjelmoinnin avulla peruskoulussa. Hsu, Chang & Hung (2018) tekivät meta-analyysin vuosina 2006–2017 julkaistuista ohjelmoinnillisen ajattelun opetusta ja oppimista koskevista tutkimuksista, joissa he määrittelivät ohjelmoinnillisen ajattelun opetuksessa yleisesti käytetyt aiheet, oppimisstrategiat, opetusvälineet ja ohjelmointikielet. He havaitsivat ohjelman suunnittelun olevan yleisin ohjelmoinnillisen ajattelun opetuksen välittämisen keino ja visuaaliset ohjelmointialustat yleisin opetusväline. Sun, Guo & Hu (2021) tekivät meta-analyysin opetusvälineiden näkökulmasta ja vahvistivat, että opetuspelit olivat tehokkaita ohjelmoinnillisen ajattelun taitojen parantamisessa.

Ohjelmoinnillista ajattelua on jonkin verran tutkittu koulumaailmassa ja nämä tutkimukset ovat osoittaneet, että ohjelmoinnillisen ajattelun opetuksesta on hyötyä kouluissa.

Tutkimuksista on löydetty positiivisia yhteyksiä matematiikan oppimiseen, ohjelmoinnin oppimiseen ja muihin ongelmanratkaisutaitoja vaativiin oppimistilanteisiin. Esimerkiksi Román-Gonzálezin, Pérez-Gonzálezin, Moreno-Leónin ja Roblesin (2018) tutkimuksessa tuloksiksi saatiin, että ohjelmoinnillisen ajattelun opettaminen ennusti positiivisesti matemaattista suoriutumista. Polatin, Hopcanin, Kucukin ja Sismanin (2021) tutkimuksessa mittaria käytettiin arvioimaan ohjelmoinnillisen ajattelun kehittymistä peruskoulun aikana. Guggemoksen, Seufertin ja Román-Gonzálezin (2019) tutkimuksessa CTt (Computational thinking test) osoitti yhteyden ohjelmoinnillisen ajattelun ja ohjelmointitaitojen välillä.

3 Ohjelmoinnillisen ajattelun arviointi

Ohjelmoinnillinen ajattelu on moniulotteinen taito, joka pitää sisällään erilaisia osa-alueita, joten sitä on tärkeää pystyä arvioimaan kokonaisvaltaisesti ja näin validoitujen mittareiden määrää on tarpeen kasvattaa. Kasvattamalla tätä mittareiden määrää paitsi mahdollistamme ohjelmoinnillisen ajattelun kokonaisvaltaisen arvioinnin, myös kehitämme sen opetusta, sillä arvioinnin avulla voidaan verrata eri opetuskäytänteillä saavutettuja oppimistuloksia asetettuihin tavoitteisiin. (Luukkainen-Markkula & Halttunen, 2019; Saari & Silander, 2015.)

3.1 Arvioinnin luotettavuus

Arvioinnin luotettavuutta käsiteltäessä on syytä keskittyä arviointimenetelmien validiteettiin sekä reliabiliteettiin. Kuitenkin jotta voi ymmärtää, mitä validiteetti ja reliabiliteetti ohjelmoinnillisen ajattelun mittareissa tarkoittaa, on ymmärrettävä, mitä mittari oikeastaan tarkoittaa. Mittari on yleensä kokonainen testipatteristo, joka kerää tarkoituksenmukaisesti tietoa tutkittavalta alueelta. Tällöin mittarista puhutaan mittavälineenä (engl. instrument). Mittarin tarkoitus onkin havainnoida jotain ilmiötä mahdollisimman objektiivisesti. (Metsämuuronen, 2009, 67.) Mittarin luotettavuudella on erittäin suuri merkitys myös tutkimusten luotettavuuteen. Metsämuuronen (2009, 74) mainitseekin, että mittarin luotettavuus on suoraan verrannollinen tutkimuksen luotettavuuteen. Kun pohditaan tutkimusten luotettavuutta, kaksi termiä nousevat esille: *reliabiliteetti* sekä *validiteetti*. Reliabiliteetilla tarkoitetaan tutkimuksissa niiden toistettavuutta. Tämä tarkoittaa sitä, että mikäli tutkimuksessa toistettaisiin sama mittaus samasta ilmiöstä, kuinka poikkeavia vastauksia saataisiin. Jos tulokset olisivat melko samanlaisia, eikä ole sattunut systemaattista virhettä eri mittauskerroilla, voidaan todeta, että mittari olisi reliaabeli. (Metsämuuronen 2009, 125.) Toisaalta kun puhutaan mittarin luotettavuudesta validiteetin näkökulmasta, mittarista tarkastellaan mittaako se juuri sitä, mitä sen on tarkoitus mitata (Metsämuuronen 2009, 74).

Mittarin eri luotettavuusominaisuuksia on tässä tutkimuksessa tarkasteltu COSMIN (**C**onsensus based **S**tandards for the selection of health **M**easurement **I**nstrument) taksonomian mukaisesti (Prinsen, Mokkink, Bouter, Alonso, Patrick, De Vet & Terwee, 2018).

3.2 COSMIN -taksonomia validiteetin ja reliabiliteetin tarkastelussa

Kuten Prinsen ym. (2018) esittää, COSMIN taksonomia sisältää yhdeksän mittauksen osatekijää, jotka ovat jaettu kolmeen pääalueeseen: (1) reliabiliteetti, (2) validiteetti ja (3) muutosherkkyys. Nämä mittauksen osatekijät esitellään taulukossa 1.

Taulukko 1.

Määritelmät mittauksen ominaisuuksille COSMIN:n mukaan (Prinsen, ym., 2018).

<i>Pääalue</i>	<i>Mittauksen osatekijä</i>	<i>Mittauksen osatekijän aspekti</i>
Reliabiliteetti	Mittauksen vapaus mittausvirheestä	Sisäinen yhdenmukaisuus (engl. Internal consistency) Miten mittarin eri osiot kykenevät mittaamaan samaa asiaa. Reliabiliteetti (engl. Reliability) Miten luotettavasti ja toistettavasti käytetty mittari mittaa haluttua ilmiötä. Mittausvirhe (engl. Measurement error) Mittauksen systemaattinen ja satunnainen virhe, joka ei johdu mitattavan asian todellisista muutoksista.
Validiteetti	Kyky mitata sitä asiaa, jota sen on tarkoitus mitata	Sisällön validiteetti (engl. Content validity) Mittarin sisällön kyky heijastaa mitattavaa asiaa riittävästi. Koettu Validiteetti (engl. Face validity) Mittarin kyky loogisesti näyttää riittävästi siltä, että se heijastaa mitattavaa asiaa. Rakennevaliditeetti (engl. Construct validity) Mittarin osioiden tulee korreloida keskenään systemaattisemmin kuin muiden muuttujien kanssa, joka perustuu siihen, että instrumentti mittaa pätevästi mitattavaa asiaa. Mittarin rakenteen validiteetti (engl. Structural validity) Mittarin kyky kuvata mitattavan asian eri ulottuvuuksia riittävän tarkasti. Hypoteesien testaus (engl. Hypotheses testing) Mittarin kohteen konstruktiiivin validiteetti Kulttuurien välinen validiteetti (engl. Cross-cultural validity) Mittarin, joka on alun perin kehitetty yhdessä kulttuurissa, kyky olla sovellettavissa ja täten merkityksellinen ja vastaava toisessa kulttuurissa. Kriteerivaliditeetti (engl. Criterion validity) Mittarin arvojen vertaaminen johonkin arvoon, joka toimii validiuden kriteerinä.
Muutosherkkyys (engl. Responsiveness)	Mittarin kyky havaita ajan kuluessa tapahtuvia muutoksia mitattavassa asiassa.	

COSMIN-taksonomian mukaan sisällön validiteetti on tärkein mittarin ominaisuus. Jokaisen mittarin tulee olla relevantti, kattava ja ymmärrettävä kohdejoukossa sekä mitattavan asian kannalta. (Prinsen ym., 2018.) Relevanttius tässä kontekstissa tarkoittaa sitä, että kaikki mittarin kysymykset ovat asianmukaisia sekä mittarin käyttötarkoituksen että kohdejoukon kannalta. COSMIN:n määritelmän mukaan mittauksella on siis kolme eri luotettavuutta lisäävää ominaisuutta. Rakennevaliditeetti (engl. construct validity) ilmaisee sen asteen, jolla mittarin pisteet ovat yhteneväiset hypoteesin testauksen kanssa, joiden mukaan mittari mittaa halutulla tavalla kyseistä käsitettä. Rakennevaliditeetti sisältää mittarin rakenteen validiteetin (engl. structural validity), hypoteesien testauksen (engl. hypothesis testing) ja eri ryhmien välisen validiteetin (engl. cross cultural validity). (Prinsen ym., 2018.)

Cosmin-taksonomian mukaan mittarin rakenteen validiteettia tulisi testata ensisijaisesti faktorianalyysillä käyttäen CTT:ta (Classical test theory). Rakennevaliditeettia mitattaessa suositetaan CFA:a (confirmatory factor analysis) enemmän kuin EFA:a (exploratory factor analysis), sillä se mahdollistaa tarkemman testauksen siitä, onko mittari yhteensopiva tehdyn hypoteesin testauksen kanssa. Molemmilla menetelmillä on paikkansa rakennevaliditeetin arvioinnissa ja molemmat voivat olla hyödyllisiä, riippuen siitä, millainen tutkimusasetelma on kyseessä ja millaisia tavoitteita tutkimuksella on. Lisäksi IRT (Item response theory) auttaa mittarin validiteetin parantamisessa. (Prinsen ym., 2018.)

Taulukossa 1 mittauksen osatekijänä esiteltiin myös reliabiliteetti, johon kuului sisäinen yhdenmukaisuus (engl. internal consistency), joka tarkoittaa sitä, miten hyvin mittarin väittämät mittaavat, sitä mitä on tarkoitus mitata (Prinsen ym., 2018). Mittarin reliabiliteetti (engl. reliability) taas osoittaa sen, kuinka suuri osuus kokonaispistemäärän vaihtelusta johtuu todellisista eroista vastaajien välillä. Mittausvirheellä (engl. measurement error) tarkoitetaan systemaattista tai satunnaista virhettä vastaajan pisteissä, joka ei johdu mitattavan asian todellisista muutoksista. (Prinsen ym., 2018.) Lopuksi taulukossa 1 esitetään mittauksen osatekijä: muutosherkkyys (engl. responsiveness), joka tarkoittaa mittarin kykyä saada kiinni ajan kuluessa tapahtuviin muutoksiin, siten että mittari olisi näiden muutosten jälkeen myös luotettava (Prinsen ym., 2018). Muutosherkkyyttä ei kuitenkaan käsitellä tässä tutkimuksessa, sillä tutkimuksemme ei keskity interventiotutkimuksiin ja niissä tapahtuviin muutoksiin.

Tässä tutkimuksessa käytetään käsitettä psykometriset ominaisuudet, jotka sisältävät nämä edellä mainitut (validiteetti ja reliabiliteetti) luotettavuutta lisäävät asiat. Psykometristen ominaisuuksien avulla tarkistetaan mittarin laadukkuutta ja toimiiko se siten, kuin on

suunniteltu. Mittareista tarkastellaan miten reliabiliteetti ja validiteetti on raportoitu katsauksen tutkimuksissa. Mittareista pyritään löytämään tapoja, joilla mittarin luotettavuutta on pyritty lisäämään. Tällaisia tapoja ovat muun muassa toistomittaus, Cronbachin alfan mittaaminen sekä asiantuntijoiden paneelit ja keskustelut sekä kommentit mittarin toimivuudesta. (Nunnally, 1978; DeVellis, 2017).

4 Tutkimusongelmat

Ohjelmoinnillisesta ajattelusta on julkaistu joitakin tutkimuksia, ja nämä ovat tuottaneet monia hyödyllisiä näkemyksiä, mutta alan kasvava kirjallisuusmäärä viittaa siihen, että tarvitaan kattavampia ja ajantasaisempia katsaustutkimuksia. Näiden avulla voidaan tunnistaa tärkeitä ja esiin nousevia tutkimusaiheita ja –teemoja. Tämän tutkimuksen tavoitteena oli viedä eteenpäin niin ohjelmoinnillisen ajattelun käsitteen tarkentumista kuin myös sen arvioimista ja tätä kautta helpottaa sen integroimista yhä useampiin opetussuunnitelmiin Suomessa ja maailmalla. Olemassa olevien mittareiden psykometrisiä ominaisuuksia ei ole tarkasteltu riittävästi ja tutkimuksen tarkoituksena oli systemaattisesti tarkastella tätä tietoa ja raportoida mittareissa käytetyt testit ja näiden tuloksina saadut reliabiliteetit ja validiteetit. Erään ohjelmoinnillisen ajattelun mittareista laaditun kirjallisuuskatsauksen perusteella suurin osa mittaamisesta keskittyi oppilaiden ohjelmointitaitoihin ja mittareiden psykometriset ominaisuudet, validiteetti ja reliabiliteetti, jäivät yli puolessa (55 %) täysin mittaamatta tai raportoimatta. (Tang, Yin, Lin, Hadad & Zhai, 2020) Näin ollen on tärkeää tutkia ohjelmoinnillisen ajattelun mittareiden psykometrisiä ominaisuuksia, jotta niitä voidaan kehittää jatkossa yhä luotettavimmiksi.

Tutkimuksen tarkoituksena oli kartoittaa, systemaattisen kirjallisuuskatsauksen avulla, minkälaisia mittareita on käytetty ohjelmoinnillisen ajattelun taitojen arviointiin alakoulussa sekä tutkia näiden olemassa olevien ohjelmoinnillisen ajattelun mittareiden psykometrisiä ominaisuuksia. Näin ollen tutkimuksen 1. pääongelmana oli selvittää, millaisia mittareita on olemassa ja 2. pääongelmana oli tutkia näiden mittareiden psykometristen ominaisuuksien raportointia, ja tämä ongelma jakautui kahdeksi alaongelmaksi validiteetin ja reliabiliteetin mukaan:

1. Millaisia mittareita on käytetty ohjelmoinnillisen ajattelun taidon arviointiin alakoulussa?
2. Mitä näiden mittareiden psykometrisistä ominaisuuksista on raportoitu?
 - 2.1. Mitä mittareiden validiteetista on raportoitu?
 - 2.2. Mitä mittareiden reliabiliteetista on raportoitu?

Näiden kysymysten avulla pyrittiin saamaan mahdollisimman kattava kuva ohjelmoinnillisen ajattelun arvioinnista ja mittareiden psykometrisistä ominaisuuksista tämän hetken tieteellisessä kirjallisuudessa.

5. Tutkimuksen toteuttaminen

Tutkimus toteutettiin systemaattisen kirjallisuuskatsauksen avulla, koska haluttiin saada laaja ja järjestelmällinen kuva aiheesta julkaisusta tutkimuksesta. Lisäksi systemaattinen kirjallisuuskatsaus antaa tutkimukselle luotettavuutta ja tekee siitä tieteellisesti perustellun, koska systemaattista kirjallisuuskatsausta varten luodaan tutkimukselle oleellisia kriteerejä, joiden avulla valitaan vain merkittäviä tutkimuksia. (Kallio 2006, 19.) Bearfield ja Eller (2008, 62) totesivat systemaattisen kirjallisuuskatsauksen hyväksi puoliksi juuri runsaan tieteellisten artikkeleiden läpikäymisen, kattavan tarkastelun sekä sen tiivistämisen kirjallisuuskatsaukseksi, josta lukijan on helppo perehtyä laajan tutkimusalueen merkittävimpiin seikkoihin. Systemaattisella kirjallisuuden läpikäynnillä pyritään saamaan mahdollisimman kokonaisvaltainen kuva ohjelmoinnillisen ajattelun mittareista ja luomaan kriteerejä, jotka tuovat tutkimukselle tieteellistä uskottavuutta. (Dixon-Woods, Booth & Sutton, 2007, 375). Kirjallisuuskatsaus tutkimustyyppinä on ollut suosittu menetelmä ohjelmoinnillisen ajattelun tutkimuksessa, sillä selkeän ohjelmoinnillisen ajattelun määritelmän puutteen myötä aikaisempien tutkimusten pohjalle nojaava tutkimus on välttämätöntä. (Winchester & Salji, 2016) Tässä tutkimuksessa kirjallisuuskatsauksen toteuttaminen toteutettiin Finkin mallin mukaisesti. Malli tarjoaa jäsennellyn lähestymistavan kirjallisuuskatsauksen tekemiseen, jossa tutkitaan laajasti aihetta käsittelevää kirjallisuutta. Lisäksi Finkin malli edellyttää johdonmukaista ja systemaattista lähestymistapaa, joka auttaa tutkijoita tekemään kattavan kirjallisuuskatsauksen. (Fink, 2005.)

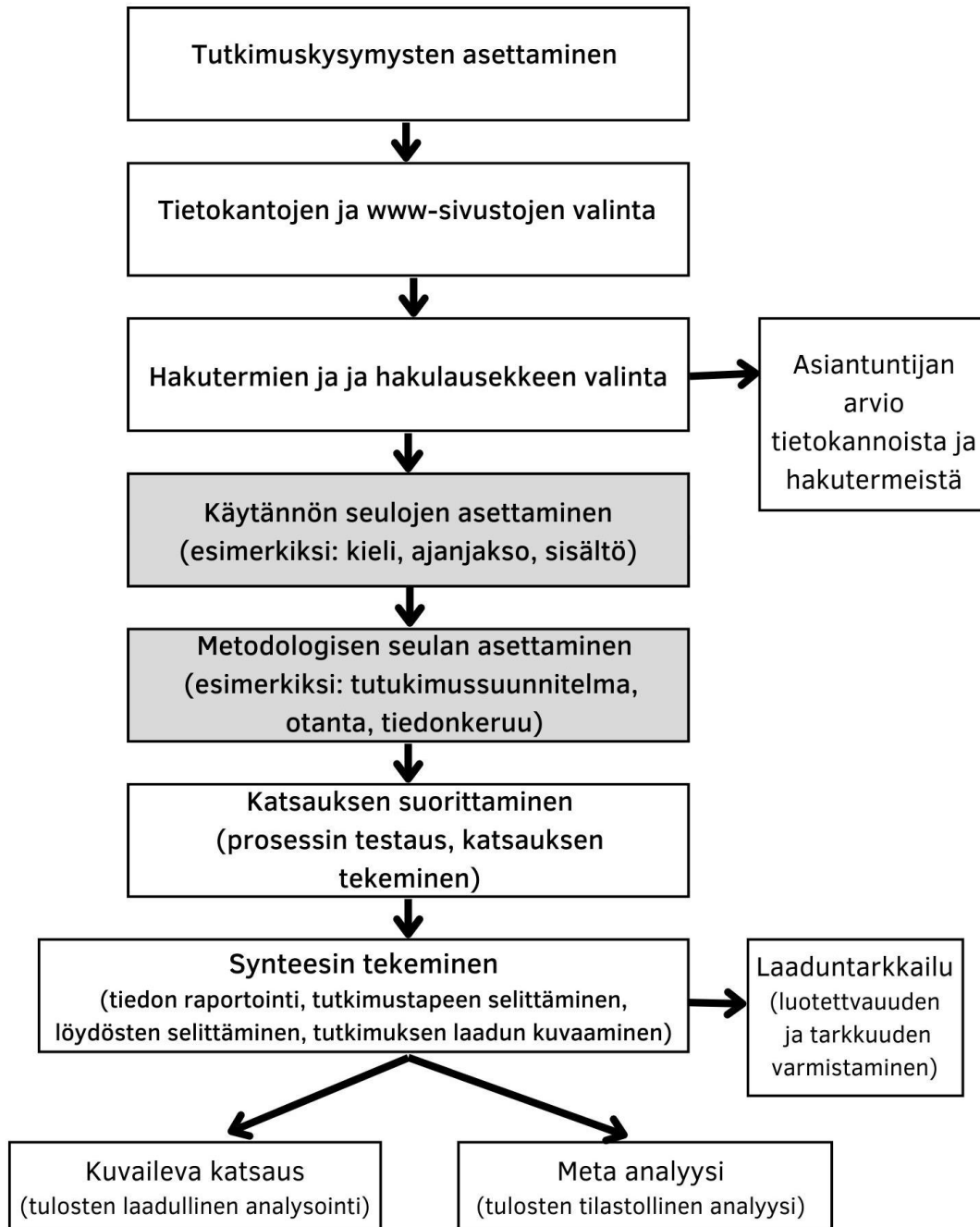
5.1 Finkin malli systemaattisen kirjallisuuskatsauksen kehikkona

Tämä tutkimus perustui Finkin (2005, 3–5) seitsenvaiheiseen malliin systemaattisen kirjallisuuskatsauksen toteuttamisesta (kuvio 3). Mallin ensimmäiset kolme vaihetta nivoutuvat yhteen, sillä niiden avulla valitaan mistä kirjallisuutta ja millä hakusanoilla tai -lausekkeilla kirjallisuutta haetaan. Ensimmäisessä vaiheessa tutkijat laativat tutkimuskysymyksen tai -kysymykset, jotka ohjaavat kirjallisuuskatsauksen toteuttamista. Toinen vaihe sisältää tietokantojen sekä mahdollisten www-sivustojen valinnan. Kolmannessa vaiheessa valitaan sopiva hakulauseke, jolla etsitään sopivia artikkeleita. Tässä vaiheessa on tärkeää rajata hakutermit ja –lausekkeen siten, että ne kattavat jokaisen oleellisen osan

tutkittavasta tutkimuskentästä. Monia hakutermejä kokeillaan parhaan mahdollisen lopputuloksen saavuttamiseksi. Kolmanteen vaiheeseen kuuluu myös asiantuntijan arvio tietokannoista sekä hakulausekkeista. (Fink, 2005, 3–5.)

Neljäs ja viides vaihe sisältää artikkelin rajaamisen seulojen avulla. Neljännen vaiheen rajauksiin kuuluvat omaan kirjallisuuskatsaukseen sopivien kielten, ajanjakson sekä sisällön valitseminen ja muiden pois rajaaminen. Tässä vaiheessa rajautuu yleensä jo suuri osa artikkeleista pois. Viidennessä vaiheessa rajataan kaikki artikkelit pois, jotka eivät sovellu tutkimuksen otantaan, tiedonkeruuseen eikä tutkimuskysymyksiin. Tavoitteena on arvioida kirjallisuuden laatua ja soveltuvuutta tutkimukseen, jotta saadaan mahdollisimman laadukas materiaali. Tässä vaiheessa tutkijat rajaavat ensin tiivistelmien ja sitten koko artikkelien lukemisen jälkeen pois ne, mitkä eivät sovellu heidän kirjallisuuskatsaukseensa. (Fink, 2005, 3–4.)

Viimeiset kaksi vaihetta sisältävät sekä katsauksen että synteesin tekemisen. Kuudes vaihe sisältää prosessin testauksen sekä katsauksen tekemisen. Tarvittaessa koulutetaan mahdolliset avustajat sekä tehdään pilottiharjoite ja arvioidaan se. Lisäksi tutkijoiden on hyvä valvoa katsauksen suorittamisen koko prosessi, jotta katsauksen luotettavuus ei kärsi. Viimeisenä vaiheena jäljellä on synteesin tekeminen eli analysointi. Kuten kuvio 3 esittää, synteesin tekeminen jakautuu kahteen osaan (meta-analyysin tekeminen sekä tulosten laadullinen analysointi), joista tutkijat valitsevat itselleen sopivimman lähestymistavan. (Fink, 2005, 3–4.) Mahdollista on myös jäädä synteesin tekemisen vaiheeseen, jolloin raportoidaan löydetyt tulokset, kuten tässä tutkimuksessa on tehty. Mallin mukaan synteesin tekemisen vaiheeseen kuuluu myös laadun tarkkailu, jonka avulla pyritään lisäämään tutkimuksen luotettavuutta. (Fink, 2005, 3–4.)



Kuvio 3. Kirjallisuuskatsauksen vaiheet (Fink 2005, 54). Kirjallisuuden rajaaminen abstraktien ja koko tekstin perusteella kuvattu harmaalla värillä.

5.2 Tutkimustehtävä ja tutkimusaineiston hankinta

Kuten Finkin mallissa, myös tässä tutkimuksessa systemaattinen kirjallisuuskatsaus alkoi tutkimuskysymysten laatimisella. Tutkimuksen pääongelmana oli selvittää, millaisia ohjelmoinnillisen ajattelun mittareita on olemassa ja tutkia näiden mittareiden psykometristen ominaisuuksien raportointia.

Tutkimuksen tietokannan valinta ohjautui EBSCOHost tietokanta-alustalle, joka kerää eri tietokantojen kirjallisuutta yhteen paikkaan. Monet testihaut, joita tietokannan valitsemisvaiheessa kokeiltiin, osoittautuivat käyttökelvottomaksi, sillä ne eivät sisältäneet tarpeeksi tutkimuksen kannalta oleellisia kirjallisuuksia. EBSCOHost –alustalta valittiin seuraavat kasvatusalan tietokannat tutkimuskohteiksi: Teacher Reference Center, Academic Search Premier, eBook Collection (EBSCOhost), ERIC sekä Education Source.

Kuten Finkin mallin kolmannessa vaiheessa, seuraavaksi lähdettiin muodostamaan hakulauseketta. Hakulauseke muotoiltiin englannin kielellä, sillä tutkijoiden kielitaito oli riittävää vain englannin- ja suomenkielisten artikkeleiden tarkasteluun. Testihakujen perusteella suomenkielisiä artikkeleita ei löytynyt EBSCOHost – alustalta ja tästä syystä englannin kieli valikoitui tämän tutkimuksen pääkieleksi. Avainsanoja tutkimuksessa olivat ohjelmoinnillinen ajattelu, mittari, peruskoulu. Näitä käsitteitä pyrittiin laajentamaan ja etsimään niille synonyymejä, jotta jokainen keskeinen artikkeli saataisiin tarkasteluun. Vapaasti suomennettuna mukaan tulivat seuraavat käsitteet: mittarin rinnalle tuli arviointi, ja peruskoulun rinnalle taas alakoulu, ala-aste sekä alkuopetus. Lopullinen hakulauseke muodostui seuraavanlaiseksi:

"computational thinking" AND (assessment OR measure OR evaluation) AND ("primary school" OR "elementary education" OR "primary education" OR "elementary school" OR "basic education").

Hakulausekkeen muodostamisvaiheessa hakusanoja ja lauseketta muokattiin ja kokeiltiin eri tietokantoihin erilaisissa muodoissa, jotta tutkimukseen saatiin mahdollisimman laaja kattaus keskeistä kirjallisuutta parasta mahdollista tulosta ajatellen.

5.3 Tutkimusaineiston valintakriteerit ja -prosessi

Seuraavaksi artikkeleista rajattiin tutkimukseen soveltumatonta kirjallisuutta pois.

Hakulausekkeella tuli tietokannasta yhteensä 185 artikkelia. Tutkimukseen haluttiin vain laadukkainta mahdollista tieteellistä kirjallisuutta, joten pois rajattiin kaikki artikkelit, jotka eivät olleet vertaisarvioituja (n=17). Tämän jälkeen artikkeleita oli jäljellä 168. Lisäksi kirjallisuuden tuli olla julkaistu viimeisen kymmenen vuoden aikana, mutta tämä ei rajannut yhtään artikkelia pois. Tämä voi johtua siitä, ohjelmoinnillisen ajattelun mittarit ovat suhteellisen uusi aihe tieteellisessä kirjallisuudessa. Hakutuloksista poistettiin myös kaksoiskappaleet (n=19), joita tuli mukaan, sillä haku toteutettiin useista eri tietokannoista EBSCOHost –alustan avulla. Vielä ennen tarkempaa tarkastelua pois rajattiin kaikki ne, jotka eivät olleet englanninkielisiä (n=2). Lopulta tarkempaan tarkasteluun jäi 147 artikkelia.

Jäljelle jääneet artikkelit vietiin Exceeliin tarkempaa analyysia varten. Molemmat tutkijat kävivät alla esitetyt valintakriteerit (A, B ja C) ensin tiivistelmien suhteen läpi ja merkkasivat NO, MAYBE tai YES, sen mukaan, mitä tutkijan mukaan tiivistelmät kertoivat tutkimuksesta. Mikäli molemmat olivat samaa mieltä, artikkeli joko hylättiin pois tai hyväksyttiin koko tekstin tarkasteluun. Jos tutkijat olivat eri mieltä, artikkelien abstrakteista keskusteltiin mukana olleen kolmannen tutkijan kanssa ja kyseiset artikkelit arvioitiin yhdessä.

Koko tekstin arviointivaiheeseen valikoitui tiivistelmien perusteella 22 artikkelia ja pois rajautui 125 artikkelia. Molemmat tutkijat lukivat jälleen kyseiset artikkelit erikseen ja arviointiin samojen kysymysten (A, B ja C) perusteella NO, MAYBE tai YES. Lopulta koko tekstin perusteella pois rajautui yhdeksän artikkelia ja tutkimukseen valikoitui 9 artikkelia, jotka sisälsivät 12 mittaria. Koko tapahtumasarja on esillä prosessikaaviossa kuviossa 4.

A) Käsitteliikö artikkeli ohjelmoinnillista ajattelua?

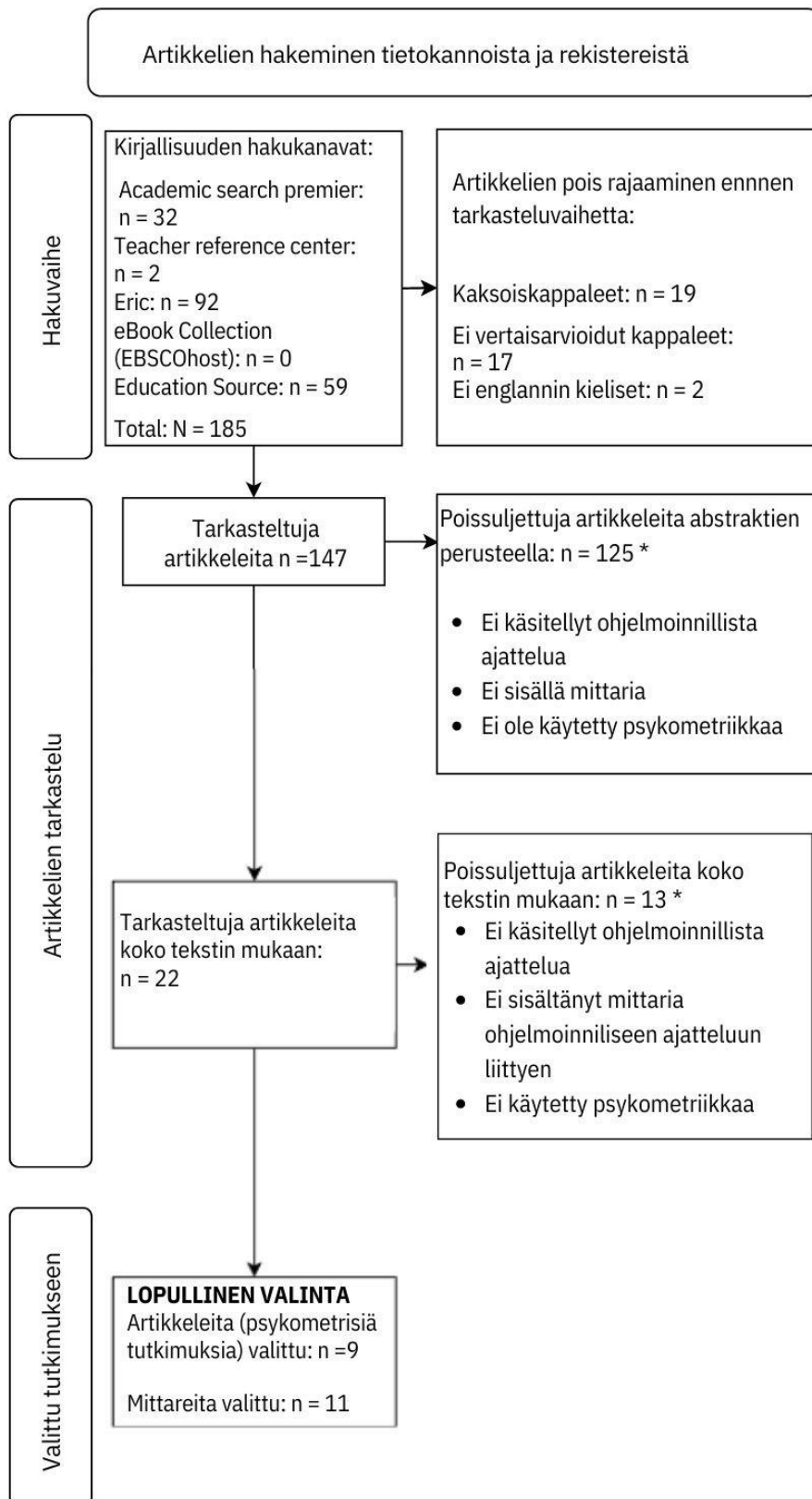
Tämän kysymyksen avulla tarkasteltiin, että käsitteliikö tutkimus ylipäättään ohjelmoinnillista ajattelua, vai esimerkiksi ohjelmointia, robotiikkaa tai ongelmanratkaisua. Tutkimukselle oli keskeistä löytää juuri ne artikkelit, joissa ohjelmoinnillinen ajattelu oli keskiössä.

B) Onko artikkelissa mittaria, joka mittaa juuri ohjelmoinnillista ajattelua?

Kriteerin tarkoituksena oli löytää sellaiset tutkimukset, joissa ohjelmoinnillista ajattelua oli mitattu ja joista löytyi selkeä mittari ohjelmoinnillisen ajattelun mittaamiseen. Tässä vaiheessa tallennettiin myös ohjelmoinnillisen ajattelun mittareiden nimet jatkoa varten.

C) Onko artikkelissa käsitelty arviointivälineen psykometrisiä ominaisuuksia

Tämän kysymyksen avulla haluttiin varmistua, että aineistoon mukaan otettavissa tutkimuksissa on kyse käytetyn arviointivälineen psykometrinen ominaisuuksien arvioinnista. Psykometrinen ominaisuuksien avulla tarkistetaan mittarin laadukkuutta ja toimiiko se siten, kun sen pitää.



Kuvio 4. Prosessikaavio: artikkelien valitseminen tietokannoista ja rekistereistä

* Poissuljettujen artikkeleiden määrää ei yksilöity, sillä osassa artikkeleista oli puutteita useammassa kuin yhdessä sisäänottokriteerissä.

Tutkimuksen aineisto valittiin erilaisia sisäänotto- ja poissulkukriteereitä käyttäen. Näiden kriteerien avulla pyrittiin välttämään systemaattisia virheitä ja varmistamaan, että aineisto oli oleellinen tutkimuskysymysten kannalta. Hyväksymiskriteerit määrittivät, mitä ominaisuuksia artikkelissa piti olla, jotta se soveltuu tutkimukseen, kun poissulkukriteerit määrittivät ne ominaisuudet, jotka estivät artikkelin mukaan ottamisen tutkimukseen. (Pudas-Tähkä & Axelin 2007, 48.) Artikkelin sisäänotto- ja poissulkukriteerit ovat nähtävissä taulukossa 2.

Taulukko 2.

Katsauksen sisäänotto- ja poissulkukriteerit

Sisäänottokriteerit	Poissulkukriteerit
<ul style="list-style-type: none"> • Artikkelin on vertausarvioitu julkaisu. • Artikkelin on julkaistu vuosien 2013–2022 aikana. • Artikkelin on julkaistu englannin tai suomen kielellä. • Artikkelin käsittelee ohjelmoinnillista ajattelua. • Artikkelissa on mittari, joka mittaa ohjelmoinnillista ajattelua. • Mittarin psykometrisiä ominaisuuksia on raportoitu artikkelissa. 	<ul style="list-style-type: none"> • Artikkelia ei ole vertaisarvioitu. • Artikkelin on julkaistu ennen vuotta 2013. • Artikkelissa käsitellään robotiikkaa, ohjelmointia tai ongelmanratkaisua, mutta ei ohjelmoinnillista ajattelua. • Artikkelissa ei ole mittaria, joka arvioi ohjelmoinnillista ajattelua. • Mittarin psykometrisiä ominaisuuksia ei ole raportoitu artikkelissa.

5.4 Tutkimusaineiston analyysi

Tässä tutkimuksessa systemaattisen kirjallisuuskatsauksen tarkoituksena oli tarkastella ohjelmoinnillisen ajattelun erilaisia mittareita sekä niiden ominaisuuksia. Tämän tutkimuksen kannalta oleelliset tieteelliset artikkelit (n=9) on esitelty taulukossa 3. Ne valittiin laajan ja tarkan seulan avulla, joka esitettiin kuviossa 4.

Taulukko 3.

Yhteenveto artikkeliin valikoiduista tutkimuksista

<i>Tutkijat</i>	<i>Julkaisu vuosi</i>	<i>Nimi</i>
Basu, Rutstein, Xu, Wang & Shear	2021	A Principled Approach to Designing Computational Thinking Concepts and Practices Assessments for Upper Elementary Grades
Chen, Shen, Barth-Cohen, Jiang, Huang & Eltoukhy	2017	Assessing elementary students' computational thinking in everyday reasoning and robotics programming.
Gane, Israel, Elagha, Yan, Luo & Pellegrino	2021	Design and Validation of Learning Trajectory-Based Assessments for Computational Thinking in Upper Elementary Grades
Kong & Wang	2021	Item response analysis of computational thinking practices: Test characteristics and students' learning abilities in visual programming contexts.
Li, Xu & Liu	2021	Development and Validation of Computational Thinking Assessment of Chinese Elementary School Students.
Relkin, de Ruiter & Bers	2020	'TechCheck': Development and Validation of an Unplugged Assessment of Computational Thinking in Early Childhood Education
Tsarava, Moeller, Román-González, Golle, Leifheit, Butz & Ninaus	2022	A cognitive definition of computational thinking in primary education.
Zapata-Cáceres, Martín-Barroso & Román-González	2021	Collaborative Game-Based Environment and Assessment Tool for Learning Computational Thinking in Primary School: A Case Study
Zhong, Wang, Chen & Li	2016	An Exploration of Three-Dimensional Integrated Assessment for Computational Thinking

Näiden artikkeleissa esiintyvien mittareiden ominaisuuksia analysoitiin sisältötasolla. Aluksi tarkasteltiin mittareiden kohderyhmä, mittareissa hyödynnetyt tehtävät ja tehtävyyttypit sekä työskentelyvälineet, kuten tietokone tai kynä ja paperi. Yleisen tarkastelun jälkeen keskityttiin mittareista raportoituihin psykometrisiin ominaisuuksiin, reliabiliteetti ja validiteetti.

Mittareiden validiteettia analysoitiin sisällön validiteetin ja rakennevaliditeetin näkökulmista Cosmin-taksonomian mukaan. Tutkijat jakoivat asiat, kuten testipilotti ja asiantuntijoiden paneelikeskustelut sisällön validiteettia tukeviin asioihin, sillä näiden avulla Cosmin-taksonomian mukaan tutkimuksissa pyritään vahvistamaan mittarin kykyä heijastaa mitattavaa asiaa riittävästi eli parantamaan mittarin sisällön validiteettia. Rakennevaliditeettia tukeviin asioihin luokiteltiin erilaiset faktorianalyysit ja näistä saadut tulokset sekä rinnakkaisvaliditeettia tai kriteerivaliditeettia vahvistavat korrelaatioanalyysit ja näiden tulokset.

Mittareiden reliabiliteettia analysoitiin Cosmin-taksonomian mukaan sisäisen johdonmukaisuuden ja reliabiliteetin kautta. Sisäisestä johdonmukaisuudesta tarkasteltiin ja raportoitiin mittareiden Cronbachin alfa-kertoimet sekä muut vastaavat mittarin eri osioiden

välistä yhdenmukaisuutta mittaavat asiat. Reliabiliteetin tarkastelussa pyrittiin löytämään raportointeja mittarin mittausten erojen osuudesta kokonaisvarianssista.

Aineiston analyysin päätavoitteena oli siis tarkastella mittareiden psykometristen ominaisuuksien raportoituja mittauksia ja tuloksia Cosmin-taksonomian reliabiliteetti- ja validiteettiosatekijöiden mukaan, jotka näkyvillä jo Taulukossa 1.

6 Tutkimustulokset

Tutkimukseen valikoitui 185 artikkelin joukosta 9 artikkelia, jotka sisälsivät 11 mittaria. Nämä mittarit avataan seuraavaksi ja tarkastellaan niiden ominaisuuksia. Mittareista esitellään tulososiossa keskeisimmät tiedot, kuten esimerkiksi: mille ikäryhmälle mittari on suunniteltu, arviointitehtävät sekä niiden kuvaukset sekä työskentelyvälineet mittarissa. Tulososiossa tarkastellaan myös mittareiden validiteetti- sekä reliabiliteettiominaisuuksia, sillä ne osoittavat mittareiden toimivan tarkoitetulla tavalla ja olevan luotettavia.

6.1 Tutkimuksissa käytetyt mittarit

Tutkimukseen valikoitui siis 11 mittaria, jotka esitellään seuraavaksi Taulukossa 3. Taulukon tarkoituksena on esitellä yhteenveto kirjallisuuskatsauksen mittareista. Ensimmäisenä taulukossa esitetään mittarin nimi, lyhenne sekä mittarin tutkijat sekä julkaisuajankohta. Tämän jälkeen taulukossa 4 esitetään kunkin mittarin kohdevuosiluokka sekä -ikäryhmä. Seuraavassa kohdassa on esillä arviointitehtävät sekä tehtäväkuvaukset, jotka osoittavat, millaisia tehtäviä mittarit sisälsivät ja kuinka paljon tehtäviä mittarit sisälsivät. Taulukon viimeisessä osassa esitellään vielä mittareiden työskentelyvälineet sekä pisteytys.

Taulukko 4.

Yhteenveto kirjallisuuskatsauksen sisältämistä mittareista

<i>Lyhenne Mittarin nimi (Tutkijat, julkaisuajankohta)</i>	<i>Vuosiluokka Ikäryhmä</i>	<i>Arviointitehtävät (määrä n): tehtäväkuvaus</i>	<i>Työskentelyvälineet Pisteytys</i>
Abbreviated CTt Abbreviated Computational Thinking test (Tsarava, ym., 2022)	Luokat 3–4 8–10-vuotiaat	Seitsemän ohjelmoinnillisen ajattelun käsitteen testi (n=21): Testi esitetään "Pac-Man"- tyyllisen sokkelon muodossa, ja vastaajan tulee valita oikea visuaalinen nuoli tai ohjelmointilohko, joka sopii tehtävään	Työskentelyvälineet: kynä ja paperi Pisteytys: 0 = väärin, 1 = oikein
BCTt The beginners CT test (Zapata-Caceres, ym., 2021)	Luokat 1–6 5–12-vuotiaat	Monivalintatehtäviä (n=25): kysymyksissä kolme vaihtoehtoa ja yksi oikea vastaus. Testi jaettu 6 osaan: ("1–6: sequences; 7–11: simple loop; 12–18: nested loop; 19–20: if- then; 21–22: if then-else; 23–25: while")	Työskentelyvälineet: kynä ja tulostetut testit Pisteytys: 0 = väärin, 1 = oikein

CTA-CES Computational Thinking Assessment for Chinese Elementary Students	Luokat 3–6 9–12-vuotiaat	Ongelmanratkaisutehtäviä (n=25): Monivalintakysymyksiä arkielämän ongelmista, joissa yksi oikea vastaus	Työskentelyvälineet: tietokoneet Pisteytys: 0 = väärin, 1 = oikein
(Li, ym., 2021)			
CTC CT concepts assessment instrument	Luokat 4-6 9-11 vuotiaat	Monivalintatehtäviä: (n= 19) kolmella eri tasolla. 4 vastausvaihtoehtoa, yksi oikea vastaus. Testi jaettu 4 osaan: (" CTC1: repetition, 6 with CTC2: conditionals, 4 with CTC3: control flow, and 5 tasks aligned with CTC4: data structures")	Työskentelyvälineet: tietokoneet Pisteytys: 0 = väärin, 1 = oikein
(Basu, ym. 2021)			
CTP CT practices assessment instrument	Luokat 4-6 9-11 vuotiaat	Monivalintatehtäviä: (n= 12) 4 vastausvaihtoehtoa, yksi oikea vastaus. Testi jaettu 4 osaan:	Työskentelyvälineet: tietokoneet Pisteytys: ER
(Basu, ym. 2021)			
CT practices test	Luokat 4-6 ER	Monivalintakysymyksiä (n=12): yhteensä kysymykset sisälsivät 47 tehtävää. Valitse oikea vaihtoehto, useita oikeita vaihtoehtoja tai raahaa ja pudota tehtäviä.	Työskentelyvälineet: tietokoneet Pisteytys: ER
(Kong & Wang, 2021)			
Grade 3 Early assessment	Luokat 3-4 ER	Avoimet kysymykset (n=10): kysymykset ohjelmoinnista. Esimerkiksi "käytä toista kolme kertaa kommentoa, jotta tehtävä toteutuu"	Työskentelyvälineet: Kynä ja tulostetut testit/ Scratch Pisteytys: ER
(Gane, ym. 2021)			
Grade 4 Early assessment	Luokat 3-4 ER	Avoimet kysymykset (n=10): kysymykset ohjelmoinnista.	Työskentelyvälineet: Kynä ja tulostetut testit/ Scratch Pisteytys: ER
(Gane, ym. 2021)			
Instrument	Luokka 5 10-11 vuotiaat	Monivalintakysymykset (n= 8) sekä avoimet kysymykset (n= 15): Arkielämän ongelmanratkaisutilanteita sekä ohjelmointiin liittyviä tehtäviä.	Työskentelyvälineet: Kynä ja tulostetut testit Pisteytys: 0 = väärin, 1 = oikein (monivalintakysymykset) 0-2 (avoimet kysymykset)
(Chen, ym. 2017)			
TechCheck	Luokat 0-3 5-9-vuotiaat	Monivalintatehtäviä (n=15): lyhimmän polun palapelit, puuttuvat symbolisarjat, sokkelot, symboliarvoitukset ja symmetriaongelmat.	Työskentelyvälineet: tietokoneet Pisteytys: 0 = väärin, 1 = oikein
(Relkin, ym. 2020)			
TDIA Three-Dimensional Integrated Assessment	Luokat 6 ER	Koodaustehtäviä (n=6): 6-vaiheinen koodausongelma, suljettuja, puolittain-suljettuja ja avonaisia	Työskentelyvälineet: ohjelmisto Pisteytys: tehtävät 1-4 asteikolla 0-5 tehtävät 5 ja 6 asteikolla 0-20
(Zhong, ym. 2016)			
ER = Ei raportoitu			

Taulukosta 4 tulee esille, että mittareita on käytetty pääosin alakoulun 3.–6. luokkalaisia oppilaille (n= 9). Poikkeuksen muodostivat Relkinin ym. (2020) TechCheck –mittari, joka kohdistui esikouluikäisistä kolmanteen vuosiluokkaan asti, sekä Zapata-Caceresin ym. (2021) BCTt, jota käytettiin kaikilla alakouluikäisillä oppilaille.

Mittareiden yleisin tehtävätyyppi oli monivalintatehtävät (n= 7), jotka mittasivat muun muassa ohjelmoinnin perusasioita sekä arkielämään liittyviä ongelmanratkaisutilanteita.

Monivalintatehtävien määrä vaihteli mittareiden välillä (n= 8–25). Vaikka Chenin ym. (2017) mittarissa oli vain kahdeksan monivalintakysymystä, siinä oli niiden lisäksi avoimia kysymyksiä (n=15). Avoimia kysymyksiä hyödynnettiin myös Ganen ym. (2021) Grade 3 early assesment ja Grade 4 early assesment –mittareissa, joissa molemmissa avoimia kysymyksiä oli 10 kappaletta. Ganen ym. mittareissa kysyttiin esimerkiksi, kuinka joku ohjelmointi toteutuu. Esimerkiksi tehtävässä saatettiin kysyä seuraavasti “Andrella on 9 keksiä, jotka hän haluaa jakaa kolmelle kaverilleen. Hän haluaa antaa kaikille yhtä monta keksiä. Kirjoita ohjeet keksien jakamiseen kolmelle kaverille. Varmista, että jokainen kaveri saa yhtä monta keksiä. Käytä komentoa “toista 3 kertaa” ainakin yhden kerran”. Avoimet kysymykset siis antoivat vastaajan päättää, miten toteuttaa tehtävä, kuitenkin niin, että tehtävän komento tulee osaksi vastausta. Tutkimukseen valikoiduista artikkeleista löytyi myös pelinomaisia mittareita. Esimerkiksi Tsaravan ym. (2022) mittarissa oppilaan tuli kulkea pacmantyyllisen sokkelon läpi komentoja antamalla.

Taulukosta 4 voidaan havaita suosituimman työskentelyvälineen olevan tietokone/tablet tietokone, joita käytettiin yhdeksässä mittarissa. Perinteistä kynää ja tulostettuja vastauslomakkeita käytettiin puolestaan viidessä mittarissa. Ganen ym. (2020) mittarissa käytettiin molempia, sillä teknologisia laitteita käytettiin Scratch -ohjelman ohjelmointitehtävissä. Pisteytys mittareissa oli useimmiten Väärin = 0 p. ja oikein = 1 p. Basu ym. (2021) ja Gane ym. (2020) mittareissa ei ollut raportoitu pisteytystä.

6.2. Mittareiden validiteetti

Taulukkoon 5 on tutkimuksissa mittareista raportoidut validiteettiarvot. Ensimmäisenä taulukossa esitetään mittarin nimi, lyhenne sekä mittarin tutkijat sekä julkaisuajankohta. Seuraavana sisällön validiteettia mittaava menetelmä ja sen tulos ja lopuksi rakennevaliditeetti, joka on jaettu mittarin rakenteen validiteettia mittaavaan menetelmään ja sen tulokseen sekä mittarin hypoteesia testaavaan menetelmään ja sen tulokseen.

Taulukko 5.

Yhteenveto artikkelien raportoiduista validiteettimuuttujista

<i>Mittari (Tutkijat, julkaisuajank ohta)</i>	<i>Sisällön validiteetti</i>		<i>Rakennevaliditeetti</i>			
	<i>Menetelmä</i>	<i>Tulokset</i>	<i>Mittarin rakenteen validiteetti</i>		<i>Hypoteesin testaaminen</i>	
			<i>Menetelmä</i>	<i>Tulokset</i>	<i>Menetelmä</i>	<i>Tulokset</i>
Abbreviated CTt Abbreviated Computational Thinking test (Tsarava, ym., 2022)	Asianmukaisuus: Muokattu aiemmin validoidusta testistä (CTt)	ER	EFA	KMO = 0.513	Rinnakkaisvaliditeetti	Kohtalainen positiivinen korrelaatio KFT <i>Vocabulary</i> subtest: ($r = 0.388$, $n = 191$, $p < .001$) Merkittävä positiivinen korrelaatio CFT <i>Series Continuation</i> and <i>Matrices</i> subtests: ($r = 0.346$, $n = 192$, $p < .001$) the HRT <i>Problem Completion</i> subtest ($r = 0.333$, $n = 192$, $p < .001$), the HRT <i>Multiplication</i> subtest ($r = 0.224$, $n = 192$, $p = .002$), and CALC performance (Calculation abilities reflecting performance on HRT subscales <i>Addition</i> and <i>Subtraction</i> , $r = 0.232$, $n = 192$, $p = .001$).
BCTt The beginners CT test (Zapata- Caceres, ym., 2021)	ER	ER	ER	ER	ER	ER
CTA-CES Computational Thinking Assessment for Chinese Elementary Students (Li, ym., 2021)	Asianmukaisuus: 8 asiantuntijan kommentointi ja 18 alakoululaisen sekä 6 yliopisto-opiskelijan pilottitestaus	ER	IRT, CTT	($p = 0.56$) keskiarvo 0.16, -2.03 – 3.85 .	Kriteerivaliditeetti	CTA-CES korreloi positiivisesti päättelykyvyn (engl. reasoning ability), $r(71) = 0.47$, $p < 0.001$, avaruudellinen hahmotus (engl. spatial ability), $r(71) = 0.43$, $p < 0.001$, ja luetun ymmärtäminen (engl. reading comprehension), $r(71) = 0.54$, $p < 0.001$
			The point-biserial correlation	Yli 0.30		
CTC CT concepts assessment instrument (Basu, ym. 2021)	Kokonaisvaltaisuus, asianmukaisuus: 5 asiantuntijan tarkastelu, kommentointi ja palautteen anto. Tehtäviä muokattiin asiantuntijoiden tarkastelun perusteella lisätäkseen	ER	IRT: Lomakkeen validointia varten osalle oppilaista ($n = 1808$) annettiin yksi CT-practices vastauslomake, jossa oli kaikki 12 tehtävää,	IRT: ER	ER	ER

	niiden selkeyttä, luettavuutta ja yhteensovittamista FKSA:n (“focal knowl-edge, skills and abilities”) kanssa.		ja heille annettiin kaksinkertaisesti niin paljon aikaa kuin muille oppilaille.			
			Faktorianalyysi	One factor analysis: RMSEA = 0.045		
CTP CT practices assessment instrument (Basu, ym. 2021)	Kokonaisvaltaisuus, asianmukaisuus: 5 asiantuntijan tarkastelu, kommentointi ja palautteen anto. Tehtäviä muokattiin asiantuntijoiden tarkastelun perusteella lisätäkseen niiden selkeyttä, luettavuutta ja yhteensovittamista FKSA:n (“focal knowl-edge, skills and abilities”) kanssa.	ER	IRT,	IRT: ER	Rinnakkaisvaliditeetti Four-factor analysis	Vahva korrelaatio (.62 to .84) yksittäisten faktorien välillä Keskinkertainen korrelaatio coefficient alpha = 0.45 Oppilaiden kyvyissä suorittaa CTP;n tehtäviä
			Faktorianalyysi	One factor analysis: RMSEA = 0.06 tai 0.062		
CT practices test (Kong & Wang, 2021)	Keskustelut asiantuntijoiden kesken ECD “Evidence-centered design”	ER	MIRT	bi-factor $\Delta-2\ln L = 411.59$, $\Delta df = 28$, $p < .001$ AIC = 474346.1 BIC = 475409.8 $\chi^2 = 1092.31$, $df = 560$, CFI = 0.95, TLI = 0.93, RMSEA = 0.04	ER	ER
Grade 3 Early assessment (Gane, ym. 2021)	Arvioinnin kolmio ECD Testipilotti	ER	IRT, CTT	Anderson LR-test, $\chi^2(6) = 11.14$, $p = .08$ B = -1.48–1.54 CTT: ER		
Grade 4 Early assessment	Arvioinnin kolmio ECD Testipilotti	ER	IRT, CTT	Anderson LR-test $\chi^2(3) = 7.09$, $p = .07$ B = -1.48 - 1.54		

(Gane, ym. 2021)				CTT: ER		
Instrument	Asianmukaisuus: Testipilotti	ER	ER	ER	ER	ER
(Chen, ym. 2017)	Asiantuntijoiden analyysi					
TechCheck	19 asiantuntijan palaute	ER	IRT, CTT	ICC: 1.25(range = -2.63,.7), 1.03(range = 0.65,1.41)	Kriteerivaliditeetti Pearsons correlation coefficient	Lineaarinen korrelaatio $r = 0.53$ ($p < .001$)
(Relkin, ym. 2020)	Fleiss' Kappa	81% yhdenmukaisuus arvioijien välillä $\kappa = 0.63$ (95 % CI) $p < 0.001$.				
TDIA Three-Dimensional Integrated Assessment	Ymmärrettävyys: Testipilotti	ER	ER	ER	ER	ER
(Zhong, ym. 2016)						

Note. AIC = Akaike Information Criterion; BIC = Bayesian Information Criterion; CTT = Classical Test Theory; ICC = Item Characteristic Curves; IIC = Item Information Curves; KMO = The-Kaiser-Meyer-Olkin; RMSEA = Root Mean Square Error of Approximation; SRMR = Residual-based Standardized root Mean Square residual; IRT = Item Response Theory; MIRT = multidimensional IRT; EFA = Eksploratiivinen faktorianalyysi; ER = ei raportoitu.

Mittareiden sisällön validiteetti oli useimmissa tapauksissa pyritty varmistamaan asiantuntijoiden ja heiltä kerättävien kommenttien avulla, varmistaen näin mittarin asianmukaisuutta ja kokonaisvaltaisuutta. Asiantuntijoiden määrä vaihteli muutamasta (CT practices test, Kong & Wang, 2021) aina 19 (TechCheck, Relkin ym., 2020) saakka. Vain yhden mittarin (TechCheck) kohdalla oli mitattu ja raportoitu asiantuntijoiden arviot, jotka olivat 81 % samanlaiset. Testipilotti oli myös useissa mittareissa suoritettu vaikkei sen tuloksista oltukaan raportoitu kattavasti. Mittareiden luomisessa käytettäviä viitekehyksiä “Arvioinnin kolmio” ja “Evidence Centered Design” (ECD) oli hyödynnetty validiteetin parantamiseksi. Abbreviated Computational Thinking test (Abbreviated CTt) - mittarin kohdalla luotettavuus perustui aiemmin validoidun CTt-testin (González ym., 2015) varaan, sillä mittari oli rakennettu tämän mittarin pohjalta.

Rakennevaliditeetti on taulukossa 5 jaettu kahteen osaan: Mittarin rakenteen validiteettiin sekä hypoteesin testaamiseen. Mittarin rakenteen validiteettia oli raportoitu useimmissa mittareissa pois lukien BCTt (Zapata-Caceres ym., 2021), Instrument (Chen ym., 2017) sekä TDIA (Zhong ym., 2016). Muissa, joissa oli raportoitu mittarin rakenteen validiteettia, suosituin analyysi, oli IRT, jota oli käytetty kuudessa mittarissa (Relkin ym., 2020; Gane ym., 2021; Gong & Wang, 2021; Basu ym., 2021 & Li ym., 2021) IRT:n lisäksi Li ym. (2021), Gane ym. (2021) ja Relkin ym. (2020) olivat käyttäneet CTT:a mittarin rakenteen validiteetin tukemiseksi ja täten mittarin luotettavuuden lisäämiseksi. Erilaisia faktorianalyseja oli käytetty kolmessa tutkimuksessa. Abrivated CTt:ssa (Tsarava ym., 2022) oli käytetty eksploratiivista faktorianalyysia; Basun ym. (2021) mittarissa oli käytetty IRT:n lisäksi myös faktorianalyysia; CT practices test:ssa (Kong & Wang, 2021) oli käytetty taas kahden faktorin analyysia. Hypoteesin testaamisesta oli raportoinut neljä tutkimusta: CTP (Basu ym., 2021) ja Abbreviated CTt (Tsarava ym., 2022) käyttivät korrelaatioanalyysia rinnakkaisvaliditeetin tarkasteluun ja TechCheck (Relkin ym., 2021) ja CTA-CES (Li ym., 2021) kriteerivaliditeetin toteamiseen.

6.3. Mittareiden reliabiliteetti

Taulukko 6 on koonti mittareiden reliabiliteettia mittaavista muuttujista. Ensimmäisenä taulukossa esitetään mittarin nimi, lyhenne sekä mittarin tutkijat sekä julkaisuajankohta. Reliabiliteetti on jaettu mittarin sisäiseen reliabiliteettiin ja yleiseen reliabiliteettiin.

Taulukko 6.

Yhteenveto artikkeleiden raportoiduista reliabiliteettiominaisuuksista.

<i>Mittari</i>	<i>Tutkimus, johon viitataan</i>	<i>Reliabiliteetti</i>			
		<i>Sisäinen johdonmukaisuus</i>		<i>Reliabiliteetti</i>	
		<i>Menetelmä</i>	<i>Tulokset</i>	<i>Menetelmä</i>	<i>Tulokset</i>
Abbreviated CTt Abbreviated Computational Thinking test	(Tsarava, ym., 2022)	Cronbach's alpha	$\alpha = 0.637$	ER	ER
		Omega	0.638, 95 % CI [0.55, 0.71, SE = 0.04]		
BCTt The beginners CT test	(Zapata-Caceres, ym., 2021)	Cronbach's alpha	$\alpha = 0.824$	ER	ER
CTA-CES Computational Thinking Assessment for Chinese Elementary Students	(Li, ym., 2021)	Cronbach's alpha	$\alpha = 0.76$	ER	ER
CTC CT concepts assessment instrument	(Basu, ym. 2021)	Cronbach's alpha	$\alpha = 0.78$	ER	ER
CTP CT practices assessment instrument	(Basu, ym. 2021)	Cronbach's alpha	Taso 1 ER Taso 2 $\alpha = 0.66$ Taso 3 $\alpha = 0.48$	ER	ER
CT practices test	(Kong & Wang, 2021)	Empirical reliability	0.81	ER	ER
Grade 3 Early assessment	(Gane, ym. 2021)	ER	ER	Cohen's Kappa	median = 0.98, min = 0.94, max = 1
Grade 4 Early assessment	(Gane, ym. 2021)	ER	ER	Cohen's Kappa	median = 0.91, min = 0.55, max = 1
Instrument	(Chen, ym. 2017)	Cronbach alpha	$\alpha = 0.808$	ER	ER
TechCheck	(Relkin, ym. 2020)	Cronbach's alpha	$\alpha = 0.68$	ER	ER
TDIA Three-Dimensional Integrated Assessment	(Zhong, ym. 2016)	ER	ER	ER	ER

Note. ER = Ei raportoitu

Useimman mittarin kohdalla oli toteutettu reliabiliteetin testaus Cronbachin alfan avulla. Poikkeuksen muodostivat Grade 3 early assesment & grade 4 early assesment (Gane ym., 2021), jossa oli testattu reliabiliteettia Cohen's Kappan avulla sekä TDIA (Zhong ym., 2016), joiden tutkimuksessa ei ollut raportoitu reliabiliteetin mittaussuomenetelmää. Cronbachin alfa arvot vaihtelivat 0.68–0.824 välillä, viitaten mittareiden vahvaan reliabiliteettiin lukuun ottamatta CTP (Basu ym., 2020), jossa tason 3 Cronbachin alfa oli kohtalainen 0.48.

7 Pohdinta

Ohjelmoinnillisen ajattelun mittaamisen haasteena on ollut kokonaisvaltaisen ja yhtenäisen määritelmän puuttuminen. Kasvavan tutkimustyön avulla näihin haasteisiin on pyritty ja osin myös pystytty vastaamaan. Mittareita on onnistuttu kehittämään ja saamaan aikaan saatu myös varsin vahvaa näyttöä niiden toiminnasta. Tämän tutkimuksen tarkoituksena oli jatkaa ohjelmoinnillisen ajattelun mittareiden tarkastelua. Tämä toteutettiin tekemällä systemaattinen katsaus olemassa olevista mittareista ja niiden psykometrisistä ominaisuuksista.

Analysoimalla näiden mittareiden psykometrisiä ominaisuuksia voidaan paremmin arvioida mittareiden toimivuutta ja havaita mahdollisia puutteita niissä. Yksi haaste mittareiden kehittämisessä on ollut se, että ohjelmoinnillinen ajattelu on monisäkeinen käsite, joka koostuu eri taidoista ja kyvyistä. Näitä taitoja ovat esimerkiksi ongelmanratkaisu, looginen ajattelu ja algoritminen suunnittelu. Mittareiden tulee olla tarpeeksi laajoja kattamaan kaikki nämä taidot, mutta myös riittävän tarkkoja erottaakseen eri taidot toisistaan.

Itse mittarien haasteet, jotka tukevat myös aikaisemmissa tutkimuksissa esille nousseita haasteita, olivat ilmeisiä myös tässä tutkimuksessa. Tutkimuksessa keskityttiin nimenomaan psykometrisiä ominaisuuksia raportoiviin artikkeleihin ja havaitsimme puutteellisen tutkimuksen myös niissä. Kirjallisuuskatsauksen aineistoksi suodattui alun perin 187 artikkelia, mutta vain 9:ssä niistä löytyi mittarin psykometriikkaa eli validiteettia ja reliabiliteettia raportoivia artikkeleita. Tangin ym. (2020) katsauksen 54 artikkelista alle puolet raportoi minkäänlaisia psykometrisiä ominaisuuksia. Tämän katsauksen perusteella tilanne on vielä tätäkin huolestuttavampi. Luotettavuus ja reliabiliteetti ovat tärkeimpiä mittarin tunnuslukuja ja ennen näiden mittaamista ja raportointia mittareita on mahdotonta ottaa laajempaan käyttöön, ja sitä kautta edistää ohjelmoinnillisen ajattelun asemaa opetusmaailmassa.

Katsauksen mittareiden vahva sisällön validiteetti ja korkea reliabiliteetin taso voidaan kuitenkin nähdä tutkimuksen positiivisina poimintoina. Sisällön validiteettia olisi hyvä vielä tehostaa kasvattamalla systemaattista otetta asiantuntijoiden kommentointien pohjalta. On tärkeää paitsi käydä keskustelua, mutta myös kerätä raportoituja tietoja näiden kommenttien ja arviointien pohjalta. Toistaiseksi vain yhden mittarin (Relkin ym., 2020) kohdalla oli

mitattu ja raportoitu asiantuntijoiden arviointien yhdenmukaisuus. Kuitenkin lähes kaikkien mittareiden kohdalla oli sisällön validiteettia kohotettu erilaisten menetelmien mukaan. Poikkeuksen muodosti BCTt (Zapata-Caceres, 2021), jossa ei ollut raportoitu mitään sisällön validiteetista. Sisällön validiteetin raportoiminen on keskeistä, sillä mittareiden sisäinen validiteetti kertoo, miten hyvin mittari pystyy tavoittamaan sen ilmiön, joka sen on tarkoitus tavoittaa. Lisäksi sisällön validiteetti kertoo, onko mittari sisällöllisesti sitä, mitä sen pitäisi olla. (Metsämuuronen, 2009, 125.) Keinoja tähän oli muun muassa asiantuntijakeskustelut ja –paneelit, validointi, eli mittarin vertaaminen sellaisiin konstrukteihin, jotka ovat todettu luotettaviksi, sekä varmistaminen, että mitattava kohde ymmärtää mittarin niin kuin se on tarkoitus ymmärtää. (Metsämuuronen, 2009, 125–131). Keskustelujen ja kommenttien tavoitteena on kehittää luotettavin mahdollinen mittari, joten selkeät tunnusluvut näiden arviointien yhdenmukaisuudesta olisivat varmasti hyödyksi tämän asian toteutumisessa.

Mittareiden rakenteen validiteettia oli myös kohotettu erilaisten menetelmien, kuten esimerkiksi CTT:n, IRT:n sekä EFA:n avulla. Cosmin taksonomian mukaan nämä ovat juuri niitä keinoja, jolla mittarista saadaan luotettavampi (Prinsen ym., 2018). Rakennevaliditeetti kertoo, kuinka hyvin mittari kykenee mittaamaan sitä, mitä sen on tarkoitus mitata. Rakennevaliditeetin menetelmät vertaavat mittareiden osioita keskenään esimerkiksi faktorianalyysin avulla. Tutkimukseen otetuista mittareista yhdeksän oli vahvistanut validiteettiaan mittarin rakenteen validiteetin tutkimisen avulla. Tämä osoittaa suurimman osan mittareista olevan rakennevalidoituja eli ne kykenevät mittaamaan kohdetta juuri niin, kuin niiden on tarkoitus mitata. Hypoteesin testaus oli kuitenkin esitelty vain neljässä tutkimuksessa (Basu ym., 2021; Li ym., 2021; Relkin ym., 2020; Tsarava ym., 2022). Hypoteesin testaamisessa uutta mittaria verrataan jo olemassa olevaan validoituun mittariin. Näitä tapoja ovat erilaiset korrelaatioanalyysit, jotka kohottivat joko rinnakkaisvaliditeettia CTP (Basu ym., 2021) ja Abbreviated CTt (Tsarava ym., 2022) tai kriteerivaliditeettia TechCheck (Relkin ym., 2021) ja CTA-CES (Li ym., 2021).

Tuloksien mukaan sekä sisällön validiteettia että rakennevaliditeettia on tutkittu vain neljässä tutkimuksessa (Basu ym., 2021; Li ym., 2021; Relkin ym., 2020; Tsarava ym., 2022). Näin ollen voidaankin todeta, että nämä neljä mittaria ovat validiteetiltaan luotettavampia kuin muut tässä katsauksessa esitellyt mittarit. Tämä ei kuitenkaan tarkoita, etteivät muut mittarit ole käyttökelpoisia nyt tai jatkokehityksen jälkeen. Kuitenkin on hyvä pohtia, että miten muita mittareita voitaisiin saada luotettavamaksi validiteetin kannalta. Onko

mahdollista, että niitä tutkittaisiin tarkemmin vielä ja kehitettäisiin siihen suuntaan, että niillä saisi kerättyä vielä luotettavampaa tietoa ohjelmoinnilliseen ajatteluun liittyen. Toki on myös mahdollista, että nämä kyseiset tutkimukset, joissa kaikkia katsauksessa esiteltyjä validiteetin tasoja ei ole raportoitu, mutta ne ovat kuitenkin testattu. Tällöin luotettavuus olisi korkeampaa validiteetin kannalta kuin tutkimukset ovat osoittaneet raportissaan.

Mittareiden reliabiliteetin raportointi oli tutkimuskatsauksen perusteella puutteellista eikä täyttänyt tieteelliselle tutkimukselle asetettavia luotettavuusvaatimuksia. Tämä johtui siitä, että yleinen reliabiliteetti jäi lähes jokaisesta mittarin tutkimuksesta raportoimatta. Tähän voi olla syynä mittareiden reliabiliteetin keskittyminen sisäiseen johdonmukaisuuteen eli sisäiseen reliabiliteettiin. Tutkimuksissa ei ole raportoitu yleistä reliabiliteettia, pois lukien Ganen ym. (2020) Grade 3 assessment ja Grade 4 assessment –mittareita. Mittareiden sisäinen reliabiliteetti on kuitenkin raportoitu lähes kaikissa tutkimuksissa pääsasiassa käyttämällä Cronbahin alfakerrointa. Poikkeuksen muodostivat Grade 3 assesment ja Grade 4 assesment (Gane ym., 2020) sekä TDIA (Zhong ym., 2016), joissa ei ollut raportoitu sisäistä reliabiliteettia ollenkaan. Tästä voidaan päätellä, että tämä menetelmä on todettu hyväksi kohottamaan reliabiliteettia ja näin ollen koko mittarin luotettavuutta.

Tämän katsauksen perusteella voidaan todeta, että Abbreviated CTt (Tsarava ym., 2022), CTA-CES (Li ym., 2021), CTP (Basu ym., 2021) sekä TechCheck (Relkin ym., 2020) olisivat psykometrisiltä ominaisuuksiltaan luotettavimpia, sillä niissä kaikissa on mitattu sisäistä validiteettia, rakennevaliditeettia sekä sisäistä johdonmukaisuutta. Näistä kaikista puuttui yleisen reliabiliteetin mittaaminen, mikä vahvistaisi vielä entisestään kyseisiä mittareita ja olisi hyvä seuraava askel mittareiden validoimiselle.

Jatkotutkimusmahdollisuuksina voisi edelleen validoida näitä olemassa olevia mittareita. Testata niitä eri kulttuureissa ja näin vahvistaa rinnakkaisvaliditeettia. Näiden uusien validointien pohjalta olisi mahdollista joko vahvistaa mittarin nykyistä asemaa tai kehittää sitä täydentämällä tai poistamalla siitä epäolennaisiksi todettuja kohtia. Tutkimuksessamme esille nousseet haasteet etenkin rakennevaliditeetin kohdalla olisi hyvä ottaa muutenkin jatkotutkimuksissa huomioon ja sitä kautta edetä olemassa olevien mittareiden kanssa. Jatkotutkimusmahdollisuuksiin lukeutuu myös olemassa olevien mittareiden tarkempi sisällön analyysi esimerkiksi ohjelmoinnillisen ajattelun sisältöalueiden kautta, kuinka hyvin mittarin sisällöt vastaavat ohjelmoinnillisen ajattelun taitojen osa-alueita.

Reliabiliteetin lisääminen on toinen seikka, joka tutkimuksemme kannalta tarvitsisi huomiota. Suurimmassa osassa mittareita Cronbachin alfa oli ainoa kerroin kuvaamaan mittarin reliabiliteettia, vaikka se onkin tärkeä mittari se ei kuitenkaan ole ainoa, jota tulisi käyttää. Sen lisäksi tulisi käyttää myös muita menetelmiä kuten toistomittaus, jotta voidaan arvioida mittarin vakautta ajan myötä tai Abbreviated CTt:ssä (Tsarava ym., 2022) käytettyä McDonaldin omegaa, jotta voidaan tarkastella useampia erilaisia mittarin ominaisuuksia.

Kaiken kaikkiaan olisi hyvä, jos jatkossa keskityttäisiin enemmän olemassa olevien mittareiden tutkimiseen, mittaamiseen ja validoimiseen. Ei ole tarve “keksiä pyörää uudestaan” kun tarjolla on lähes valmiita mittareita, jotka tarvitsevat lisää tutkimusta kehittyäkseen.

7.1 Tutkimuksen luotettavuus ja tutkimusetiikka

Tutkimuksessa käytetty systemaattinen kirjallisuuskatsaus jo itsessään vahvistaa tutkimuksen luotettavuutta. Kahden tutkijan yksilöllinen aineiston arviointi ja näiden keskinäisten arviointien vertailu yhdessä kolmannen tutkijan kanssa paransi tutkimuksen luotettavuutta. Tässä kirjallisuuskatsauksessa hyödynnettiin useita tietokantoja ja näin oli mahdollista saavuttaa laajempi aineisto ja näin lisätä luotettavuutta.

Tutkimuksen luotettavuutta pohtiessa on kuitenkin hyvä tiedostaa myös rajoitteet ja luotettavuutta heikentävät seikat. Tutkimuksen aineistonkeruu toteutettiin englannin kielellä tutkijoiden omista kielirajoitteista johtuen, joten monet kansainvälisestikin arvostetut artikkelit, joita ei ole käännetty englanniksi jäivät kokonaan pois katsauksesta. Katsaukseen otettiin mukaan myös ainoastaan vertaisarvioidut ja julkaistut artikkelit, joten opinnäytetyöt ja muut vertaisarvioimatta olevat artikkelit jäivät pois tutkimuksesta jo heti alkuvaiheessa. Tutkimuksen prosessin aikana tiivistelmien pohjalta jätettiin pois 125 artikkelia, joten näiden artikkelien tutkijoiden kirjoittamat tiivistelmät ja niissä korostetut asiat vaikuttivat suurelta osin tämän tutkimuksen lopputuloksiin.

Lopuksi täytyy myös tunnustaa tutkijoiden mahdollinen oman kokemattomuuden vaikutus. Tämä oli ensimmäinen systemaattinen kirjallisuuskatsaus, jonka tutkijat toteuttivat, joten menetelmä oli heille kokonaan uusi. Ohjelmoinnillinen ajattelu oli myös molemmille tutkijoille jossain määrin vieras, joten siitä kirjoitettujen artikkeleiden sisäistäminen paitsi vei

aikaa, saattoi myös tuottaa virheitä artikkeleiden sisäänottoja ja poissulkuja mietittäessä, vaikka yksilöllisillä arvioinneilla pyrittiinkin tähän virhemarginaaliin vaikuttamaan.

Lähdeluettelo

- Ahtola, A. (2016). Koulu hyvinvoinnin rakentajana. Teoksessa Ahtola, A.(toim.) Psykkinen hyvinvointi ja oppiminen. Jyväskylä: PS-kustannus, 12-21.
- A.V. Aho Computation and computational thinking *Computer Journal*, 55 (2012), pp. 832-835, 10.1093/comjnl/bxs074
- Balanskat, A., & Engelhardt, K. (2015). Computing our future *Computer programming and coding Priorities, school curricula and initiatives across Europe*. European Schoolnet.
- Basu, S., Rutstein, D. W., Xu, Y., Wang, H., & Shear, L. (2021). A principled approach to designing computational thinking concepts and practices assessments for upper elementary grades. *Computer Science Education*, 31(2), 169-198.
- Fink, A. (2019). *Conducting research literature reviews: From the internet to paper*. Sage publications.
- Bearfield, D. A., & Eller, W. S. (2008). Writing a literature review: the art of scientific literature. *Public administration and public policy*. New York, 134, 61.
- Chen, G., Shen, J., Barth-Cohen, L., Jiang, S., Huang, X., & Eltoukhy, M. (2017). Assessing elementary students' computational thinking in everyday reasoning and robotics programming. *Computers & education*, 109, 162-175.
- Computer Science Teachers Association (CSTA). (2017). CSTA K-12 Computer Science Standards. Retrieved from <https://www.csteachers.org/page/standards>
- Curzon, P., Dorling, M., Ng, T., Selby, C., & Woollard, J. (2014). *Developing computational thinking in the classroom: a framework*.
- DeVellis, R. F. (2017). *Scale development: Theory and applications*. Sage Publications.
- Dixon-Woods, M., Booth, A., & Sutton, A. J. (2007). Synthesizing qualitative research: a review of published reports. *Qualitative Research*, 7(3), 375–422.

- Ehsan, H., Rehmat, A. P., & Cardella, M. E. (2021). Computational thinking embedded in engineering design: capturing computational thinking of children in an informal engineering design activity. *International Journal of Technology and Design Education*, 31(3), 441-464.
- European Schoolnet. (2017). Developing Computational Thinking in Compulsory Education. Retrieved from https://www.eun.org/documents/10180/186546/ComputationalThinking_Report.pdf/579dfa24-69a9-4929-bf2d-7f167058dd98
- Gane, B. D., Israel, M., Elagha, N., Yan, W., Luo, F., & Pellegrino, J. W. (2021). Design and validation of learning trajectory-based assessments for computational thinking in upper elementary grades. *Computer Science Education*, 31(2), 141-168.
- González, M. R. (2015). Computational thinking test: Design guidelines and content validation. In *EDULEARN15 Proceedings* (pp. 2436-2444). IATED.
- Guggemos, J., Seufert, T., & Román-González, M. (2019). The relationship between students' computational thinking skills, their motivational variables, and their use of digital media. *International Journal of Educational Technology in Higher Education*, 16(1), 1-15.
- Hsu, T. C., Chang, S. C., & Hung, Y. T. (2018). How to learn and how to teach computational thinking: Suggestions based on a review of the literature. *Computers & Education*, 126, 296-310.
- Ioannidou, A., Bennett, V., Repenning, A., Koh, K. H., & Basawapatna, A. (2011). Computational Thinking Patterns. Online Submission.
- Kallio, T. J. (2006). Laadullinen review -tutkimus metodina ja yhteiskuntatieteellisenä lähestymistapana. Tampere: Hallinnon tutkimuksen seura.
- Kalelioglu, F., Gulbahar, Y., & Kukul, V. (2016). A framework for computational thinking based on a systematic research review.
- Kong, S. C., & Wang, Y. Q. (2021). Item response analysis of computational thinking practices: Test characteristics and students' learning abilities in visual programming contexts. *Computers in Human Behavior*, 122, 106836.

- Li, Y., Xu, S., & Liu, J. (2021). Development and validation of computational thinking assessment of Chinese elementary school students. *Journal of Pacific Rim Psychology*, 15, 18344909211010240.
- Luukkainen-Markkula, R., & Halttunen, L. (2019). Assessment and evaluation in Finnish education. In *Finnish Education* (pp. 65-81). SensePublishers, Rotterdam.
- Mannila, L., & Paksula, M. (2021). Measuring programming skills: a critical review of programming skill assessment methods. *Education and Information Technologies*, 26(4), 3999-4026.
<https://doi.org/10.1007/s10639-021-10446-w>
- Manyika, J. (2017). *Technology, jobs and the future of work*.
- Metsämuuronen, J. (2009). *Tutkimuksen tekeminen ihmistieteissä*. Vaajakoski: Gummerus Kirjapaino Oy. 4. painos.
- National Research Council. (2010). *Committee for the Workshops on Computational Thinking: Report of a workshop on the scope and nature of computational thinking*. Washington, DC: National Academies Press. http://www.nap.edu/catalog.php?record_id=12840.
- National Science Foundation (NSF). (n.d.). *Computer Science for All*. Retrieved from https://www.nsf.gov/funding/pgm_summ.jsp?pims_id=505020
- Nunnally, J. C. (1978). *Psychometric theory*. McGraw-Hill.
- OECD. (2019). *The Future of Education and Skills: Education 2030*. Retrieved from https://www.oecd-ilibrary.org/education/the-future-of-education-and-skills_9789264279395-en
- Papert, S. (1980). *Personal computing and its impact on education. The computer in the school: Tutor, tool, tutee*, 197–202.
- Papert, S. (1991). *Situating constructionism*. In I. Harel & S. Papert (Eds.), *Constructionism* (pp. 1–11). Norwood: Ablex
- Passey, D. *Computer science (CS) in the compulsory education curriculum: Implications for future research* *Education and Information Technologies*, 22 (2) (2017), pp. 421-443, 10.1007/s10639-016-9475-z

- Prinsen, C. A., Mokkink, L. B., Bouter, L. M., Alonso, J., Patrick, D. L., De Vet, H. C., & Terwee, C. B. (2018). COSMIN guideline for systematic reviews of patient-reported outcome measures. *Quality of life research*, 27, 1147-1157.
- Polat, Ö., Hopcan, S., Kucuk, M., & Sisman, B. (2021). A comparative analysis of the computational thinking skills of middle school students in Turkey and Spain. *Education and Information Technologies*, 26(2), 1183-1200.
- Relkin, E., de Ruiter, L., & Bers, M. U. (2020). TechCheck: Development and validation of an unplugged assessment of computational thinking in early childhood education. *Journal of Science Education and Technology*, 29(4), 482-498.
- Román-González, M., Pérez-González, J. C., & Jiménez-Fernández, C. (2017). Validation of the computational thinking test (CTt) in a Spanish-speaking sample. *Computers & Education*, 114, 106-118.
- Román-González, M., Pérez-González, J. C., Moreno-León, J., & Robles, G. (2018). Measuring computational thinking skills: A systematic review. *Computers & Education*, 126, 201-211.
- Saari, H., & Silander, P. (2015). Assessment practices in Finnish schools. In *Second International Handbook of Educational Change* (pp. 411-426). Springer, Dordrecht.
- Shute, V. J., Sun, C., & Asbell-Clarke, J. (2017). Demystifying computational thinking. *Educational Research Review*, 22, 142-158.
- Sun, L., Guo, Z., & Hu, L. (2021). Educational games promote the development of students' computational thinking: a meta-analytic review. *Interactive Learning Environments*, 1-15.
- Tang, X., Yin, Y., Lin, Q., Hadad, R., & Zhai, X. (2020). Assessing computational thinking: A systematic review of empirical studies. *Computers & Education*, 148, 103798.
- Tang, K. Y., Chou, T. L., & Tsai, C. C. (2020). A content analysis of computational thinking research: An international publication trends and research typology. *The Asia-Pacific Education Researcher*, 29(1), 9-19.
- Tanhua-Piironen, E., Viteli, J., Syvänen, A., Vuorio, J., Hintikka, K. A., & Sairanen, H. (2016). Perusopetuksen oppimisympäristöjen digitalisaation nykytilanne ja opettajien valmiudet hyödyntää digitaalisia oppimisympäristöjä.

- Terwee, C. B., Prinsen, C. A., Chiarotto, A., Westerman, M. J., Patrick, D. L., Alonso, J., ... & Mokkink, L. B. (2018). COSMIN methodology for evaluating the content validity of patient-reported outcome measures: a Delphi study. *Quality of Life Research*, 27, 1159-1170.
- Tikva, C., & Tambouris, E. (2021). Mapping computational thinking through programming in K-12 education: A conceptual model based on a systematic literature Review. *Computers & Education*, 162, 104083
- Tsarava, K., Moeller, K., Román-González, M., Golle, J., Leifheit, L., Butz, M. V., & Ninaus, M. (2022). A cognitive definition of computational thinking in primary education. *Computers & Education*, 179, 104425.
- UNESCO. (2011). UNESCO ICT Competency Framework for Teachers. Retrieved from <https://unesdoc.unesco.org/ark:/48223/pf0000214117>
- Voogt, J., Fisser, P., Good, J. et al. Computational thinking in compulsory education: Towards an agenda for research and practice. *Educ Inf Technol* 20, 715–728 (2015). <https://doi.org/10.1007/s10639-015-9412-6>
- Winchester CL, Salji M. (2016) Writing a literature review. *Journal of Clinical Urology*. 9(5):308-312. doi:[10.1177/2051415816650133](https://doi.org/10.1177/2051415816650133)
- Wing, J. (2017). Computational thinking's influence on research and education for all. *Italian Journal of Educational Technology*, 25(2), 7-14
- Wing. (n.d.). Towards• Computational• Thinking. <https://doi.org/10.1093/itnow/bwx113>
- Wing, J. M. (2006). Computational thinking. *Communications of the ACM*, 49(3), 33-35
- Zapata-Cáceres, M., Martín-Barroso, E., & Román-González, M. (2021). Collaborative game-based environment and assessment tool for learning computational thinking in primary school: a case study. *IEEE Transactions on Learning Technologies*, 14(5), 576-589.
- Zhang, L., & Nouri, J. (2019). A systematic review of learning computational thinking through Scratch in K-9. *Computers & Education*, 141, 103607.
- Zhong, B., Wang, Q., Chen, J., & Li, Y. (2016). An exploration of three-dimensional integrated assessment for computational thinking. *Journal of Educational Computing Research*, 53(4), 562–590.