

Õppimist toetava e-tasemetöö väljatöötamine III kooliastmele õpilaste loodusteadusliku pädevuse hindamiseks

Katrin Vaino^{a1}, Triin Rosin^a, Ülle Liiber^{a,b}, Regina Soobard^a,
Moonika Teppo^a, Ana Valdmann^a, Elle Reisenbuk^c, Miia Rannikmäe^a

^a Tartu Ülikooli ökoloogia ja maateaduste instituudi
loodusteadusliku hariduse keskus

^b Tartu Ülikooli ökoloogia ja maateaduste instituudi geograafia osakond

^c Haridus- ja Noorteamet

Annotatsioon

Selle disainipõhise uuringu eesmärk oli välja töötada e-tasemetöö, millega saaks hinnata riiklikus õppekavas määratletud III kooliastme lõpuks omandatud loodusteadusliku pädevuse taset. Ühtlasi pidi tasemetöö andma kirjeldavat tagasisidet nii õpilasele, õpetajale, lapsevanemale kui ka haridusüldsusele. Protsessi kompleksisusest tingituna keskendutakse selles artiklis kitsamalt tasemetöö disainimise protsessile ja saadud lõpptulemusele ning põhjendatakse tehtud disainiotsuseid. Uurijate ja praktikute vahelises koostöös (2018–2022) jõuti uudse lahenduse – kontekstipõhise tasemetööni, mida saab kasutada üle-eestiliseks põhiuuringuks. Õpilaste saavutatud loodusteadusliku pädevuse taset kirjeldatakse neljal tasemel üheksa tunnuse abil, mis on ülevaatlikkuse nimel rühmitatud omakorda neljaks: loodusteaduslikud teadmised, uurimuslikud oskused, probleemi lahendamise ja otsuse tegemise oskused ning kommunikatsioonioskused. Tasemetöö koosneb neljast alatestist ning kokku 35 ülesandest.

Võtmesõnad: loodusteaduslik pädevus, kontekstipõhine e-tasemetöö, diagnostiline test, disainipõhine uuring

¹ Ökoloogia ja maateaduste instituut, Tartu Ülikool, Vanemuise 46, 51003, Tartu; katrin.vaino@ut.ee.

Sissejuhatus

Loodusainete õppimise tulemusena peaks õpilastel põhikooli lõpuks kujunema loodusteaduslik pädevus, mille all mõistetakse loodusteaduslikke teadmisi, uurimuslikke ja probleemilahendamise oskusi, kommunikatsioonioskusi ning jätkusuutlikkust väärtustavaid hoiakuid (PRÕK, 2023, lisa 4). Taoliselt määratletud loodusteadusliku pädevuse kujunemise eelduseks on kognitiivselt aktiveeriva õpetamispraktika kasutamine, mida toetab ka Eestis juba aastaid kehtinud pädevuspõhine õppekava. Samas näitavad uuringud (Valdmann *et al.*, 2012; Henno *et al.*, 2017; OECD, 2019), et loodusteaduste tundides domineerib siiani struktureeritud (õpetaja juhitud) õpetamine, mis keskendub madalamat järku mõtlemisoskuste kujundamisele. Seega võib väita, et Eesti loodusteaduste tundides kaldutakse õpetamisel, sh hindamisel keskendumata vaid osale taotletavast loodusteaduslikust pädevusest.

Vastavalt põhikooli riiklikule õppekavale (2023) peaks õpilane olema „õppeprotsessis aktiivne osaleja, kes võtab võimetekohaselt osa oma õppimise eesmärgistamisest, õpib iseseisvalt ja koos kaaslastega, õpib oma kaaslasi ja ennast hindama ning oma õppimist analüüsima ja juhtima“ (§ 5). Selleks, et toetada õpilase loodusteadusliku pädevuse ning ennastjuhtiva õppija kujunemist, peaksid mitmekesiste õppemeetodite rakendamise kõrval olema olulisel kohal ka õppimist toetavad hindamismeetodid, mis aitavad koguda andmeid õpilase arengu kohta ning annavad tagasisidet, mis toob välja nii õpilase olemasolevad teadmised, oskused, hoiakud (sh tugevused) kui ka arendamist vajavad aspektid (PRÕK, 2023, § 19). Samas näitab Eesti koolides läbiviidud uuring (Aksen *et al.*, 2018), et õpilastele antava sõnalise kirjeldava tagasiside määr väheneb igas järgnevas kooliastmes.

Eestis on alates 2006. aastast 15–16-aastaste õpilaste seas korduvalt läbi viidud PISA uuringut, sh hinnatud nende loodusteaduslikku pädevust. Siiski ei anna selle tulemused tagasisidet õpilase või õpilaste rühma pädevuse taseme ega Eesti põhikooli riiklikus õppekavas (2023, lisa 4) määratletud loodusteadusliku pädevuse taseme kohta. See tingis vajaduse Eesti õppekavast lähtuva riikliku tasemetöö järele, mis annaks võimalikult objektiivset ning võrreldavat tagasisidet, et hinnata õppekavas seatud eesmärkide saavutamist ning teha hariduspoliitilisi otsuseid (Haridus- ja Teadusministeerium, 2014). Teiselt poolt peaks see andma spetsiifilist tagasisidet õpilasele ja õpetajale, et selle pinnalt planeerida oma edasist õppimist ning õpetamist (*ibid.*).

Riiklikud tasemetööd on õppeaine või ainevaldkonna ülesannetest koosnevad hindamisvahendid, mida Haridus- ja Noorteamet igal aastal riikliku õpitulemuste välishindamise osana korraldab, ning selle vormiks on test (Harno, 2023a). Testi defineeritakse kui vahendit või protseduuri, mille abil

kogutakse standardiseeritud kujul andmeid isikute võimekusest lahendada teatud ülesandeid.

Loodusteadusliku pädevuse määratlus

Pädevust võib defineerida kui oskuste, teadmiste, käitumisviiside, väärtuste ja hoiakute kogumit, mida inimene vajab koolis ja töömaailmas ülesannete täitmiseks ning igapäevaelus hakkama saamiseks (OECD, 2018; Pereira, 2022). Loodusteadusliku pädevuse kaasaegne käsitlus hõlmab õpilaste loodusteaduslike teadmiste, nende rakendamise ning uurimuslike oskuste kõrval loovust, valmidust tegeleda mitme lahendiga probleemidega ning oskust teha põhjendatud otsuseid (OECD, 2020; PRÕK, 2023, lisa 4). Mitme lahendiga probleemideks on näiteks strateegia valiku, disaini-, dilemma- ning muud avatud ehk nn halvasti defineeritud probleemid (Jonassen, 2011). Choi jt (2011) toovad loodusteadusliku pädevuse määratlemisel välja ka väärtused ja isikuomadused, mis toetavad vastutustundlikku käitumist globaalse kodanikuna. Lisaks koostöö- ja suhtlemisoskustele, mida läheb vaja loodusteadusliku kandepinnaga probleemide lahendamisel, tõstavad nad esile ka oskuse analüüsida ja reflekteerida oma käitumist ning oskuse ennast õppijana ise juhtida (*ibid.*). Thier ja Daviss (2002) nimetavad loodusteaduslikku pädevust määratlevate kommunikatsioonioskuste hulgas ka oskust sõnastada oma seisukohti teistele. Kommunikatsioonioskuste puhul on järjest enam hakatud rõhutama oskust hinnata info usaldusväärsust, et aidata õpilastel paremini orienteeruda (sotsiaal)meedias levivas ning sageli vastuolulises infos (Osborne *et al.*, 2022). Nimetatud oskuste kujundamisest üksnes sotsiaalainete valdkonnas ei piisa, sest sageli levib vastuoluline info ka seoses loodusteaduslike teemadega (Höttecke & Allchin, 2020).

Diagnostilised testid: suundumused ennastjuhtiva õppija toetamisel

Nii Eesti kui ka paljude teiste riikide õppekavad on muutunud pädevuspõhiseks, mis omakorda eeldab teistsugust hindamiskäsitlust, kus kõrge panusega hindamise (näiteks eksamid) kõrval on tähtis roll hindamisel, mis keskendub õpetamise ja õppimise toetamisele ning ennastjuhtiva õppija kujunemisele (Assessment Reform Group, 1999; Haridus- ja Teadusministeerium, 2014). Ennastjuhtiv õppija (ingl *self-directed learner*) võtab ise vastutuse püstitada õpieesmärged, valida vahendeid nende saavutamiseks ning monitoorida nende täitmist (Bolhuis, 2003). Ennastjuhtivaks õppijaks saamist toetab ennastreguleeriv õppimine (ingl *self-regulated learning*), mille kohaselt on õpilased

oma õppeprotsessis aktiivsed osalejad nii metakognitiivsel, motivatsioonilisel kui ka käitumuslikul tasandil (Zimmerman, 1989).

Eelkirjeldatud hariduslik suundumus on tinginud vajaduse rakendada kõrge panusega hindamise kõrval järjest rohkem diagnostilisi madala panusega teste, mis võivad lisaks riiklikule õpitulemuste saavutamise monitoorimisele pakkuda ka konstruktiivset tagasisidet õpilase olemasolevate teadmiste, oskuste, hoiakute, aga ka arengukohtade kohta (Csapó & Molnár, 2019). Konstruktiivne tagasiside võiks aidata õpetajal tuvastada nii üksikõpilaste kui õpilaste grupi õpivajadusi ning kujundada õpetamisstrateegiaid, mis lähtuvad õpilaste tegeliku arengu tasemest (Schut *et al.*, 2020). Ka aitaks see paremini adresseerida Loogma jt (2020) esile tõstetud individuaalsete õpiteede kujundamise vajadust, pidades silmas meie järjest enam mitmekesistuvat õpilaskonda. Eelnevalt lähtudes toetuvad taolised uue põlvkonna diagnostilised testid sageli Vögotski loodud sotsiaalkultuurilisele teooriale (Csapó & Molnár, 2019), mille kohaselt on õppimine võimalik, kui see toimub õpilase lähima arengu tsoonis (Vygotsky & Cole, 1978; Daniels, 2007). Testist saadav tagasiside peaks aitama ühelt poolt luua õpetajal sobivad „tellingud“ (Wood *et al.*, 1976 tähenduses) õpilase uute teadmiste, oskuste ning hoiakute kujunemiseks tema lähima arengu tsooni piires. Teiselt poolt peaks tagasiside toetama õpilase eneseregulatsiooni, mille kaudu ta reflekteerib infot, et püstitada teadlikult endale järgmisi õpieesmärke (Hattie & Timperley, 2007). Nimetatud oskuse eelduseks on õpetaja teadlik juhendamine, arendamaks õpilase eneseregulatsiooni oskust, et ta suudaks iseisvalt mõelda sellest, miks ta midagi teeb (Saks & Leijen, 2015). Sellisel kujul muutub õppija testimise protsessis ning oma õpitee kujundamisel aktiivseks tegutsejaks (Üldhariduskoolide välishindamise ..., 2014; Toomaneejinda, 2017) ehk ennastjuhtivaks õppijaks.

Elektroniliste testide eelised ja piirangud

Nii riiklikud eksamid, tasemetööd kui ka muud testi vormis toimuvad hindamised on viimasel aastakümnel muutunud järjest rohkem elektrooniliseks. Elektroonilistes testides saab kasutada erinevaid meediarikkaid stiimuleid: graafilisi, heli-, video- või animeeritud materjale, simulatsioone jne (nt Boyle & Hutchison, 2009). Välishindamise elektroonilise testi puhul eeldatakse, et selle antud hinnangud õpilaste pädevuste kohta on õpetaja antud hinnangutest valiidsemad ja usaldusväärsemad (Vitello & Williamson, 2017). Elektroonilise testi üheks peamiseks piiranguks peetakse aga asjaolu, et pikemaid vabavastuseid tuleb endiselt kodeerida käsitsi (Zhang *et al.*, 2022). Samas toimuvad uuringud, arendamaks välja taoliste vastuste automaatse hindamise süsteemi (Gweon & Schonlau, 2022; He & Schonlau, 2020).

Kontekstipõhine lähenemine

Õpilase huvi puudumine tasemetöö lahendamisel kujutab endast tõsist ohtu selle tulemustest tehtavate järelduste kehtivusele (Finn, 2015). Konteksti rolliks õpilaste testimisel on tekitada õpilases situatiivset huvi, et ergutada teda vastama ülesande enda, mitte võimaliku tulemuse pärast (Fensham & Rennie, 2013).

Kontekstipõhise lähenemise korral tuleb õpilastel rakendada oma teadmisi igapäevaelulises olukorras ettetulevate praktiliste ülesannete lahendamiseks, aga kontekstid võivad hõlmata ka laiemaid ühiskondlikult olulisi probleeme ning nendega seotud väärtusi (Sjöströmi & Talanqueri, 2014). Kontekstipõhise hindamisvahendi korral on õpilasel võimalus näidata loodusteaduslikku pädevust viisil, mis võimaldab tal saada oma sooritusele sisukat tagasisidet. See aitab õpilasel mõista oma pädevuse ulatust, kujundab oskust mõelda oma pädevustele iseseisvalt (ennastjuhtvalt) ja vajadusel neid arendada.

Loodusvaldkonna tasemetöö lähtekontseptsioon

Tasemetöö lähteülesande püstitas Haridus- ja Teadusministeerium. Loodusvaldkonna e-tasemetöö kontseptsioon tugineb Pedaste jt (2017) raportile „Loodusvaldkonna õpitulemuste hindamine“, sh loodusteadusliku pädevuse kaasaegsele määratlusele, mida tutvustati eespool. Raport esitab koos tasemetöö kontseptsiooniga võimaliku hindamismudeli ning tasemetööde koostamise ja läbiviimise esialgse protseduuri. Tasemetöö peaks hindama õpilaste loodusteadusliku pädevuse taset kümne seda iseloomustava tunnuse kaudu: (1) loodusnähtuste ja protsesside ning nende vaheliste põhjuse-tagajärjeseoste selgitamine; (2) loodusteaduslike mõistete, sümbolite ja ühikute kasutamine; (3) arusaamine loodusteaduslikust tekstist ning selle koostamine; (4) arusaamine loodusteaduslikest mudelitest; (5) probleemide lahendamine ja otsuste tegemine, toetudes loodusteaduslikele teadmistele, oskustele ning väärtushinnangutele; (6) info analüüsimine ning järelduse tegemine; (7) probleemi ning (8) uurimisküsimuse või hüpoteesi püstitamine; (9) katse kavandamine ja hindamine; (10) info otsimine ning selle usaldusväarsuse hindamine. Nimetatud kümme tunnust on tuletatud põhikooli riikliku õppekava loodusvaldkonna õppe-eesmärkidest (PRÕK, 2011, lisa 4, § 1.1). Tunnuseid kirjeldatakse neljal tasemel: baastase, kesktase, kõrgtase ja tiiptase. III kooliastme tasemetöös hinnatakse õpitulemusi vastavalt Andersoni ja Krathwohli taksonoomiale (2001) alates „teadmiste rakendamise“ tasemest, st madalamat järku õpitulemusi (teadmine, arusaamine) ei hinnata.

Vastavalt lähtekontseptsioonile (Pedaste *et al.*, 2017) saab õpilane sõnalist tagasisidet oma loodusteadusliku pädevuse tugevatest ja nõrkadest külgedest, mis võimaldab tal ennast õppimise käigus paremini juhtida. Lapsevanem saab

kasutada seda infot oma lapse arengu toetamisel. Õpetaja saab selle põhjal oma tööd paremini planeerida ning õpetamist individualiseerida. Tasemetöö väljaarendamise käigus uurisid Rosin jt (esitatud) ka loodusainete õpetajate sellekohast valmisolekut, millest ilmnes, et kuigi osa õpetajatest on hakanud tasemetööst saadavat tagasisidet õppe planeerimisel rakendama, vajab enamik selle tõlgendamisel rohkem või vähem tuge.

Samale kontseptsioonile toetudes töötakse välja I, II ja IV kooliastme tasemetöid, neist esimese kohta saab lugeda Pedaste jt (Pedaste *et al.*, 2021) uurimusest.

Disainipõhise uuringu rakendamisest hariduses

Disainipõhine uuring (ingl *design-based research*, aga ka *design research*, *research by design*) püüab ületada lõhet haridusuuringute ning reaalse koolipraktika vahel, suurendada väljatöötatud disaini töökindlust (van der Akker *et al.*, 2006) ning täiustada olemasolevat hariduspraktikat süsteemse, paindliku ning korduva analüüsi, edasiarendamise, rakendamise ning korrigeerimise kaudu (Cobb *et al.*, 2003; Barab & Squire, 2004). Vajadust taolise süsteemse lähenemise järele on tinginud ka asjaolu, et sageli jäävad haridusinnovatsiooni elluviijate kaalutlused disainiotsuste tegemisel vaid nende enda teada. Vaja oleks aga sellest protsessist välja sõeluda olulisemad õpimomendid ning disainipõhimõtted, mis aitaksid muuta järgmised sarnased disainiprotsessid elluviijatele energia- ja ajatõhusamaks ning parandada lõpplahenduse vastavust püstitatud ootustele (van der Akker *et al.*, 2006).

Disainiprotsessi käigus kogutavad andmed võivad olla nii kvalitatiivsed kui ka kvantitatiivsed (The Design-Based Research Collective, 2003): kogutakse erinevate kasutajate kogemusi ja mõtteid ning jälgitakse nende käitumist ja interaktsiooni loodavate disainidega (Anderson & Schatuck, 2012), mis dokumenteeritakse (Cobb *et al.*, 2003). Uurimise tulemuseks on nii lõplik disain (uudne lahendus) kui ka protsessi käigus tuletatud disainimise põhimõtted ehk viisid, kuidas disainida (The Design-Based Research Collective, 2003; Cobb *et al.*, 2003). Nagu igale disainimisele, on ka disainipõhisele uuringule omane tsüklikus, kuigi Zhen (2015) toob oma süstemaatilises kirjanduse ülevaates esile, et pooltes 2003–2014 vaatluse all olnud disainipõhistes uuringutes viidi läbi vaid üks disainitsükkel. Sel juhul aga ei pruugi ilmnedada disainipõhise uuringu tõeline potentsiaal (*ibid.*).

Igas disainitsükklis sisaldub rida korduvaid etappe (tegevusi). Kirjanduses on enamasti esitatud 4-etapilised tsükliid (nt Reeves, 2006; McKenney & Reeves, 2012). Erinevate autorite pakutud etapid on üksteisega olemuselt üsna sarnased, sisaldades probleemi analüüsi, disainimist, testimist ning disaini hindamist (reflekterimist).

Disainipõhise uuringu formaati pole pädevusi (teadmisi, oskusi) hindavate testide väljaarendamiseks eriti kasutatud, küll aga on seda rakendatud digitaalsete õpikeskkondade, õppematerjalide, kursuste jms väljatöötamiseks (vt nt Anderson & Scatuck, 2012; Zheng, 2015; Tinoca *et al.*, 2022).

Uuringu eesmärk ja uurimisküsimused

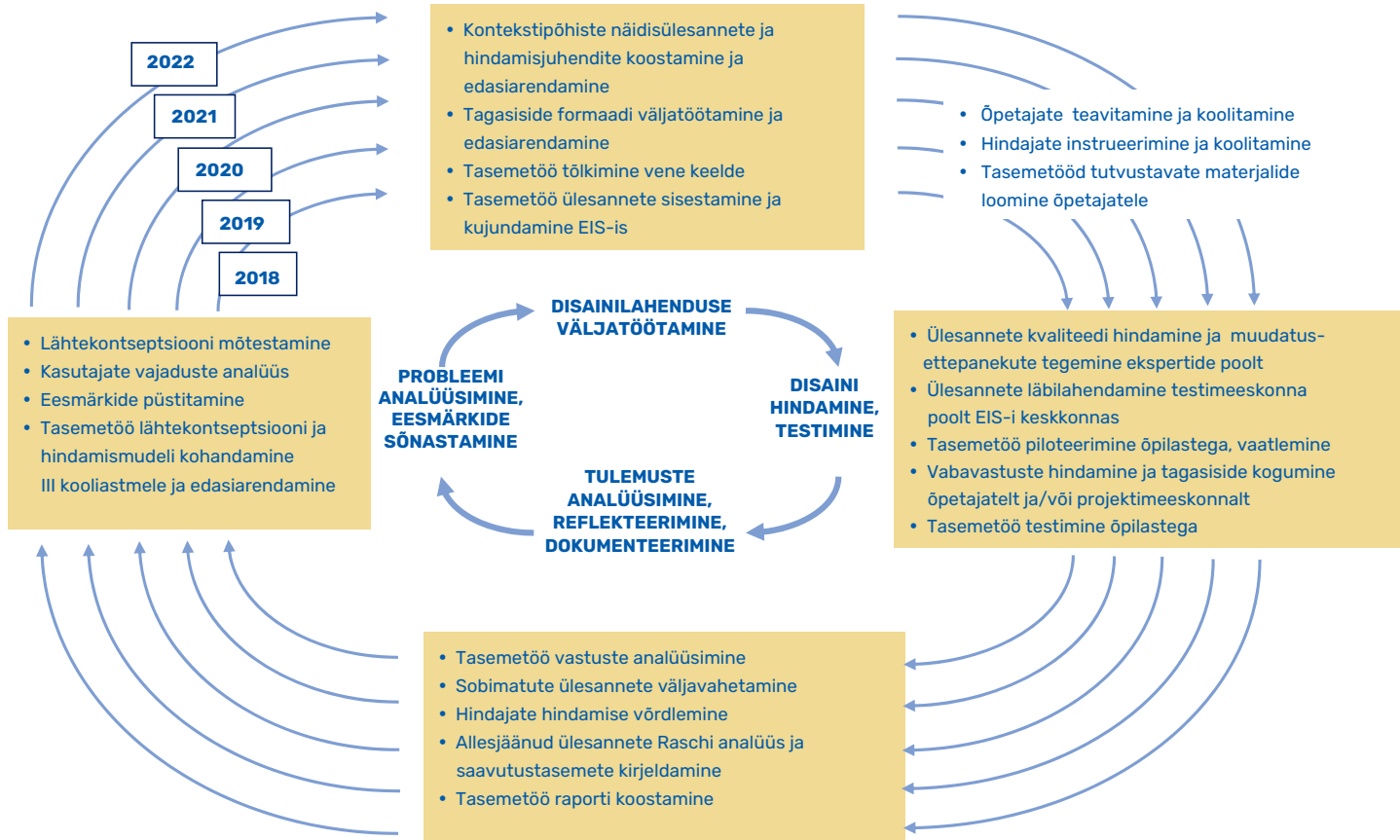
Selle disainipõhise uuringu eesmärk oli välja töötada e-tasemetöö (edaspidi tasemetöö), mis hindaks õpilaste III kooliastme lõpuks omandatud loodus-teadusliku pädevuse taset, nagu see on määratletud riiklikus õppekavas (PRÕK, 2023), ning annaks selle kohta kirjeldavat tagasisidet nii õpilasele, õpetajale, lapsevanemale kui ka haridusüldsusele. Tasemetöö väljatöötamise protsess toimus kooskõlas riikliku õppekava valdkonnakavade uuendamisega (2017–2023) ning uurijate ja praktikute (õpetajate, aineekspertide, testikeskkonna disainerite) omavahelises koostöös, et jõuda uudsete tasemetöö arendamise põhimõtteni.

Selle artikli eesmärk on kirjeldada tasemetöö disainimise protsessi, põhjendada tehtud disainiotsuseid ja iseloomustada põhjalikumalt saadud lõpptulemust, vastates kahele uurimisküsimusele.

- Millised on e-tasemetöö väljaarendamise peamised väljakutsed ning lahendused?
- Kuidas interakteeruvad õppijad loodud e-tasemetööga?

Metoodika

Tasemetöö töötati välja Tartu Ülikooli ning Haridus- ja Noorteameti (endise SA Innove) vahelises koostöös. Disainipõhisest uurimismeetodist tulenevalt koguti tasemetöö disaini sobivuse hindamiseks mitmekesiseid andmeid. Protsessi kompleksisusest tingituna on selles töös saadud andmetest ning tulemustest esitatud vaid osa. Tasemetöö disaini sobivust hinnati testitulemustest saadud statistiliste mudelite näitajate ning selle kaudu, kuidas protsessis osalejad saadud disainiga interakteeruvad. Vastavalt Reeves'ilt (2006) adapteeritud mudelile koosnes tasemetöö väljatöötamise protsess igas disainitsüklis neljast etapist (vt joonist 1).



Joonis 1. Disainimise tsüklid aastatel 2018–2022.

(1) Probleemi analüüsimine nii uurijate kui praktikute poolt, eesmärkide sõnastamine

Esimese tsükli puhul tähendas see tasemetöö kontseptsiooni (Pedaste *et al.*, 2017) ning seniste parimate praktikate mõtestamist, kasutajate (õpetajate, õpilaste, haridusüldsuse) vajaduste analüüsi ning eesmärkide sõnastamist TÜ uurijate, sh tegevõpetajate ning Harno tasemetöö arendajate koostöös. Kuna kolmandas kooliastmes õpetatakse loodusteadusi nelja iseseisva õppeainena (bioloogia, geograafia, füüsika, keemia), siis tuli esialgset, ainult ühele ainele (loodusõpetusele) suunatud lähtekontseptsiooni rakendada laiemas kontekstis, mis lisas tasemetöö disainilahenduse väljatöötamisele uusi väljakutseid.

Igas järgnevas tsüklis vaadati üle eelmises tsüklis esile kerkinud või lahendamata jäänud probleemid, seejärel modifitseeriti testimiste tulemuste analüüsi põhjal esialgset hindamismudelit ja sõnastati uue arendustsükli eesmärgid. Näiteks seati 2020. a eesmärgiks siduda III kooliastme tasemetöö hindamismudel paremini II kooliastme tasemetöö omaga.

(2) Tasemetöö ülesannete ja tagasiside formaadi (disainilahenduse) väljatöötamine

Tasemetöö disaini teine etapp (2018) algas hinnatavate tunnuste ja hindamismudeli eri tasemete kohta näidisülesannete (TÜ aineekspertidid) ning nendele tuginevate esimese põlvkonna tasemetöö ülesannete (loodusainete ekspertõpetajad koostöös TÜ aineekspertidega) koostamisega. Tuginedes läbiviidud testi tulemuste analüüsile, täiendati ja täpsustati ülesandeid kõigis järgnevas tsüklites, muutes sõnastust või ülesande tüüpi. Mõnest ülesandest tuli halbade statistiliste näitajate põhjal loobuda ja nende asemele uued koostada. Samal ajal koostasid TÜ aineekspertidid hindamisjuhendid avatud vastustega ülesannetele, mida järgnevas tsüklites pidevalt vastavalt hindajate tagasisidele ning õpilaste vastustele täiendati.

2019. aastal alustati testi tagasiside formaadi väljatöötamist nii õpilasele kui õpetajale, toetudes väljatöötatud hindamismudelile ning I ja II kooliastme tasemetöö tagasiside formaadile. Hilisemates tsüklites võeti tagasiside formaadi aluseks testitulemustest tuletatud Wrighti kaartidel (vt järgmist alapunkti) põhinev ehk empiiriline mudel, mida igal järgneval aastal vastavalt eelmise aasta testi tulemustele mingil määral modifitseeriti.

Tasemetöö disain hõlmas ka ülesannete sisestamise ja visuaalsete lahenduste ning tagasiside formaadi tehnilise lahenduse väljatöötamist Eksamite Infosüsteemi (EIS) keskkonnas, mida arendas Harno tasemetööde meeskond.

(3) Tasemetöö hindamine ja testimine

Väljaarendatud tasemetööd hindasid projektimeeskonnast väljaspool seisvad aineekspertid: 2–3 ekspertõpetajat ning 2–3 ainedidaktikut igast loodusainest. Saadud tagasiside põhjal tegi projektimeeskond ülesannetes vajalikud täiendused. Järgnevalt lahendas EIS-i keskkonnas tasemetöö läbi TÜ loodusteadusliku hariduse keskuse ja Harno laiendatud projektimeeskond, kelle ettepanekute põhjal tehti samuti vajalikud muudatused. Samamoodi hinnati tasemetöö ülesannete kvaliteeti iga järgneva tsükli jooksul. Seejärel piloteeriti tasemetööd õpilastega, kellelt koguti ühtlasi tagasisidet testi raskuse, huvitavuse, aga ka muude näitajate kohta, mida selles artiklis ei ole käsitletud.

Loodusvaldkonna tasemetöö teste on aastatel 2018–2022 läbi viidud nii eesti- kui venekeelsetes üldhariduskoolides ja pärast tasemetöö üleviimist 10. klassi ka kutseõppeasutustes. Ülevaate testimisest annab tabel 1. Kõigil aastatel oli eesmärk moodustada representatiivne valim, mida ei saavutatud vaid 2020. aastal COVID-19 pandeemia tõttu.

Tabel 1. Testimisel rakendatud ülesannete arv ning osalenud õpilaste arv vastavalt soorituskeelele ja koolitüübile

Aasta	Klass	Ülesannete arv	Testi sooritanud õpilaste arv vastavalt soorituskeelele ning õppeasutusele	Testi ametlik nimetus Eksamite Infosüsteemis
2018	9.	51	1226 (eesti); 274 (vene)	Põhikatsetus 1
2019	9.	37	539 (eesti); 269 (vene), üldharidus	Põhikatsetus 2
2020	9.	35	108 (eesti), üldharidus	Lisakatsetus
2021	10.	35	1363 (eesti); 453 (vene); üldharidus 1323; kutseharidus 493	Katseline tasemetöö
2022	10.	35	532 (eesti); 260 (vene); üldharidus 713, kutseharidus 79	Katseline tasemetöö

Igas disainitsükli küsiti Harno koostatud küsimustiku abil avatud vastuseid hinnanud õpetajate (15–30 (olenevalt aastast)) tagasisidet ülesannete ja hindamisjuhendite asjakohasuse ja arusaadavuse kohta.

(4) Tulemuste analüüsimine, reflekteerimine ja dokumenteerimine uurimisraportitena (Rannikmäe *et al.*, 2018, 2019, 2020, 2021a, 2021b, 2022)

Tasemetöö prototüübi piloteerimise tulemusi analüüsi, kasutades (1) kirjeldavat statistikat ning alates 2019 (2) uurivat ja kinnitavat faktoranalüüsi ning (3) üksikvastuste teoorial (IRT – *item response theory*) põhinevat Raschi

analüüsi illustreeritud Wrighti kaartidega. Andmed sisestati arvutuste tegemiseks standardiseerimata kujul ehk sisestati ülesannete eest saadud punktid, kus maksimumskoorid jäid kõigis ülesannetes 1 ja 3 punkti vahele. Uurivat faktoranalüüsi kasutati suure hulga testis olevate mõõdetavate tunnuste (üksik-küsimuste) põhjal latentsete tunnuste (faktorite) väljaselgitamiseks. Kinnitava faktoranalüüsi testiti saadud faktorstruktuuri (faktormudeli) statistilist headust. Raschi analüüsi kasutati nii testi väljatöötamisel kui analüüsimisel. Raschi analüüsil põhinev illustreeritud Wrighti kaart võimaldab viia õpilaste võimekuse (omandatus tase) ja ülesannete raskusastme ühisele skaalale ning hinnata testi üksik-küsimuste suhtelist raskusastet (Bond, 2015).

Avatud vastuste hindamisjuhendite asjakohasust uuriti kahe hindaja vahelise kooskõla põhjal, milleks leiti Coheni kapp (κ). 2018. aastal hinnati üle ligikaudu 10% sooritatud tasemetöödest, sama oli kavas ka 2023. aastal. Kooskõla loeti suurepäraseks, kui $\kappa > 0,8$; oluliseks, kui $0,6 < \kappa < 0,8$; mõõdukaks, kui $0,4 < \kappa < 0,6$; rahuldavaks, kui $0,2 < \kappa < 0,4$, ning puuduvaks, kui $0 < \kappa < 0,2$ (Harvey, 2021). Hindamisjuhendit otsustati üle vaadata ja muuta, kui $\kappa < 0,6$. Ülejäänud aastatel hindas projektmeeskond õpetajate antud punktide vastavust hindamisjuhendile jooksvalt vabavastuste kvalitatiivse analüüsimise käigus.

Tulemused ja arutelu

Disainipõhised uurimused on sageli avatud ja eksploratiivsed, nende eesmärk on mitte ainult hinnata sekkumise mõju, vaid ka aidata kaasa teoreetiliste raamistike ja hea praktika arendamisele hariduses (Cobb *et al.*, 2003), kõnealusel juhul loodusvaldkonna III kooliastme e-tasemetöö väljaarendamisele. Järgnevalt püütakse vastata kahele uurimisküsimusele: (1) *Millised on e-tasemetöö väljaarendamise peamised väljakutsed ning lahendused?*; (2) *Kuidas interakteeruvad õppijad loodud e-tasemetööga?*. Kuna õpilaste interakteerumine testiga oli samuti osa testiarendajate ees seisvatest väljakutsetest, siis esitatakse tulemused ja arutelu mõlema küsimuse kohta koos.

Tasemetöö ülesannete ning hindamisjuhendite disainimine

Arendusprotsessi alguses ehk „disainilahenduste väljatöötamise“ etapis oli projektmeeskonnale tõsiseks väljakutseks koostada selliseid ülesandeid, mis sobiksid valitud kontekstiga, vastaksid hindamismudeli hinnatavate tunnuste kirjeldustele ja hõlmaksid ühtaegu eri tasemeid. Sobivate ülesannete väljatöötamiseks korraldati ühiseid seminare, milles osalesid ülesandeid disainivad õpetajad ning ülikooli ja Harno tasemetöö arendajad. Töö käigus selgus, et Sjöströmi ja Talanqueri (2014) tutvustatud kontekstitüüpidest sobisid

probleemülesannete koostamiseks paremini sellised, mis võimaldasid sisse tuua ühiskondlikult olulisi probleeme (nt vaksineerimine). Samas sobisid teadmiste praktilisele rakendamisele suunatud kontekstid (nt katlakivi) hästi loodusteaduslikke teadmisi hindavate ülesannete koostamiseks.

Igas testi arendamise tsüklis täiustati olemasolevaid ülesandeid või koostati uusi, lähtudes põhimõttest, et need vastaksid paremini hindamismudelile ja et iga taseme hindamiseks oleks tagatud piisav arv ülesandeid. Viimast aitas saavutada ka arendajate otsus, et avatud vastusega ja mitme õige valikvastusega ülesanded võimaldavad määratleda mitut taset.

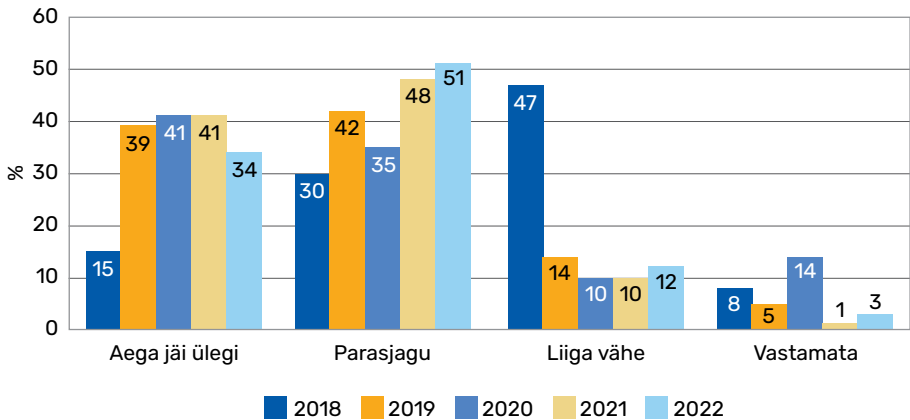
Ülesannete koostamisel kasutati erinevaid meediarikkaid stiimuleid, nagu fotod, joonised, videod, simulatsioonid jms. Paraku jäid näiteks videot ja arvuti-simulatsiooni sisaldavad ülesanded mitterahuldavate statistiliste näitajate tõttu tasemetöö lõppversioonist välja. Kindlasti tasub neid uuesti kasutada tasemetöö järgmise versiooni väljaarendamisel.

Hindamisjuhendite täiendamisel toetuti kolmele tagasisideallikale. Väikeseid täpsustusi tehti hindamisjuhendites igas disainitsükli „tulemuste analüüsimise ja hindamise“ faasis, kui hindajate töös leiti süsteemseid kõrvalekaldeid juhendist. 2018. a disainitsüklis hinnati avatud vastusega ülesannete hindamisjuhendeid kahe hindaja vahelise kooskõla (Coheni kapp) leidmisega. Kaheksa ülesande vastuste ülehindamisest selgus, et oluline ehk tugev kooskõla esines kolme, mõõdukas kolme ning nõrk ühe ülesande puhul, ühe ülesande puhul oli see puuduv. Tuginedes 2018. a püstitatud kriteeriumitele, töötati „disainilahenduse väljatöötamise“ etapis seetõttu ümber enamik avatud vastustega ülesandeid ning hindamisjuhendeid. Kolmandaks koguti „disaini hindamise ja testimise“ etapis hindamisjuhendite kohta tagasisidet hindajatelt küsimustiku abil ning selle põhjal täiendati juhendeid, et neid oleks võimalikult mugav kasutada.

Tasemetöö ülesannete arv

Esimene tasemetöö versioon koosnes 51 ülesandest. Nii mahukas test oleks aidanud tagada tulemuste suurema usaldusväärsuse (Kline, 1986) ning loodusainete õppesisu ja hinnatavate oskuste parema esindatuse (sisu valiidsuse), kuid ülesannete arvukuse tõttu ei jõudnud üle poole õpilastest kogu testi tasemetöök ettenähtud 120 minuti jooksul sooritada (vt joonist 2). Ülesannete arvu võimaldas 2019. aastal vähendada üleminek empiirilisele hindamismudelile „tulemuste analüüsi, reflekteerimise ja dokumenteerimise“ etapis, kui loobuti tingimusest, et iga taset ja tunnust ($N = 10$) peaks hindama kaks korda. Ülesannete arvu vähendati „disainilahenduse väljatöötamise“ etapis ka selliste ülesannete arvelt, mille tulemused korreleerusid kõrgelt mõne teise ülesandega, neist jäeti testi alles asjakohasem (kas alatesti läbiva „loo“ terviklikkuse või

hinnatava õppesisu olulisuse seisukohalt). Tasemetöö lõppversioon koosneb 35 ülesandest, mille lahendamise aega pidas 2022. aastal „disaini hindamise ja testimise“ etapis piisavaks 85% õpilastest, 34%-l jäi aega ülegi. Kuna testimist läbiviivatel õpetajatel on õpilaste eripärasid arvestades õigus testi tegemise aega vajadusel pikendada, otsustati jääda 35 ülesandega versiooni juurde.



Joonis 2. Õpilaste vastus küsimusele „Kas testi lahendamiseks oli piisavalt aega?“.

Avatud vastusega ülesanded vs. automaatselt hinnatavad ülesanded

Avatud vastusega ülesanded võimaldavad hinnata paremini õpilaste kõrgemat järku mõtlemist, sh probleemilahendusoskusi ja loovmõtlemist (Champagne, 2013). Samas eeldab avatud vastuste rohkus hindamisel suuremaid ressursse (Zhang *et al.*, 2022), nagu kvalifitseeritud hindajate olemasolu ning hindamisele kuluv aeg. Ressursside kokkuhoiduks muudeti iga järgneva disainitsükli „disainilahenduse väljatöötamise“ etapis osa madalamat järku kognitiivseid mõtlemisprotsesse (teadmiste rakendamist) eeldavaid avatud vastustega ülesandeid automaatselt hinnatavateks ülesanneteks. Samuti muudeti valikvastuseliseks mõned hästi defineeritud probleemide lahendamisele suunatud ülesanded, mille puhul eksisteeris n-ö õige ehk parim lahendus, mis aga vaatamata ülesande vormile nõudsid lahendajalt kõrgemat järku mõtlemisprotsesside rakendamist – õpilasel tuli arvesse võtta mitme loodusteadusliku nähtuse koosmõju ja lisaks igapäevaelulist konteksti – ning mis seetõttu osutusid katsetamisel isegi kõrg- või tipp-taseme ülesanneteks.

Avatud vastustega ülesannete hulga vähendamine oli disaini loomulik osa: esmalt tuli selgitada välja õpilaste tüüpilised vastused, et siis nende põhjal koostada valikvastused või täiendada avatud vastustega ülesannete hindamisjuhendeid näidisvastustega. Toetudes Champagne’ile (2013), ei loobutud avatud vastustest probleemide lahendamise ja loodusteadusliku teksti kirjutamise ning

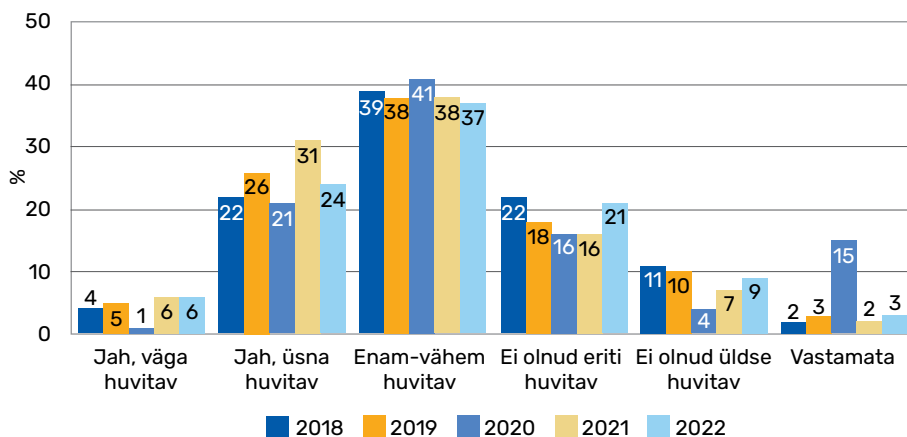
eriti just kõrg- ja tipptaseme ülesannetes, mis eeldasid probleemilahendamise oskusi ja loovmõtlemist.

Testimise üleviimine 9. klassist 10. klassi algusesse

Aastatel 2018–2020 viidi test läbi 9. klassi õpilaste seas õppeaasta teises pooles („disaini hindamise ja testimise“ etapp). Valitud aeg osutus aga keerukaks lõpuksamiteks valmistumise ning gümnaasiumi sisseastumiskatsete tõttu. Seega viidi testimine Harno eestvedamisel 10. klassi algusesse, mida toetab asjaolu, et järgmise kooliastme alguses viiakse tasemetööd läbi ka I ja II kooliastme õpilastele. Gümnaasiumi alguses on õpetajatel võimalik testimise tagasisidest saada hea ülevaade (uute) õpilaste teadmistest ja oskustest, et planeerida edasist õpetamist.

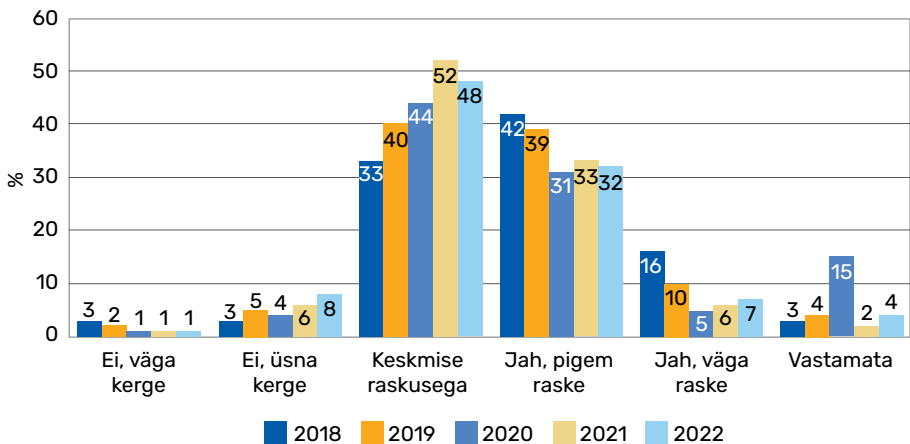
Tajutud huvi ja raskus testi sooritamisel

Kuna madala panusega testide puhul võib järelduste kehtivust ohustada õpilaste madal huvi testi lahendada (Finn, 2015), siis uuriti seda iga tsükli „disainilahenduse hindamise ja testimise“ faasis tasemetöö lõpus küsimusega „Kas test oli sinu jaoks huvitav?“. Õpilaste vastused on olnud läbi aastate suhteliselt stabiilsed (vt joonist 3). Erandiks on vaid 2020. aasta tulemused, kus küsimusele jättis vastamata 15% õpilastest. Kui kõnealuse aasta vastused vaatluse alt välja jätta, võib öelda, et ligikaudu kaks kolmandikku vastajatest on tasemetööd pidanud enam-vähem või väga huvitavaks, mis omakorda tähendab, et ühe kolmandiku jaoks on test olnud pigem ebahuvitav või üldse mitte huvitav. Üheks võimalikuks huvi tõstmise võimaluseks oleks edaspidi nn meediarikaste stiimulite mitmekesistamine tasemetöö ülesannetes (Boyle & Hutchison, 2009).



Joonis 3. Õpilaste vastus küsimusele „Kas test oli sinu jaoks huvitav?“.

Joonise 4 põhjal on näha, et õpilased on testi aastate vältel tajunud keskmiselt pigem raskena. Selle põhjuseid võib olla mitu. Õpilastele võis olla harjumatu nii tasemetöö ülesannete kontekstipõhisus, mis eeldab teadmiste ülekanndmist ja rakendamist elulises kontekstis, kui ka asjaolu, et lisaks aineteadmistele hinnati tasemetöös ka uurimuslikke, avatud probleemide lahendamise ning kommunikatsioonioskusi, mis uurimustele (Henno *et al.*, 2017; Valdmann *et al.*, 2012) toetudes ei ole siiani saanud veel loodusainete tundides kujundatavate õpitulemuste enesestmõistetavaks osaks.



Joonis 4. Õpilaste vastus küsimusele „Kas test oli sinu jaoks raske?”.

Teisalt võib näha, et nende õpilaste arv, kes tajuvad testi üsna või väga raskena, on aja jooksul vähenenud. Selle taga võib olla nii asjaolu, et tasemetöö ise on hakanud mõju avaldama õpetajate tunnis rakendatavatele õpetamis- ja hindamismeetoditele (Rosin *et al.*, esitatud), kui ka see, et aja jooksul on vähenenud ülesannete, sh avatud vastustega ülesannete arv testis ning seega õpilaste võimalik kognitiivne ülekoormus. Kuna tasemetöö sisaldab ka raskemaid ülesandeid, et tuvastada tippsooritajaid, võisid just raskemad ülesanded panna osa õpilasi tajuma testi pigem või väga raskena. Tagamaks testi eristusvõimet, ei olnud eraldi eesmärk testi ülesannete raskusastet vähendada, kuid õpilaste poolt tajutud tasemetöö raskust tasub ka edaspidi tasemetöö läbiviimisel monitorida: ootuspärane oleks, et nende õpilaste osakaal, kes tajuvad tasemetööd raskena, väheneks ka edaspidi tulenevalt muutustest õpetamisel.

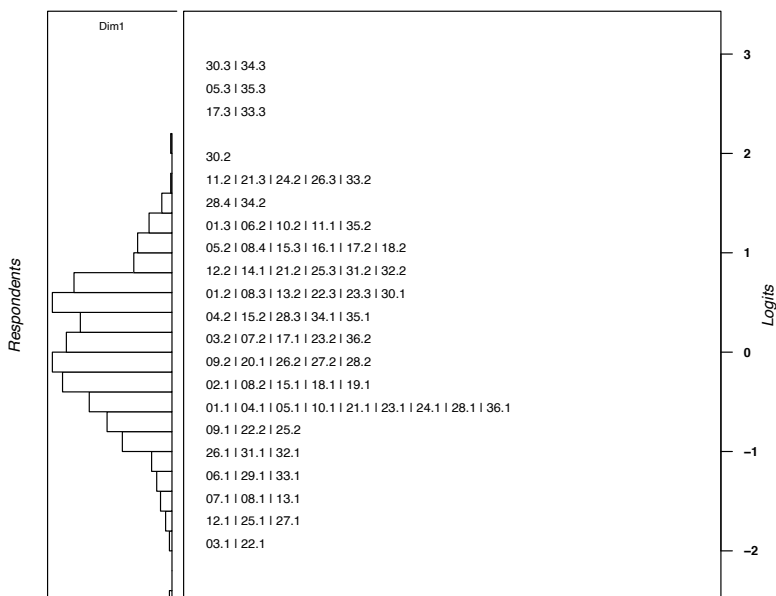
Õpilaste tasemetööde ülesannete vastuste analüüsi tulemused

Tuginedes tasemetöö tulemustele, loodeti uuriva faktoranalüüsi käigus leida mitmefaktoriline lahendus, et vähendada mõõdetavate tunnuste arvu. Kahjuks saadi mitme tsükli vältel (2019–2022) „tulemuste analüüsimise, reflekteerimise

ja dokumenteerimise“ etapis uuriva ja kinnitava faktoranalüüsi tulemusena korduvalt mitmefaktorilised mudelid, millel olid küll suhteliselt head sobivuse näitajad, kuid mille faktorid polnud sisuliselt tõlgendatavad. Seepärast otsustati jääda ühefaktorilise lahenduse juurde, mille sisereliaablust ehk küsimuste sise- mist tähenduslikku kooskõla hinnati Cronbachi alfa kordajaga, mis on olnud erinevate testivariantide korral iga kord >0,80, testi viimase versiooni puhul (2022) oli see 0,88. Ühefaktorilise mudeli headuse näitajad jäid analüüsitava perioodi (2019–2022) jooksul lubatavasse vahemikku ($0,03 < RMSEA < 0,08$, $CFI \geq 0,90$, $TLI \geq 0,90$), mis näitab mudelite head sobivust nende andmestike puhul, ning need näitajad on muutunud iga aastaga paremaks.

Kuigi testi ülesannete sisse- või väljajätmise üle otsustamisel võeti esialgu faktorkaalu alumiseks piiriks 0,40, otsustati ülesannete sisust ning testi kontseptsioonist tulenevatel põhjustel igal aastal testi alles jätta paar ülesannet, mis olid kehtestatud piirist madalama faktorkaaluga ($0,30 < \dots < 0,40$).

Sarnaselt Pedaste jt (2021) uurimusega tuletati Wrighti kaardi põhjal (vt joonist 5) üksikülesannete vastavus tasemetele (baas-, kesk-, kõrg- ja tipp- tase) ja ülesannete iseloomust ning saadud tulemusest lähtuvalt vastava taseme õpitulemuste kirjeldused, mida hiljem kasutati õpilasele antava tagasiside koostamiseks, kirjeldades saavutatud taset ning andes soovitusi järgmise pädevustaseme saavutamiseks. Selleks, et saavutada mingi tunnuse arvestuses taset x, pidi õpilasel olema saavutatud vähemalt pool selle taseme ülesannete punktidest.



Joonis 5. Tasemetöö ülesannete Wrighti kaart.

Kui III kooliastme tasemetöö väljaarendamisel võeti aluseks I ja II kooliastme õpitulemuste hindamismudel, mille hinnatavate tunnuste tasemete kirjeldustele lisati kõrgemaid tasemeid, siis osutus II kooliastmega sarnase tagasiside väljatöötamine kummagi testi tulemuste faktorstruktuuride erinevuse tõttu – II kooliastmes saadi 4-faktoriline sisuliselt tõlgendatav lahendus (Pedaste *et al.*, 2021) – ülimalt keerukaks. Sarnane faktorstruktuur (sh faktorite sisuline tõlgendus) võimaldaks anda läbivalt ühtlasemat tagasisidet kooliastmete kaupa ning muuta õpilase arengu (või õpilasel oma arengu) jälgimise ühelt kooliastmelt teisele lihtsamaks.

Sisuliselt tõlgendamatu faktorite saamise põhjuseks mitmedimensiooniliste lahendite puhul pakuvad autorid asjaolu, et tasemetöö ülesannetes esineb hulk muutujaid, mis ülesannetes omavahel kombineeruvad ning mõjutavad seega faktorstruktuuri. Lisaks ainete õppesisule, mis pärineb (1) neljast erinevast loodusainest, on muutujaks ka (2) ülesande kontekst; ülesanded hindavad loodusteaduslikku pädevust, mis hõlmab (3) üheksat hinnatavat tunnust, ning seetõttu peab õpilane ülesannete lahendamisel rakendama väga erinevaid kognitiivseid operatsioone. Muutujaks on tasemetöös ka (4) ülesannete vorm, hõlmates nii arvutihinnatavaid valik- ja lühivastuseid kui ka käsitsi hinnatavaid avatud vastuseid.

Laia valdkonna teadmiste kontrollimiseks kasutatavate standardiseeritud testide kalduvust anda ühedimensionaalseid lahendusi on toodud välja ka teistes uurimustes (Wainer *et al.*, 2000; Gaffney *et al.*, 2010). Gaffney jt (2010) rõhutasid seejuures, et kui mitmedimensiooniline lahendus ka saadakse, on saadud faktorite ning üksikküsimuste aluseks olnud teoreetiliste kategooriate omavaheline vastavus pettumust valmistavalt madal. Samuti püstitasid nad intrigeeriva küsimuse, kas n-ö mitme õppeaine kombineeritud testid ei mõõda tegelikult vaid ühte latentset tunnust. Kõnealuse tasemetöö kontekstis võiks selle nimetada õpilase loodusteaduslikuks pädevuseks.

Lõplik hindamismudel

Tasemetöö aluseks olev lõplik hindamismudel (vt Harno, 2023b) hõlmab üheksat hinnatavat tunnust, mis on ülevaatlikkuse mõttes liigendatud neljaks sisu alusel eristuvaks komponendiks: loodusteaduslikud teadmised, uurimulikud oskused, probleemilahendusoskused, kommunikatsioonioskused. Lõplik hindamismudel on saadud lähtekontseptsioonis esitatud teoreetilise mudeli korrigeerimisel, toetudes õpilaste testimise tulemuste analüüsile. Võrreldes lähtekontseptsiooniga on selles üks hinnatav tunnus vähem, kuna protsessi käigus ühendati sisulise sarnasuse alusel tunnused „probleemi püstitamine“

ja „uurimisküsimuse ning hüpoteesi püstitamine“, et vähendada hinnatavate tunnuste ning seega ka tasemetöö ülesannete arvu.

Saadud hindamismudel sisaldab nii uuendatud Bloomi (Anderson & Krathwohl, 2001) kui ka SOLO taksonoomia (Biggs & Collis, 2014) õpitulemustele iseloomulikke jooni. Näiteks sobib Bloomi taksonoomia hästi uurimuslike oskuste tasemete kirjeldamiseks: kesktasemel suudab õpilane välja valida sobivad katsevahendid vastamaks etteantud uurimisküsimusele (rakendamine), kõrgtasemel analüüsima etteantud katse käiku ning tipptasemel hindama katse kvaliteeti. Samas tekkis avatud vastusega probleemilahenduse ning loodusteadusliku teksti kirjutamise oskuste kirjeldamisel vajadus lahutada ülesande tase vastamise tasemest, mida võimaldab paremini just SOLO taksonoomia (Adeniji *et al.*, 2022), sest nendes ülesannetes sõltub vastuse tase selle põhjalikkusest, korrektsusest, seoste loomisest erinevate valdkondade vahel, mitte niivõrd küsimusest endast (mis on avatud). SOLO taksonoomia sobib paremini ka loodusteaduslike nähtuste ja protsesside selgitamise hindamiseks: baastasemel suudab õpilane selgitada üksikut loodusnähtust (SOLO järgi „unistrukturealne mõtlemine“), kõrg- ja tipptaseme puhul aga keerukamaid põhjus-tagajärg-seoseid erinevate loodusnähtuste vahel („seostatud mõtlemine“) ning igapäevaelulises või globaalses kontekstis („üldistatud mõtlemine“).

Hindamismudeli tasemete kirjeldused võtavad mingil määral arvesse ka ülesande vormi. Näiteks kirjeldatakse õpitulemust „valib õige selgituse“ juhul, kui tegemist on valikvastuselise ülesandega, ning „selgitab“ või „põhjendab“, kui õpilane vastab avatud vastusega küsimusele oma sõnadega.

Tasemetöö lõplik versioon

Tasemetöö lõplik versioon, milleni jõuti viie disainitsükli tulemusena, koosneb neljast kontekstist (alatestist), millest igaüks esindab ühte konteksti (lugu), ning kokku 35 ülesandest, mis kokku määratlevad õpilaste loodusteadusliku pädevuse taseme (baas-, kesk-, kõrg- või tipptase). Kui õpilase sooritus jääb ka baastasemest nõrgemaks, saab ta nulltaseme. Kõiki teisi tasemeid kirjeldavad 9 teoreetilist tunnust, mis sisemise loogika ning ülevaatlikkuse huvides on liigendatud neljaks komponendiks (vt tabelit 2). Iga alatesti ülesanded koonduvad ümber ühe konteksti, millest lähtuvalt esitatakse õpilastele ülesanded lahendamiseks. Kontekstid on seotud igapäevaeluga ning hõlmavad endas nii teadmiste rakendamisele suunatud kontekste kui ka ühiskondlikult oluliste loodusteadusliku sisuga probleemide käsitlemist.

Tabel 2. Tasemetöös hinnatavad loodusteadusliku pädevuse komponendid, tunnused ja seda hindavate ülesannete tüübid (AV – avatud vastus, LV – lühivastus, VV – valikvastus) ning vastavad näited

LK pädevuse komponendid ja tunnused	AV	LV	VV	Kokku	Näited peamistest ülesannete tüüpidest
1. Loodusteaduslik ainesisu					
Õpilane:					
Loodusnähtuste ja protsesside ning nende vaheliste põhjus-tagajärg seoste selgitamine			3	3	Reastab protsessid õiges järjekorras.
Loodusteaduslike mõistete, sümbolite ja ühikute kasutamine		1	4	5	Valib etteantud olukorda sobivad ühikud; arvutab seose põhjal vastuse.
Arusaamine loodusteaduslikest mudelitest	2		4	6	Lohistab mõisted mõistekaardile; selgitab oma sõnadega mudelit (joonist).
2. Uurimuslikud oskused					
Uurimisküsimuse või hüpoteesi püstamine			2	2	Valib situatsioonikirjelduse põhjal sobiva uurimisküsimuse; sõnastab graafiku põhjal hüpoteesi.
Katse kavandamine ja selle kvaliteedi hindamine	3		2	5	Reastab situatsioonikirjelduse põhjal katse etapid; selgitab oma sõnadega, miks jätab kirjeldatud katse kvaliteet soovida.
Info analüüsimine ning järelduse tegemine	1	1	2	4	Analüüsib andmeid, mis on antud tabeli või graafiku kujul, ning teeb järeldused.
3. Probleemi lahendamine ja otsuse tegemine					
Probleemide lahendamine ja otsuste tegemine	3		4	7	Selgitab etteantud situatsiooni põhjal probleemi olemust ning järjestab tegevused, et kõrvaldada kirjeldatud süsteemis rike (Jonassen (2011) põhjal hästi defineeritud probleem); kaalutleb kahe sotsiaalse otsustuse variandi vahel (dilemma) ning põhjendab tehtud otsust mitmekülgeselt, sh väljendades oma väärtushinnanguid (halvasti defineeritud probleem (<i>ibid.</i>)).
4. Kommunikatsioonioskused					
Info otsimine ning selle usaldusväarsuse hindamine	1		1	2	Leiab internetist vajaliku info; hindab etteantud veebilehe usaldusväarsust, toob välja aspekte, mis toetavad tema hinnangut.
Arusaamine loodusteaduslikust tekstist ning selle koostamine	1			1	Koostab etteantud tekstilise ja pildilise info põhjal oma sõnadega kokkuvõtte (50–60 sõna).

Avatud vastusega probleemilahendamise ja kommunikatsioonioskuste ülesannetes määratleb õpilase vastuse taseme (null-, baas-, kesk-, kõrg- või tipptase) vastuse sisukus (väljatoodud aspektide arv, eriliigilisus, loogilisus ning loodusteaduslik korrektsus), mitte niivõrd ülesanne ise.

Tagasiside õpilasele, õpetajale ning haridusüldsusele

Tasemetöö tagasiside annab õpilasele teada tema loodusteadusliku pädevuse taseme (null-, baas-, kesk-, kõrg-, tipptase) ning vastavalt tasemele tema olemasoleva pädevuse kirjelduse üheksa tunnuse kaudu. Lisaks annab tagasiside õpilasele soovitusi edasiste õpieesmärkide püstitamiseks – nendeks on õpitulemused, mis on hindamismudelil saavutatud tasemest ühe astme võrra kõrgemal ning mis suure tõenäosusega asuvad õpilase lähima arengu tsoonis. Seega saab õpilane võimaluse tagasisides pakutavate õpieesmärkide kaudu reguleerida ise aktiivselt oma edasist õppimist (Hattie & Timperley, 2007; Zimmerman, 2000). Näiteks saab baastaseme saavutanud õpilane kommunikatsioonioskuste ühe tunnuse, loodusteadusliku teksti kirjutamise oskuse kohta järgmist tagasisidet (joonis 5).

...

Sa oskad

- koostada lihtsa sisuga lühikest loodusteaduslikku teksti, kasutades valdavalt tavakeelt.

Sinu järgmiseks eesmärgiks võiks olla

- osata koostada teksti, kasutades loodusteaduslikke mõisteid.

...

Joonis 5. Väljavõte õpilasele antavast tagasisidest.

Õpilasel ja tema vanemal on tagasiside kaudu ligipääs ning võimalus tutvuda loodusteadusliku pädevuse erinevatele tasemetele vastavate näidisülesannetega. Õpetaja saab samuti tagasisidet nii õpilase kui õpilaste rühma (klassi) kohta, st ta saab teada nende saavutatud tasemed, tasemete kirjeldused ning võimaluse tutvuda iga taseme näidisülesannetega. Lisaks saab õpetaja soovitusel, kuidas tagasisidet rakendada. Välja on töötatud kirjalik juhend ja video tasemetöö tulemuste kohta antava tagasiside tõlgendamiseks, mõlemad on õpetajatele kättesaadavad Harno kodulehe kaudu.

Tasemetööde kokkuvõtted valimis olnud õpilaste loodusteadusliku pädevuse tasemest, trendidest, sh suuremate vastajate rühmade vahelistest eripäradest (poisid vs. tüdrukud, vene keeles vastanud vs. eesti keeles vastanud õpilased) on laiema üldsuseni jõudnud Harno kodulehel olevate tasemetöö kokkuvõtete ning töörühma esinemiste kaudu hariduskonverentsidel ja teabepäevadel.

Järeldused ja ettepanekud

Kvaliteetse tasemetöö väljatöötamine on keerukas, süstemaatilist tegevust nõudev disainiülesanne, mille juures tuleb arvesse võtta arvukaid lähtetingimusi ja piiranguid, teha kompromisse sisuliselt ja statistiliselt näitajatelt parimate lahenduste vahel, kogudes disaini kasutajatelt (peamiselt õpilastelt) andmeid, et jõuda kasutajasõbraliku lahenduseni. Selline tasemetöö koosneb optimaalsest arvust ülesannetest, mida õpilased tajuvad mõõdukalt raskena ja mille lahendamise vastu tunneb huvi enamik õpilastest. Seejuures tuleb protsessi käigus kogu aeg silmas pidada õppekava eesmärke ning võtta arvesse õpetajate olemasolevaid arusaamu hindamisest, et tasemetöö ja selle tulemuste kohta käivat teavitustegevust sobivaks kohandada.

Erinevalt enamikust disainipõhistest haridusuuringutest, milles tehti läbi vaid üks disainitsükkel, viidi selles töös läbi viis tsükli, mis aitab autorite arvates kaasa saadud tasemetöö lahenduse asjakohasusele. Arendustegevuse tulemusena töötati välja disainilahendus kontekstipõhisele e-tasemetööle, mis hindab õpilaste loodusteaduslikku pädevust ning mida saab kasutada õpilaste üleriiklikuks hindamiseks (monitoorimiseks). Lisaks annab see sõnalist tagasisidet õpilasele tema olemasoleva taseme kohta, teeb soovitusi järgmiste õpi-eesmärkide püstitamiseks ning õpetajale üksiku õpilase ja õpilasrühma taseme ning spetsiifiliste õpivajaduste kohta. Lapsevanemal aitab see toetada lapse edasisi hariduslikke valikuid. Uudsetest disainipõhimõtetest, mida selles töös kirjeldati ja põhjendati, võiks olla kasu järgmiste kontekstipõhiste tasemetööde arendajatel nii loodusteadusliku pädevuse kui ka muude valdkondlike pädevuste diagnostilisel hindamisel. Kuigi väljatöötatud tasemetöö versiooni n-õ eluiga saab peagi läbi, võimaldab väljatöötatud meetodika arendada järgmist tasemetöö versiooni tunduvalt energia- ja ajasäästlikumalt.

Tasemetööd disainides ilmnes ka üks põhimõtteline probleem. Nimelt ei olnud õpilaste üksikülesannete lahendamise tulemuste põhjal läbi viidud uuriva ja kinnitava faktoranalüüsi abil võimalik ühelgi aastal saada sellist mitmemensioonilist faktorstruktuuri, mille faktorid oleksid sisuliselt tõlgendavad. Autorite arvates on põhjuseks eelkõige rida muutujaid (õppeaine, hinnatava õpitulemuse olemus, ülesande vorm jm), mis ülesannetes omavahel erinevalt kombineeruvad, muutes töökindlate ning sisuliselt koherentsete faktorite saamise ning latentse tunnuse defineerimise väga raskeks. Seetõttu pakutakse selles töös välja ühedimensiooniline lahend, mille latentse tunnuse võibki nimetada loodusteaduslikuks pädevuseks. Taseme määratlemise ja tagasiside aluseks oleva ühefaktorilise mudeli sobivusnäitajate põhjal saab järeldada mudeli head sobivust testi andmetega.

Kuna I-II ning III kooliastme tasemetööd arendasid välja osaliselt erinevad meeskonnad, olid tulemuseks ka vormiliselt mõnevõrra erinevad tagasiside vormid. Järgnevates disainitsüklites võiks otsida võimalusi ühtlustada II ja III kooliastme tasemetöö tagasiside vorme, et nii õpilasel endal, õpetajal kui ka lapsevanemal oleks lihtsam jälgida õpilase loodusteadusliku pädevuse arengut kooliastmete kaupa.

Seda kogemust tasemetöö väljaarendamisel võib Eesti tingimustes pidada unikaalseks, kuna see võimaldas:

- kujuneda tervel ekspertide kogukonnal (ülikooli loodusainete teadlased ja didaktikud, Harno tasemetööga seotud spetsialistid, üldhariduskoolide ekspertõpetajad ning oma kooli õpilaste vastuste hindamisse kaasatud loodusainete õpetajad), kes püüdsid mõtestada viie disainitsükli jooksul kaasaegset nägemust pädevuspõhisest ja õppimist toetavast hindamisest ning arendada oma sellekohaseid pädevusi; kogukonna arvukus tagab ühtlasi tasemetöö edasiste arenduste jätkusuutlikkuse;
- luua õppekava suhtes valideeritud raamistiku kontekstipõhiste ülesannete koostamiseks, mis oli võimalik tänu õpilaste vastuste süstemaatilisele analüüsile ning hindamismudeli järkjärgulisele peenhäälestamisele;
- luua disainiprotsessi käigus väljakukkunud ülesannetest panga, mida saab kasutada edasiarendamiseks ning pädevuse tasemeid illustreerivate avalike ülesannete loomiseks;
- teha hariduspoliitilisi otsuseid õpetajate koolitusvajaduste kohta, et tõhustada tasemetööst saadava tagasiside rakendamist.

Kuigi tasemetöö on loonud eeldused õpilaste ennastjuhtiva õppimise toetamiseks, vajab tegelik praktika, sh kuidas õpetaja saaks õpilast aidata edasiste õpieesmärkide püstitamisel, ning põhjalikum mõistmine, mida õpilane tasemetööst arvab ja kuidas seda neile asjakohasemaks, huvitavamaks ning olulisemaks muuta, veel uurimist.

Piirangud

Õpilaste lähima arengu tsoon on hüpoteetiline konstrukt, mille ulatus võib Vögotski järgi (Vygotsky & Cole, 1978) õpilaseti olulisel määral varieeruda. Kuna selle ulatust on kõnealuse tasemetöö raames raske uurida, siis soovitab tasemetöö tagasiside kõigile õpilastele n-ö võrdse sammu võrra edasi liikunud õpieesmärki (v.a tipptasemel).

Aastate jooksul on valimite suurused ja esinduslikkus üldkogumi suhtes kõikunud. Näiteks oli 2020. aastal COVID-19 puhangu tõttu pilootuuringus väga vähe testi tegijaid, mis vähendab saadud tulemuste ja tehtud järelduste üldistusjõudu.

Tänuõnad

Loodusvaldkonna III kooliastme e-tasemetöö väljatöötamine toimus digipöörde programmi (2018–2022) tegevuse „Kaasaegse ja uuendusliku õppevara arendamine ja kasutuselevõtt“ raames, mida rahastab Euroopa Sotsiaalfond. Lisaks tänavad autorid Haridus- ja Noorteameti e-tasemetöö arendajaid ning TÜ ökoloogia ja maateaduste instituudi loodusteadusliku hariduse keskuse kolleege meeldiva koostöö eest. Suur tänu kõigile, kes on osalenud tasemetöö ülesannete koostamisel, tasemetöö piloteerimisel, hinnanud avatud vastuseid, andnud tagasisidet ülesannete ning hindamisjuhendite asjakohasusele ning panustanud seeläbi lõpliku tasemetöö valmimisse.

Kasutatud kirjandus

- Aksen, M., Jürimäe, M., Nõmmela, K., Saarsen, K., Sillak, S., Eskor, J., Vool, E., & Urmann, H. (2018). Eesti üldhariduskoolides kasutatavad hindamissüsteemid. Tartu Ülikool. https://www.hm.ee/sites/default/files/uuringud/hindamine_lopparuanne_15.okt_loplik.pdf.
- Anderson, L. W., & Krathwohl, D. R. (toim) (2001). *A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives*. New York: Addison Wesley Longman, Inc.
- Anderson, T., & Shattuck, J. (2012). Design-based research: A decade of progress in education research? *Educational Researcher*, 41(1), 16–25. <https://doi.org/10.3102/0013189X11428813>
- Assessment Reform Group (1999). *Assessment for Learning: Beyond the Black Box*. Cambridge: University of Cambridge School of Education.
- Baker, P., & Schmude, M. (2022). Structure of the observed learning outcomes (SOLO) model: A mixed-method systematic review of research in mathematics education. *Eurasia Journal of Mathematics, Science and Technology Education*, 18(6), 1–17. <https://doi.org/10.29333/ejmste/12087>
- Barab, S., & Squire, K. (2004). Design-based research: Putting a stake in the ground. *The Journal of the Learning Sciences*, 13(1), 1–14. https://doi.org/10.1207/s15327809jls1301_1
- Biggs, J. B., & Collis, K. F. (2014). *Evaluating the Quality of Learning: The SOLO Taxonomy (Structure of the Observed Learning Outcome)*. Academic Press.
- Bolhuis, S. (2003). Towards process-oriented teaching for self-directed lifelong learning: A multidimensional perspective. *Learning and Instruction*, 13(3), 327–347. [https://doi.org/10.1016/S0959-4752\(02\)00008-7](https://doi.org/10.1016/S0959-4752(02)00008-7)
- Bond, T. (2015). *Applying the Rasch Model: Fundamental Measurement in the Human Sciences* (3rd ed.). Routledge.

- Boyle, A., & Hutchison, D. (2009). Sophisticated tasks in e-assessment: What are they and what are their benefits?. *Assessment & Evaluation in Higher Education*, 34(3), 305–319. <https://doi.org/10.1080/02602930801956034>
- Champagne, A. B. (2013). Content to be assessed across the history of the national assessment of educational progress. D. Corrigan, R. Gunstone, & A. Jones (Eds.), *Valuing Assessment in Science Education: Pedagogy, Curriculum, Policy*. (pp. 119–151). Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-94-007-6668-6_7
- Choi, K., Lee, H., Shin, N., Kim, S.-W., & Krajcik, J. (2011). Re-conceptualization of scientific literacy in South Korea for the 21st century. *Journal of Research in Science Teaching*, 48(6), 670–697. <https://doi.org/10.1002/tea.20424>
- Cobb, P., diSessa, A., Lehrer, R., & Schauble, L. (2003). Design experiments in educational research. *Educational Researcher*, 32(1), 9–13. <https://doi.org/10.3102/0013189X032001009>
- Csapó, B., & Molnár, G. (2019). Online diagnostic assessment in support of personalized teaching and learning: The eDia system. *Frontiers in Psychology*, 10 (1522), 1–14. <https://doi.org/10.3389/fpsyg.2019.01522>
- Daniels, H. (2007). Pedagogy. In H. Daniels, M. Cole, & J. V. Wertsch (Eds.), *The Cambridge companion to Vygotsky*. (pp 307–331). Cambridge University Press. <https://doi.org/10.1017/CCOL0521831040.013>
- Fensham, P. J., & Rennie, L. J. (2013). Towards an authentically assessed science curriculum. D. Corrigan, R. Gunstone, & A. Jones (Eds.), *Valuing Assessment in Science Education: Pedagogy, Curriculum, Policy* (pp. 69–100). Dordrecht: Springer. https://doi.org/10.1007/978-94-007-6668-6_5
- Finn, B. (2015). Measuring motivation in low-stakes assessments. *ETS Research Report Series*, 2015(2), 1–17. <https://doi.org/10.1002/ets2.12067>
- Gaffney, T. W., Cudeck, R., Ferrer, E., & Widaman, K. F. (2010). On the factor structure of standardized educational achievement tests. *Journal of Applied Measurement*, 11(4), 384.
- Gweon, H., & Schonlau, M. (2022). Automated classification for open-ended questions with BERT. *arXiv preprint arXiv:2209.06178*.
- Haridus- ja Noorteamet (Harno) (2023a). *Tasemetööd*. <https://harno.ee/tasemetood>.
- Haridus- ja Noorteamet (Harno) (2023b). *III kooliastme testi materjalid 2022: Loodus-teadusliku kirjaoskuse komponendid ja tasemed koos näidisülesannetega III kooliastmes*. <https://projektid.edu.ee/pages/viewpage.action?pageId=132157214>.
- Haridus- ja Teadusministeerium (2014). *Eesti elukestva õppe strateegia 2020*. https://www.haridusfoorum.ee/images/haridusstrateegia/Eesti_elukestva_õppe_strateegia_loplik.pdf.
- Harvey, N. D. (2021). A Simple guide to inter-rater, intra-rater and test-retest reliability for animal behaviour studies. <https://doi.org/10.31219/osf.io/8stpy>
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81–112. <https://doi.org/10.3102/003465430298487>

- Henno, I., Kollo, L., & Mikser, R. (2017). Eesti loodusainete õpetajate uskumused, õpetamispraktika ja enesetõhusus TALIS 2008 ja 2013 uuringu alusel. *Eesti Haridusteaduste Ajakiri*, 5(1), 268–296. <https://doi.org/10.12697/eha.2017.5.1.09>
- Hsieh, H.-F., & Shannon, S. E. (2005). Three approaches to qualitative content analysis. *Qualitative Health Research*, 15(9), 1277–1288. <https://doi.org/10.1177/1049732305276687>
- Höttecke, D. & Allchin, D. (2020). Reconceptualizing nature-of-science education in the age of social media. *Science Education*, 104(4), 641–666. <https://doi.org/10.1002/sce.21575>
- Jonassen, D.H. (2011). *Learning to Solve Problems: A Handbook for Designing Problem-Solving Learning Environments*. New York: Routledge.
- Kline, P. (1986). *A Handbook of Test Construction: Introduction to Psychometric Design*, Methuen, London.
- Loogma, K., Erss, M., Ümarik, M., & Aasa, M. (2020). Õpetaja professionalismi võimalikud tulevikustsenaariumid aastaks 2035. *Eesti Haridusteaduste Ajakiri*, 8(1), 180–212. <https://doi.org/10.12697/eha.2020.8.1.08>
- OECD (2018). *The future of education and skills: Education 2030*. Paris: OECD Publishing.
- OECD (2019). TALIS 2018 Results (Volume I): Teachers and school leaders as lifelong learners, TALIS, OECD Publishing. <https://doi.org/10.1787/1d0bc92a-en>
- OECD (2020). *PISA 2024 strategic vision and direction for science (Issue March)*. <https://www.oecd.org/pisa/publications/PISA-2024-Science-Strategic-Vision-Proposal.pdf>
- OECD (2021). *The assessment frameworks for cycle 2 of the programme for the international assessment of adult competencies*, OECD Skills Studies, OECD Publishing, Paris.
- Osborne, J., Pimentel, D., Alberts, B., Allchin, D., Barzilai, S., Bergstrom, C., Coffey, J., Donovan, B., Kivinen, K., Kozyreva, A., & Wineburg, S. (2022). *Science Education in an Age of Misinformation*. Stanford University, Stanford, CA.
- Pedaste, M., Baucal, A., & Reisenbuk, E. (2021). Towards a science inquiry test in primary education: Development of items and scales. *International Journal of STEM Education*, 8, 1–19. <https://doi.org/10.1186/s40594-021-00278-z>
- Pedaste, M., Brikker, M., Rannikmäe, M., Soobard, R., Mäeots, M., & Reiska, P. (2017). Loodusvaldkonna õpitulemuste hindamine. Raport, Tartu.
- Pereira, T., Amaral, A., & Mendes, I. (2022). A Competency definition based on the knowledge, skills, and human dispositions constructs. In *International Conference on Internet of Everything* (pp. 29–38). Cham: Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-25222-8_3
- Põhikooli riiklik õppekava (2023). <https://www.riigiteataja.ee/akt/108032023005>.
- Põhikooli riiklik õppekava (2023). Lisa 4. https://www.riigiteataja.ee/akt/lisa/1080/3202/3005/18m_pohi_lisa4.pdf#.
- Rannikmäe, M., Vaino, K., Teppo, M., Soobard, R., Rosin, T., & Must, O. (2018). Loodusvaldkonna õpitulemuste hindamine III kooliastmes. Raport. Tartu.

- Rannikmäe, M., Soobard, R., Vaino, K., Rosin, T., Teppo, M., & Must, O. (2019). Loodusvaldkonna õpitulemuste hindamine III–IV kooliastmes. Raport. Tartu.
- Rannikmäe, M., Soobard, R., & Vaino, K. (2020). Loodusvaldkonna õpitulemuste e-hindamine põhikooli kolmandas astmes ja gümnaasiumis. Kontseptsioon. Tartu.
- Rannikmäe, M., Soobard, R., Vaino, K., & Rosin, T. (2021a). Loodusvaldkonna õpitulemuste e-hindamine põhikooli kolmandas astmes ja gümnaasiumis. Kontseptsioon. Tartu.
- Rannikmäe, M., Soobard, R., Vaino, K., Teppo, M., Valdmann, A., & Rosin, T. (2021b). Loodusvaldkonna õpitulemuste hindamine. Raport. Tartu
- Rannikmäe, M., Vaino, K., Soobard, R., Teppo, M., & Reisenbuk, E. (2023). Lühikokkuvõte 2022/2023. õppeaasta loodusainete III kooliastme katselise tasemetöö tulemustest.
- Reeves, T. (2006). Design research from a technology perspective. In J. van den Akker, K. Gravemeijer, S. McKenney, & N. Nieveen (Eds.), *Educational design research* (pp. 52–66). London: Routledge.
- Rosin, T., Vaino, K., Soobard, R., & Rannikmäe, M. (esitatud). Understanding science teachers' beliefs about the teaching and assessment of scientific competences: Explaining the reasons for implementing or not implementing science e-test feedback.
- Saks, K., & Leijen, Ä. (2015). Kognitiivsete ja metakognitiivsete õpistrateegiate toetamine tehnoloogiaga tõhustatud keeleõppes. *Eesti Haridusteaduste Ajakiri*, 3(2), 130–155. <https://doi.org/10.12697/eha.2015.3.2.05>
- Schut, S., Heeneman, S., Bierer, B., Driessen, E., van Tartwijk, J., & van Der Vleuten, C. (2020). Between trust and control: Teachers' assessment conceptualisations within programmatic assessment. *Medical Education*, 54(6), 528–537. <https://doi.org/10.1111/medu.14075>
- Sjöstrom, J., & Talanquer, V. (2014). Humanizing chemistry education: From simple contextualization to multifaceted problematization. *Journal of Chemical Education*, 91(8), 1125–1131. <https://doi.org/10.1021/ed5000718>
- Zimmerman, B. J. (1989). A social cognitive view of self-regulated academic learning. *Journal of Educational Psychology*, 81(3), 329–339. <https://doi.org/10.1037/0022-0663.81.3.329>
- Zhang, L., Huang, Y., Yang, X., Yu, S., & Zhuang, F. (2022). An automatic short-answer grading model for semi-open-ended questions. *Interactive Learning Environments*, 30(1), 177–190. <https://doi.org/10.1080/10494820.2019.1648300>
- Zheng, L. (2015). A systematic literature review of design-based research from 2004 to 2013. *Journal of Computers in Education*, 2, 399–420. <https://doi.org/10.1007/s40692-015-0036-z>
- Thier, M., & Daviss, B. (2002). *The New Science Literacy: Using Language Skills to Help Students Learn Science*. Portsmouth, NH: Heinemann.
- Tinoca, L., Piedade, J., Santos, S., Pedro, A., & Gomes, S. (2022). Design-based research in the educational field: A systematic literature review. *Education Sciences*, 12(6), 410. <https://doi.org/10.3390/educsci12060410>

- Tire, G., Puksand, H., Lepmann, T., Henno, I., Lindemann, K., Täht, K., Lorenz, B., & Silm, E. (2019). PISA 2018 Eesti tulemused: Eesti 15-aastaste õpilaste teadmised ja oskused funktsionaalses lugemises, matemaatikas ja loodusteadustes. <https://www.innove.ee/uuringud/pisa-uuring/pisa-2018/>.
- The Design-Based Research Collective. (2003). Design-based research: An emerging paradigm for educational inquiry. *Educational Researcher*, 32(1), 5–8. <https://doi.org/10.3102/0013189X032001005>
- Toomaneejinda, A. (2017). Zone of proximal development, dynamic assessment and learner empowerment. *LEARN Journal: Language Education and Acquisition Research Network*, 10(1), 176–185.
- Valdmann, A., Holbrook, J., & Rannikmäe, M. (2012). Evaluating the teaching impact of a prior context-based, professional development programme. *Science Education International*, 23(2), 166–185.
- Van den Akker, J., Gravemeijer, K., McKenney, S., & Nieveen, N. (2006). *Educational Design Research*. Routledge.
- Vitello, S., & Williamson, J. (2017). Internal versus external assessment in vocational qualifications: A commentary on the government's reforms in England. *London Review of Education*, 15(3), 536–548. <https://doi.org/10.18546/LRE.15.3.14>
- Vygotsky, L. S., & Cole, M. (1978). *Mind in Society: Development of Higher Psychological Processes*. Harvard University Press.
- Wainer, H., Sheehan, K. M., & Wang, X. (2000). Some paths towards making Praxis scores more useful. *Journal of Educational Measurement*, 37, 113–140. <https://doi.org/10.1111/j.1745-3984.2000.tb01079.x>
- Wood, D., Bruner, J., & Ross, G. (1976). The role of tutoring in problem-solving. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, 17, 89–100. <https://doi.org/10.1111/j.1469-7610.1976.tb00381.x>

The development of a national e-test on science competence for the third school level: An assessment to support learning

Katrin Vaino^{a1}, Triin Rosin^a, Ülle Liiber^{a,b}, Regina Soobard^a,
Moonika Teppo^a, Ana Valdmann^a, Elle Reisenbuk^c, & Miia Rannikmäe^a

^a Centre for Science Education,

Institute of Ecology and Earth Sciences, University of Tartu

^b Geography Department, Institute of Ecology and Earth Sciences,
University of Tartu

^c Estonian Education and Youth Board

Summary

Theoretical framework

The contemporary concept of science competence includes scientific knowledge and its application, inquiry skills, creativity, the ability to solve problems and make well-considered decisions (OECD, 2020), sustainable values, collaboration and communication skills, and the ability to self-regulate one's learning (Choi et al., 2011).

In Estonia, since 2006, the PISA survey has been conducted among 15–16-year-old students, including an assessment of their science competence. However, its results do not provide feedback on the level of competence of an individual or group of students, nor specifically on the level of science competence defined in the Estonian national curriculum for basic schools (2023).

The given circumstances necessitated the need for a nationwide test based on the Estonian national curriculum, which would, on the one hand, provide objective and comparable feedback about the level of Estonian students' scientific competence to assess the effectiveness of the educational system and, on the other hand, provide specific feedback to the student and the teacher (Ministry of Education and Research, 2014).

Education internationally has moved towards competence-based curricula, which in turn requires a changed approach to assessment, which focuses on supporting teaching and learning and developing a self-directed learner (Assessment Reform Group, 1999). This trend has also increased interest in

¹ Institute of Ecology and Earth Sciences, University of Tartu, Vanemuise 46, 51003, Tartu; katrin.vaino@ut.ee.

the potential and use of diagnostic (low stake) tests. The new generation of diagnostic tests can also offer constructive feedback that defines the student's existing knowledge, skills, attitudes, and areas for improvement (Csapó & Molnár, 2019). Therefore, this type of test is often based on the sociocultural theory (*Ibid.*), according to which learning is possible if it occurs in the student's zone of proximal development (Vygotsky & Cole, 1978; Daniels, 2007).

Therefore, this study aims to develop a nationwide e-test that is able to assess the level of science competence achieved by students at the end of the third school level and provide feedback to the student, teacher, parent, and, in a generalised manner, to the educational community. This paper aims to describe the design process, explain the various design decisions made throughout (e.g. how the users – students interact with the created design), and, in more detail, the resulting final design (e-test).

The conceptual framework of the science e-test is based on the report of Pedaste et al. (2017), including the contemporary definitions of science competence as introduced above. The report presents a potential assessment model and the initial procedure for preparing and executing the task. Accordingly, the test should assess students' science knowledge and inquiry, communication, and problem-solving skills via ten characteristics aligning with the science curriculum as part of the Estonian national curriculum for basic schools (2023). According to the report (Pedaste et al., 2017), the test should consist of context-based tasks, the uses of which should help, amongst other benefits, to maintain the student's interest, as the absence of their interest poses a serious threat to the validity of the conclusions drawn from the test results (Finn, 2015).

Methodology

The test was developed in cooperation with the University of Tartu and the Education and Youth Board (Republic of Estonia). In this study, the design-based research approach was used, which tries to bridge the gap between educational research and actual school practice to increase the viability of the developed design (van der Akker et al., 2006) and to improve the existing educational practice through systematic, flexible and repeated analysis, further development, implementation, and correction (Cobb et al., 2003). As a result of the study, diverse data were collected to assess the appropriateness of the test design. Due to the complexity of the process, only a part of the results is presented in this paper. The characteristics of the final design are described in more detail.

The test development process in each of the five design cycles consisted of four steps (adapted from Reeves, 2006):

- (1) Problem analysis by both researchers and practitioners, goal setting of activities.
- (2) Development of the test tasks and the format of feedback.
- (3) Test evaluation and piloting (see Table 1).

Table 1. The number of test tasks and the number of participating students according to the language of performance and the type of school

Year	Grade	Number of tasks	Number of participating students according to the language of performance and the type of school
2018	9.	51	1226 (Estonian); 274 (Russian); all general education
2019	9.	37	539 (Estonian); 269 (Russian); all general education
2020	9.	35	108 (Estonian); general education
2021	10.	35	1363 (Estonian); 453 (Russian); general education 1323; vocational education 493
2022	10.	35	532 (Estonian), 260 (Russian), general education 713, vocational education 79

(4) Analysing, reflecting, and documenting the results. The results of the prototype test were analysed using (i) descriptive statistics from 2019, (ii) exploratory and confirmatory factor analysis, and (iii) Rasch analysis based on item response theory (IRT) illustrated with Wright maps. Correspondence of individual items to specific levels (base, middle, high, or top) was derived from Wright's maps. Based on the nature of the items at each level and the respective scores of the items, the learning outcomes per level were described.

Results and discussion

As in the case of a low-stakes test, the validity of conclusions may be jeopardised by students' low interest in solving the tasks (Finn, 2015). This aspect was particularly examined in the study. It has been found that, over the years, two-thirds of respondents have perceived the test as interesting. One potential way to increase the response of uninterested students would be to enrich the tasks with media-rich stimuli (Boyle & Hutchison, 2009) in future tests.

Additionally, over the years, students have perceived the test as rather challenging, which can be explained by the fact that solving tasks requires transferring acquired knowledge into everyday life contexts. Another explanation could be that the assessment of inquiry skills has not yet become a self-evident part of science learning, as indicated earlier by Henno et al. (2017).

On the other hand, the number of students perceiving the test as difficult has decreased over time. This may be attributed to both the fact that the test itself has started to impact teaching and assessment methods in schools (Rosin et al., forthcoming) and because the number of tasks in the test has decreased, reducing the potential cognitive overload for students.

While reducing the test difficulty was not a direct goal for the team to ensure the test's discriminative power, monitoring the perceived difficulty of the tests by students in future administrations is advisable.

Based on exploratory factor analysis of the test results, a multifactorial solution was hoped to be found to reduce the number of variables to explain and interpret the results. In reality, based on exploratory and confirmatory factor analysis of several test results (2019–2022), multifactorial models were obtained, which had relatively good fit indicators but whose factors were essentially uninterpretable. Therefore, it was decided to stick to a one-factor solution. According to Cronbach's alpha coefficient, the internal consistency of the factor items was 0.89 (final test), which could be considered "good." Therefore, this study proposes a one-dimensional solution; the latent trait behind the factor items can be called "science competence."

The final assessment model, which successive versions have guided the development of the whole test, includes nine attributes, which are categorised into four components: subject knowledge, inquiry skills, problem-solving skills, and communication skills (see Harno, 2023b). The model includes a description of the learning outcomes characteristic of both – the updated Bloom's (Anderson & Krathwohl, 2011) and SOLO taxonomy (Biggs & Collis, 2014). The latter was primarily used to describe learning outcomes collected through open-ended responses.

The final version of the test consists of four sub-tests, each representing one context ("story") and a total of 35 items designed to determine the level of students' science competence, described by nine attributes, and for the sake of clarity, grouped into four categories (see Table 2).

Table 2. Assessed components, respective attributes, the number, type and examples of the items

Components and attributes	Open-ended	Short answer	Multiple choice	Total	Examples
1. Science knowledge					
The student:					
Explaining natural phenomena			3	3	Arranges natural processes in the correct order.
Use of scientific concepts, symbols, and units		1	4	5	Selects appropriate units for the given situation.
Understanding of scientific models	2		4	6	Drags concepts to the concept map. Explains the model in their own words.
2. Inquiry skills					
Posing a research question or hypothesis			2	2	Selects a research question based on the given situation; formulates a hypothesis based on the graph.
Designing experiment and evaluating its quality	3		2	5	Arranges the stages of the experiment based on the given situation; explains in their own words why the described experiment falls short in quality.
Analyzing information and drawing conclusions	1	1	2	4	Analyzes data presented in the form of a table or graph and draws conclusions.
3. Problem-solving and decision-making					
Problem-solving and decision-making	3		4	7	Explains the nature of the problem based on the given situation and organizes activities to eliminate the malfunction in the described system; considers a dilemma between two social decision alternatives and justifies the decision, including expressing personal values.
4. Communication skills					
Finding information and assessing its reliability	1		1	2	Searches for relevant information on the internet; evaluates the reliability of the given source, highlighting aspects that support their evaluation.
Understanding of and writing science text	1			1	Summarizes textual and visual information in their own words (50–60 words).

The provided feedback informs the student of their level of science competence (zero, base, medium, high, top) and, according to the level, describes their existing competence through nine assessed attributes. In addition, the feedback provides recommendations for setting the next learning goals – these are posed one step higher than the achieved level. The teacher receives feedback on the level of both individual and groups of students that enables him/her to identify students' learning needs more effectively and, based on that, modify his/her teaching.

Keywords: science competence, diagnostic e-test, feedback, design-based research, context-based