

Tartu Ülikool  
Sotsiaalteaduste valdkond  
Haridusteaduste instituut  
Õpetajahariduse õppekava

Valle Morel ja Agnes Vask

ÜKSIKVASTUSTE TEOORIA RAKENDAMINE JÕU MÕISTELISE  
TESTI ANALÜÜSIMISEL

Magistritöö

Vastutav juhendaja: vanemteadur Kaido Reivelt  
Juhendaja: lektor Ingrid Koni

Tartu 2023

## **Üksikvastuste teooria rakendamine jõu mõistelise testi analüüsimisel**

Jõu mõisteline test on USA-s välja töötatud valikvastustega test, mida kasutatakse sealsete ülikooli tudengite arusaamade kaardistamiseks enne mehaanika kursuse läbimist ja pärast mehaanika kursuse läbimist. Jõu mõistelise testi eel- ja järeltesti tulemuste analüüsimisega saab uurida, milline on kursuse kvaliteet ehk kuidas õpilaste arusaam jõust (Newtoni mehaanikast) kursuse jooksul muutub.

Üksikvastuste teooria on meetod valikvastusteliste testide arendamiseks, kuid seda on ka kasutatud analüüsimaks kursuse jooksul toimunud õpilaste teadmiste juurdekasvu. Teste ja teadmiste juurdekasvu on analüüsitud ka klassikalise testiteooriaga.

Käesoleva magistritööga leiti, milline üksikvastuste teooria mudel sobib kõige paremini Eesti gümnaasiumiõpilaste sooritatud jõu mõistelise testi analüüsimiseks, kas jõu mõistelist testi üldse saab kasutada Eesti gümnaasiumiõpilaste peal ja milliseid teadmiste juurdekasvu näitavad klassikalise testiteooria ja üksikvastuste teooria meetodid.

CERCS: S270 Pedagoogika ja didaktika

Märksõnad: jõu mõisteline test, üksikvastuste teooria, klassikaline testiteooria

## **Applying Item Response Theory for analysing Force Concept Inventory**

Force Concept Inventory is a multiple-choice test developed in the United States, used to assess the understanding of university students before and after taking a mechanics course. By analyzing the results of the pre-test and post-test of force concept inventory, it is possible to investigate the quality of the course, that is, how students' understanding of force (Newtonian mechanics) changes throughout the course.

Item response theory is a method used for developing multiple-choice tests, but it has also been used to analyze the growth of students' knowledge during a course. Tests and knowledge growth have also been analyzed using classical test theory.

This master's thesis aimed to determine which item response theory model is best suited for analyzing the results of Force Concept Inventory taken by Estonian high school students, whether Force Concept Inventory can be used with Estonian high school students, and which knowledge growth methods are indicated by classical test theory and item response theory.

CERCS: S270 Pedagogy and didactics

Keywords: Force Concept Inventory, Item Response Theory, Classical Test Theory

# Sisukord

<b>Kokkuvõte</b>	<b>2</b>
<b>Sissejuhatus</b>	<b>5</b>
<b>1 Teoreetiline ülevaade</b>	<b>6</b>
1.1 Klassikaline testiteooria ja üksikvastuste teooria valikvastusteliste testide arendamisel . . . . .	6
1.1.1 Klassikaline testiteooria . . . . .	6
1.1.2 Üksikvastuste teooria - IRT . . . . .	7
1.1.3 IRT mudelid . . . . .	8
1.2 Mõistetestidest ( <i>concept inventory</i> ) ja mõistetestide koostamise meetodikatest .	12
1.3 FCI . . . . .	15
<b>2 Metoodika</b>	<b>18</b>
2.1 Valim . . . . .	18
2.2 Andmekogumine . . . . .	18
2.3 Andmeanalüüs . . . . .	19
<b>3 Tulemused</b>	<b>20</b>
<b>4 Arutelu</b>	<b>26</b>
<b>Tänu sõnad</b>	<b>29</b>
<b>Autorsuse kinnitus</b>	<b>30</b>
<b>Kirjandus</b>	<b>31</b>
<b>Lisa 1. 1PL mudel eeltest a=3.4 küsimused 1–7</b>	<b>34</b>
<b>Lisa 2. 1PL mudel eeltest a=3.4 küsimused 8–14</b>	<b>35</b>
<b>Lisa 3. 1PL mudel eeltest a=6 küsimused 1–7</b>	<b>36</b>
<b>Lisa 4. 1PL mudel eeltest a=6 küsimused 8–14</b>	<b>37</b>
<b>Lisa 5. 1PL mudel eeltest a=9 küsimused 1–7</b>	<b>38</b>
<b>Lisa 6. 1PL mudel eeltest a=9 küsimused 8–14</b>	<b>39</b>
<b>Lisa 7. 2PL mudel eeltest küsimused 1–7</b>	<b>40</b>

<b>Lisa 8. 2PL mudel eeltest küsimused 8–14</b>	<b>41</b>
<b>Lisa 9. 3PL mudel eeltest küsimused 1–7</b>	<b>42</b>
<b>Lisa 10. 3PL mudel eeltest küsimused 8–14</b>	<b>43</b>
<b>Lisa 11. 1PL mudel järeltest <math>a=4.7</math> küsimused 1–7</b>	<b>44</b>
<b>Lisa 12. 1PL mudel järeltest <math>a=4.7</math> küsimused 8–14</b>	<b>45</b>
<b>Lisa 13. 1PL mudel järeltest <math>a=6.4</math> küsimused 1–7</b>	<b>46</b>
<b>Lisa 14. 1PL mudel järeltest <math>a=6.4</math> küsimused 8–14</b>	<b>47</b>
<b>Lisa 15. 1PL mudel järeltest <math>a=7.6</math> küsimused 1–7</b>	<b>48</b>
<b>Lisa 16. 1PL mudel järeltest <math>a=7.6</math> küsimused 8–14</b>	<b>49</b>
<b>Lisa 17. 2PL mudel järeltest küsimused 1–7</b>	<b>50</b>
<b>Lisa 18. 2PL mudel järeltest küsimused 8–14</b>	<b>51</b>
<b>Lisa 19. 3PL mudel järeltest küsimused 1–7</b>	<b>52</b>
<b>Lisa 20. 3PL mudel järeltest küsimused 8–14</b>	<b>53</b>

## Sissejuhatus

Igasuguste õpilaste teadmisi kontrollivate testide puhul on oluline, et test oleks kvaliteetne ja mõttekas. Kvaliteetne ja mõttekas test suudab testile vastajaid eristada nende võimete järgi, küsimused on õpilasele paraja keerukusega ja õpilane võiks testile vastata siiralt, mitte vastuseid huupi pakkudes (Gipps, 1994).

Teadmisi kontrollivaid teste on mitmeid: valikvastustega testid, vabavastustega testid, vastavastustestid, lünktekstid jne (Õpiedukustestid, sa). Valikvastustega teste hinnatakse, kuna neid on lihtne teha arvutis sest testi tulemused tulevad vähese tööjõukuluga ja kiirelt. Aina enam soovitakse teadmiste kontrolle muuta arvutipõhiseks (Rats, 2023).

Valikvastustega testide puhul on mitmeid matemaatilisi meetodeid analüüsimaks testi ennast. Mõned nendest on klassikaline testiteooria ehk CTT (*classical test theory*) ja üksikvastuste teooria ehk IRT (*item response theory*). Mõlemal meetodil on omad tugevused ja nõrkused (Wallace and Bailey, 2010), mida käesolevas magistritöös analüüsitakse.

Füüsika on keskkooliõpilaste arvates raske aine (Shirazi, 2017) ja füüsika põhimõistetest arusaamise kontrollimiseks on USA teadlased koostanud mitmeid valikvastustega mõisteteste (*concept inventory*) (McColgan et al., 2017; Jacobi et al., 2003; Wallace and Bailey, 2010; Hestenes et al., 1992). Esimene ja tuntuim neist on jõu mõisteline test ehk FCI (*Force Concept Inventory*), mis loodi juba aastal 1992 (Hestenes et al., 1992). FCI kontrollib õpilaste arusaamisi Newtoni kolmest seadusest, mida õpetatakse põhjalikumalt näidete, ülesannete ja katsete näol Eesti gümnaasiumides (GRÕK, 2011). FCI arendati USA ülikoolide tudengite arusaamiste kontrollimiseks, seega on vaja teada, kas FCI-d saab kasutada ka Eesti gümnaasiumiõpilaste peal. FCI-d on juba kasutatud Eesti põhikooliõpilaste väärarusaamade kaardistamiseks Newtoni mehaanika mõistmisel (Aarna, 2020), aga FCI sobilikkust Eesti keskkooliõpilaste Newtoni mehaanika teadmiste kontrollimiseks pole analüüsitud.

Käesoleva magistritöoga rakendatakse Eesti gümnaasiumiõpilaste sooritatud FCI testide analüüsimisel CTT ja IRT meetodit.

# 1 Teoreetiline ülevaade

## 1.1 Klassikaline testiteooria ja üksikvastuste teora valikvastusteliste testide arendamisel

Õpetamisel on oluline saada tagasisidet selle kohta, kuidas õpilased teemast aru saavad. Olenevalt eesmärgist, mille kohta soovitakse tagasisidet koguda, sobivad selleks erinevad meetodid. Intervjuud või tagasiside küsimustikud võimaldavad õpilasel oma arvamust avaldada selle kohta, mis meeldis ja mida muuta. Kui soovitakse õpilaste teadmiste hindamist kasutada tagasiside saamiseks, sobivad selleks testid. Testidega saab tagasisidet nii õpetamise efektiivsuse, kui ka üksiku õpilase teadmiste kohta. Õpilaste teadmiste kontrollimisel on vajalik, et meetod, millega kontrolli teostatakse, oleks kvaliteetne. Õpilaste tulemused sõltuvad õpilaste oskustest aga ka testi/kontrolltöö kvaliteedist. Õpilaste tulemusi mõjutab, kas küsimused on arusaadavad, kas küsimused olid liiga lihtsad või liiga rasked jne. Näiteks, kui testi küsimusele õigesti vastanute protsent on madal, siis võib see väljendada õpilaste madalat teadmiste taset või hoopis seda, et küsimus oli raskesti sõnastatud, küsimusest oli võimalik mitmeti aru saada või oli küsimus selle grupi õpilaste jaoks liiga raske. (Krull, 2018)

Testi kvaliteeti saab kontrollida näiteks klassikalise testiteooriaga (CTT - *classical test theory*) või üksikvastuste teooriaga (IRT - *item response theory*). Mõlemad kontrollmehhanismid on olulised testide tulemuste mõtestamisel ja testide küsimuste arendamisel. (Wallace and Bailey, 2010)

### 1.1.1 Klassikaline testiteooria

Klassikaline testiteooria (Classical Test Theory, CTT) on traditsiooniline analüüsimeetod testidele ja küsitlustele. CTT väidab, et testi tegija tulemus ei näita tema tõelist tulemust vaid testile vastaja vaadeldav tulemus ( $X_p$ ) on tema tõelise tulemuse ( $T_p$ ) ja juhusliku vea ( $E_p$ ) summa (Wallace and Bailey, 2010):

$$X_p = T_p + E_p$$

CTT-l on mitmeid statistilisi parameetreid, mis aitavad hinnata ja iseloomustada testile vastaja tõelist tulemust ja vastamisel tehtud juhuslikku viga.

CTT annab juhised, kuidas hinnata testi reliaablust, mis näitab, mil määral indiviidi tõelist tulemust peegeldab tema testitulemus ehk annab hinnangu, kuidas juhuslik viga mõjutab testi tulemust. Testi reliaabluse hindamisel kasutatakse tihti Cronbachi alfat (Wallace and Bailey, 2010), mis on aga probleemne, kuna Cronbachi alfa ei ole sama, mis testi reliaablus, vaid väljendab täpsemalt reliaabluse madalaimat võimalikku väärtust (Sijtsma, 2009). Cronbachi alfat saab leida valemiga  $\alpha = \frac{K\bar{c}}{\bar{v} + (K-1)\bar{c}}$ , kus  $\bar{c}$  on üksiktunnuste vahelise kovariatsioonikordajate keskmine;  $\bar{v}$  on üksitunnuste keskmine dispersioon ja  $K$  on tunnuste

arv (Tooding, 2020).

CTT kasutab oma analüüsis p-väärtusi, mis on küsimusele õigesti vastanute arvu ja kogu vastajate arvu suhe. p-väärtusi kasutatakse CTT-s, et hinnata testi raskust. CTT mõõdab ka testile vastaja testitulemuste ja üksikule küsimusele vastamise korrelatsiooni, et hinnata küsimuste eristatavust. CTT puuduseks on, et CTT sõltub suuresti valimist ja selle homogeensusest (valimis on ühesuguse taustaga vastajad, nt 10. klassi õpilased). Valimi homogeensus mõjutab Cronbachi alfa tulemust ja ka p-väärtusi ja korrelatsiooni testi üksikküsimuse ja testiskoori vahel. Cronbachi alfa kasutamine testi reliaabluse ja sisemise struktuuri järjepidevuse (kas testi küsimused hindavad üht ja sama asja) hindamisel on kahtluse alla seatud (Sijtsma, 2009). Lisaks ei saa CTT tulemuste abil ennustada uue grupi jaoks testi eristatavust ja reliaablust (Wallace and Bailey, 2010).

CTT abil on võimalik hinnata õpilaste teadmiste juurdekasvu. Teadmiste juurdekasvu ühikuks on kasutatud logiteid. Kui lasta õpilastel suuritada enne õpetatavat kursust eeltest ja pärast õpetatavat kursust järeltest, saab leida eel- ja järeltesti logitite vahe (valem (1)) (Wallace and Bailey, 2010):

$$\ln\left(\frac{\text{järeltest}(\%)}{100\% - \text{järeltest}(\%)}\right) - \ln\left(\frac{\text{eeltest}(\%)}{100\% - \text{eeltest}(\%)}\right) \quad (1)$$

Vaatamata piirangutele kasutatakse CTT-d ikka veel laialdaselt haridus- ja psühholoogilises uurimistöös ning see pakub raamistikku hindamistulemuste mõistmiseks ja analüüsimiseks (Sijtsma, 2009).

### 1.1.2 Üksikvastuste teooria - IRT

Item response theory (IRT) ehk üksikvastuste teooria on meetod, mis aitab hinnata õpilase oskusi või iseloomujooni, mis on mõõdetud testide või küsitluste abil. See aitab mõista, kui hästi testi üksikud küsimused mõõdavad huvipakkuvat konstruktsiooni ning kuidas erinevad inimesed neid küsimusi tajuvad. IRT mudelid ennustavad küsimustele õigesti vastamise tõenäosust, võttes arvesse õpilase teadmiste taset ja testi küsimuse omadusi (Wang and Bao, 2010; Wallace and Bailey, 2010).

IRT abil on võimalik kaardistada, millise tõenäosusega vastab mingi tasemega õpilane testi konkreetsele küsimusele õigesti. Õpilased vastavad valikvastustega küsimustele ja nad jaotatakse testi tulemuste järgi tasemegruppideks. Näiteks maksimaalse 100-punktise testi puhul võiks moodustada 10 gruppi: 0-10 punkti saanud, 11-20 punkti saanud, 21-30 punkti saanud õpilased jne. Seejärel arvutatakse iga grupi puhul, millise tõenäosusega vastas näiteks küsimusele nr 7 õigesti näiteks 21-30 punkti saanud keskmine grupi liige (Wang and Bao, 2010; Planinic et al., 2010).

IRT üheks peamiseks eeliseks on see, et see võimaldab hinnata õpilase oskustaset isegi juhul, kui ta pole testi täielikult läbinud. Selle põhjuseks on, et IRT mudelid seostavad õpilase oskustaseme ja õige vastuse tõenäosuse iga testipunkti kohta. Seetõttu kasutatakse IRT-d tihti

hariduses ja psühholoogilises testimises, samuti turu-uuringutes ja küsitluste meetodoloogias (Wang and Bao, 2010) ja näiteks Tartu Ülikooli akadeemilise testi adaptiivseks muutmisel (Silm, 2023).

### 1.1.3 IRT mudelid

Mõned IRT mudelid on üheparameetiline logistiline mudel, kaheparameetiline logistiline mudel ja kolmeparameetiline logistiline mudel. Igal mudelil on kindlad eeldused ja milline mudel valida, sõltub uurimusküsimusest ja andmete omadustest. Üheparameetiline IRT logistilise mudeli abil on võimalik hinnata vaid küsimuse keerukust, kaheparameetrilise mudeli abil on võimalik hinnata ka küsimuse eristavust (st, kas küsimus suudab õpilased enam-vähem ühtlaselt erinevateks tasemeteks eristada) ja kolmeparameetrilise mudeli abil on võimalik hinnata lisaks küsimuse keerukusele ja eristatavusele, kui palju panevad õpilased küsimusele vastates huupi (Wang and Bao, 2010).

IRT enamlevinumaks üheparameetriliseks mudeliks on Taani matemaatiku Georg Raschi järgi nimetatud mudel (valem (2)) (Rasch, 1980):

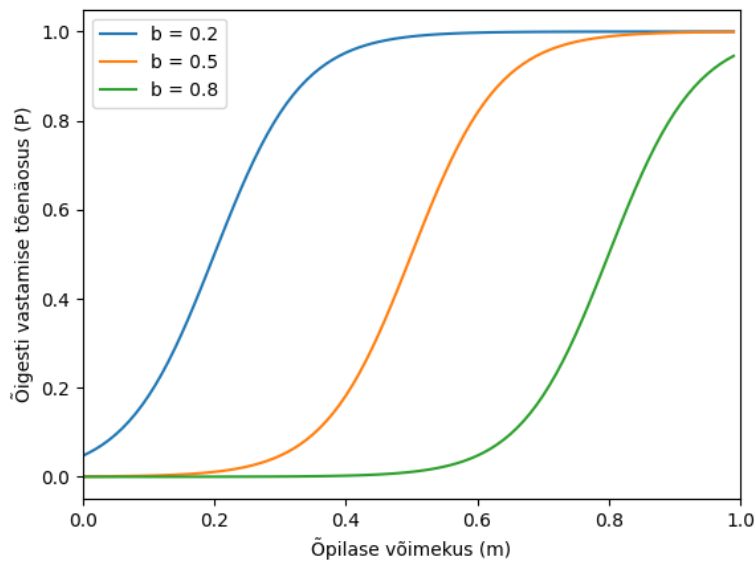
$$P = \frac{e^{m-b}}{1 + e^{m-b}}, \text{ kus} \quad (2)$$

- P - küsimusele õigesti vastamise tõenäosus,
- m - õpilase võimekus,
- b - küsimuse keerukus.

Parameeter b saadakse katsepunktidele parima lähenduse leidmise abil.

Valikvastustega testi vastuseid analüüsid Raschi mudeliga, valmib testi iga küsimuse kohta üks logistilise regressiooni graafik, kus graafiku abstsiss-teljel kujutatakse õpilase võimekust protsentides ja ordinaatteljel tõenäosust, kas vastava võimekusega õpilane vastab analüüsitud küsimusele õigesti (joonis 1). Joonisel 1 on näha, kuidas kõige raskemat küsimust iseloomustav kõver on joonisel kõige paremal pool ja kõige kergemat küsimust iseloomustav küsimus joonise kõige vasakpoolsemal osal. Raschi mudeli graafikud pakuvad kiiret visuaalset hinnangut testi kohta.

Raschi mudel lähendab õpilaste testitulemusi logistilise regressiooni abil. Kasutades IRT edasiarendusi (kaheparameetrilist ja kolmeparameetrilist IRT versiooni) on võimalik veel analüüsida, kui eristav on küsimus ja kui suure tõenäosusega vähima võimekusega õpilane huupi pakkudes vastab küsimusele õigesti (Wang and Bao, 2010).



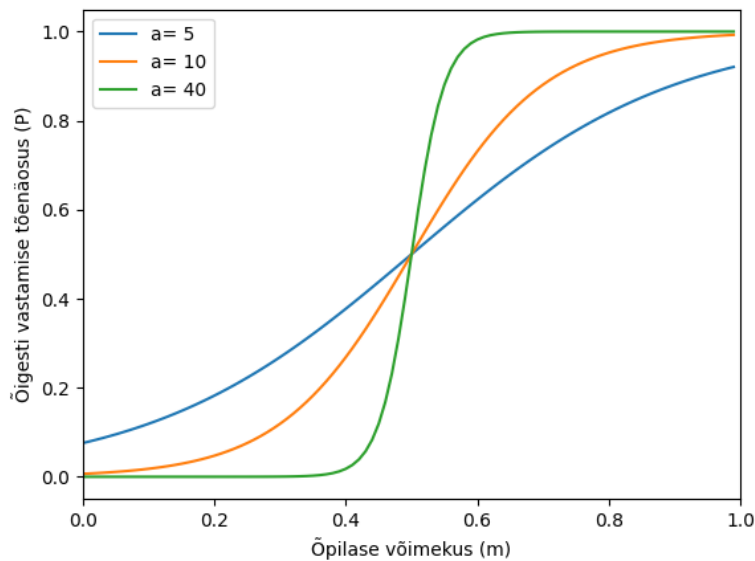
Joonis 1: Teoreetiline üheparameetriline Raschi mudel kolmele erineva raskusastmega küsimusele.

Kui mudelis muudetakse küsimuse eristatavust, saadakse kaheparameetriline Raschi mudel (Wallace and Bailey, 2010):

$$P = \frac{e^{a(m-b)}}{1 + e^{a(m-b)}}, \text{ kus} \quad (3)$$

- P - küsimusele õigesti vastamise tõenäosus (protsentides);
- m - õpilase võimekus (protsentides);
- a - küsimuse eristatavus;
- b - küsimuse keerukus.

Logistilise regressiooni graafikul paistab see välja graafiku tõusuna (joonis 2). Kui tõus on suur (läheneb lõpmatusse), siis on tegu hästi eristava küsimusega, kus võimekamad testitegijad vastavad selgelt küsimusele õigesti ja vähemvõimekas küsimusele valesti. Kui tõus on olematu (läheneb nullile), siis pole küsimus eristav. See tähendab, et vähem võimekad ja rohkem võimekad vastavad küsimusele samasuguse õigestivastamise tõenäosusega. Üheparameetrilise Raschi mudeli puhul (valem (2)) võetakse küsimuse eristatavus konstantseks.



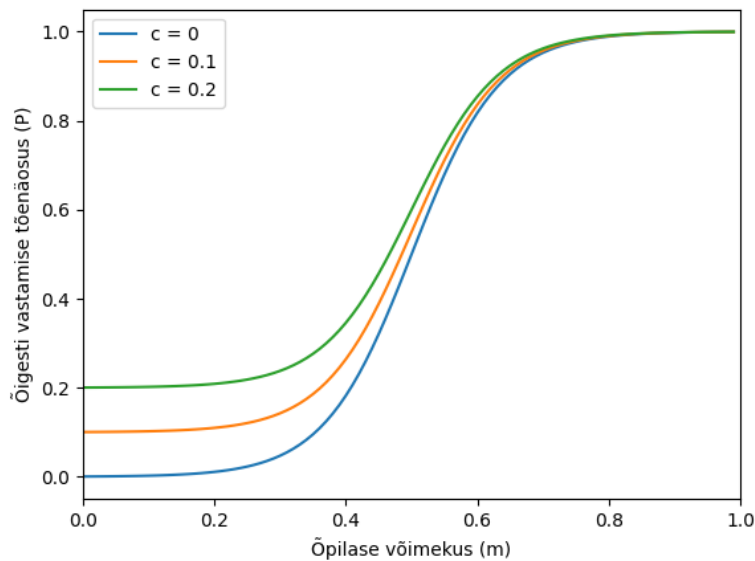
Joonis 2: Teoreetiline kaheparameetriline Raschi mudel kolmele erineva eristatavusega küsimusele.

Kui mudelis muudetakse peale küsimuse eristatavuse ja küsimuse raskuse ka “huupi” vastamise osakaalu, saadakse kolmeparameetriline Raschi mudel (Wallace and Bailey, 2010):

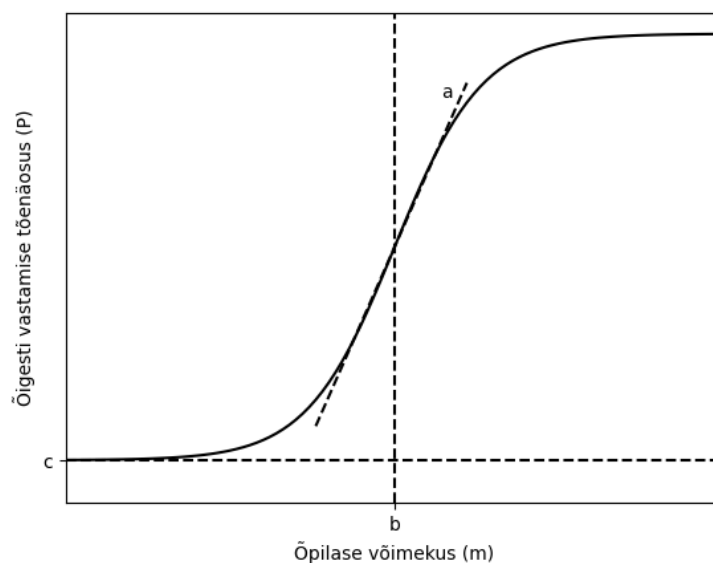
$$P = c + (1 - c) \frac{e^{a(m-b)}}{1 + e^{a(m-b)}}, \text{ kus} \quad (4)$$

- P - küsimusele õigesti vastamise tõenäosus (protsentides);
- m - õpilase võimekus (protsentides);
- a - küsimuse eristatavus;
- b - küsimuse keerukus;
- c - huupi pakkumise parameeter.

Nimelt, mõnikord on küsimus nii raske või esitatud nii segaselt, et õpilane ei saa küsimuselt üldse aru ja vastab valikvastuselisele küsimusele huupi. Sellisel juhul pole ka kõige nõrgemate õpilaste küsimusele õigesti vastamise tõenäosus enam null ja logistilise regressiooni graafikul on näha, et graafiku tunnusjoon on y-telge pidi tõstetud (joonis 3). Kui küsimusel on viis vastusevarianti, siis nõrgemad õpilaste grupi keskmine on huupi pannes vähemalt 20%.



Joonis 3: Teoreetiline kolmeparameetriline Raschi mudel kolmele erineva "huupi" pakumise osakaaluga küsimusele.



Joonis 4: Parameetrite mõju õigesti vastamise tõenäosuse ja õpilase võimekuse kõverale: a - küsimuse eristatavus; b - küsimuse keerukus; c - huupi pakumise parameeter

Kokkuvõttes: üheparameetrilise mudeli kõveral on huupi pakumise parameeter (c) null, mis tähendab, et kõver algab punktist õpilase võimekusega (m) null ja õigesti vastamise tõenäosusega (P) null (joonis 4). Küsimuse eristatavus (a) on kõikide küsimuste puhul sama ehk kõvera tõus jääb muutumatuks. Lähenduses muudetakse küsimuse keerukust (b), mis muudab kõvera keskpunkti asukohta õpilase võimekuse teljes. Kaheparameetrilise mudeli puhul jääb huupi pakumise parameeter nulliks, lähenduses muudetakse nii küsimuse

eristatavust kui ka küsimuse keerukust ehk muutub nii keskkoha asukoht (b) kui ka kõvera tõus (a). Kolmeparameetrilise mudeli puhul muudetakse lähenduses kõiki kolme parameetrit, mille tulemusena muutub lisaks kõvera tõusule (a) ja keskkoha asukohale (b) ka kõvera alguspunkti asukoht õigesti vastamise tõenäosuse teljes (c).

Samamoodi nagu CTT-ga, on IRT abil võimalik hinnata õpilaste teadmiste juurdekasvu. Teadmiste juurdekasvu ühikuks kasutatakse logiteid. Kui lasta õpilastel sooritada enne õpetatavat kursust eeltest ja pärast õpetatavat kursust järeltest, saab leida eel- ja järeltesti logitite vahe. Erinevus CTT-ga on, et logiteid arvutatakse mitte kasutades õpilase küsimusele õigesti vastamise protsenti, vaid õpilase küsimusele õigesti vastamise tõenäosust (valem (5)) (Wallace and Bailey, 2010):

$$\ln\left(\frac{\text{järeltest}(\%)}{100\% - \text{järeltest}(\%)}\right) - \ln\left(\frac{\text{eeltest}(\%)}{100\% - \text{eeltest}(\%)}\right) \quad (5)$$

## 1.2 Mõistetestidest (*concept inventory*) ja mõistetestide koostamise meetodikatest

Concept inventory testid ehk mõistest arusaamise testid loodi USA haridusteadlaste poolt 1980-ndate aastate lõpus. Esimene mõistest arusaamise test oli jõu mõisteline test (inglise keeles Force Concept Inventory ehk FCI), mis loodi aastal 1992 ja oli edasiarendus 1985. aastal ilmunud uuringust “The initial knowledge state of college physics students/Mechanics diagnostics” (Hestenes et al., 1992). “The initial knowledge. . .” küsimustik loodi, et kaardistada ülikooli astuvate tudengite väärarusaamasid Newtoni mehaanikast. Küsimustiku vastuste variandid saadi, kui anti tudengitele avatud vastustega küsimused. Levinumatest tudengite väärarusaamistest moodustati küsimustiku vastuste variantideks nn “peibutuspardid” (Halloun and Hestenes, 1985).

FCI eeskujul on loodud mitmeid mõisteteste füüsikas:

- Electricity and Magnetism Conceptual Assessment (EMCA): käsitleb elektrivälju, elektrilist potentsiaali, vooluringe, magnetväljasid (Broder et al., 2015; McColgan et al., 2017)
- Heat and Thermodynamics Concept Inventory (HTCI): käsitleb soojuse ülekannet, soojustasakaalu ja termodünaamika seadusi (Jacobi et al., 2003)
- Thermodynamics Concept Inventory (TCI): käsitleb õpilaste arusaamist termodünaamikast, energiast, entalpiast ja entroopiast. (Midkiff et al., 2001)
- Quantum Mechanics Conceptual Survey (QMCS): käsitleb näiteks duaalsusprintsipi, superpositsiooni, kvantmõõtmisi (McKagan et al., 2008)
- Star Properties Concept Inventory (SPCI): käsitleb astronoomiaga seotud teemasid (Wallace and Bailey, 2010)

Mõistete teste on arendatud ka näiteks keemias, bioloogias ja matemaatikas:

- Chemistry Concept Inventory (CCI): käsitleb aatomi struktuuri, keemilisi sidemeid, stöhhiomeetriaat ja termodünaamikat (Mulford and Robinson, 2002)
- Enzyme-Substrate Interactions Concept Inventory (ESICI): käsitleb biokeemilisi reaktsioone, uurib ensüümide-substraatide interaktsioone (Bretz and Linenberger, 2012)
- Biology Concept Inventory (BCI): käsitleb bioloogilisi protsesse, bioloogilisi struktuure ja juhuslikkust (Garvin-Doxas and Klymkowsky, 2007)
- Genetics Concept Assessment (GCA): käsitleb geneetikat, genoome, mutatsioone ja geenihäigusi (Smith et al., 2008)
- Basic Skills Diagnostic Test (BSDT): käsitleb suurusjärkude hindamist, baasalgebrat, aritmeetikat ja mõõtmisi (Epstein, 1997)
- Calculus Concept Inventory (CCI): käsitleb funktsioone, tuletisi, piirväärtusi, arvsuhteid ja funktsiooni pidevust (Epstein, 2013)

Mõisteteste on koostatud seega palju. Humphreys (2015) kaardistas mõisteteste loodusteadustes ja leidis 51 mõistetesti: füüsikas (11 testi), astronoomias (neli testi), bioloogias (19 testi), biokeemias (kolm testi), keemias (13 testi) ja materjaliteaduses (üks test). Humphreys kaardista USA ülikoolides loodud mõisteteste. (Humphreys, 2015).

Mõistetestid on populaarsed, kuna nad on lühikesed, valikvastustega ja kaardistavad lihtsalt ja kiirelt õpilaste väärarusaamasid. Mõistetestide populaarsus kasvas, kui Hake'i (1998) suure FCI metauuringuga tõestati, et traditsioonilise loenguformaadiga paranesid USA ülikoolide tudengite teadmised Newtoni mehaanikast märkimisväärselt vähe. Seevastu interaktiivsete õpetamismeetoditega kursuste käigus paranesid tudengite teadmised palju (Hake, 1998). Sellest alates on mitmed ülikoolid rakendanud FCI-d oma kursuste tudengite arusaamiste kasvu kaardistamiseks. Ka teised erialad on teinud oma kursuste põhilisi arusaamisi kontrollivaid mõisteteste (Epstein, 2013).

Head mõistmistestid on loodud nii, et kõigepealt vastavad õpilased avatud küsimustele, mille abil kaardistavad uurijad õpilaste seas enamlevinud väärarusaamad, mida saab kasutada mõistetesti vastusevariantides. Mõistetestide arendamisel on kasutatud nii CTT-d kui IRT-d. Mõistetestid on tihti jaotatud eel- ja järeltestideks, et kaardistada kursuse kvaliteeti, kuidas õpilased on kursuse jooksul teadmisi omandanud ja väärarusaamade asemel omandanud õige, teaduslikult kinnitatud maailmapildi (Epstein, 2013).

Viimasel viieteistkümnel aastal on tekkinud mõistmisteste, mis kattuvad testitava sisu poolest, pole jaotatud eel- ja järeltestiks, mille arenduses pole kasutatud põhjalikke psühhomeetrilisi analüüsimeetodeid (näiteks CTT-d või IRT-d) ja mis on nii uued, et nendele testide täpsust ja mõju pole teiste teadusgruppide poolt analüüsitud (Humphreys, 2015).

Mitmeid mõistmistest haldab USA füüsikaõpetajaid ja füüsikaharidusega tegelevaid teadlasi ja õppejõude koondav internetiportaal PhysPort. PhysPort-is on välja toodud hea mõistmistesti standard, mis annab raamistiku, millised teaduslikud uuringud/katsed peavad olema mõistmistesti puhul läbi viidud. Portaalil antakse seitse etappi, mis peab mõistmistestide puhul läbi viidud (Madsen et al., 2022).

1. Õpilaste ideede kogumine antud teema kohta, tavaliselt intervjuude või avatud küsimustega kirjalike küsimustike abil.
2. Õpilaste ideede kasutamine valikvastustega mõistmistestide koostamiseks, kus valede vastuste hulka kuuluvad õpilaste kõige tavalisemad väärarusaamad, kasutades õpilaste tegelikke sõnastusi.
3. Küsimuste testimine teise õpilasarühmaga. Tavaliselt kasutavad teadlased intervjuusid, kus õpilased arutavad igat küsimust.
4. Nende küsimuste testimine erialaekspertidega, et tagada küsimuste olulisuse ja vastuste õigsuse osas üksmeel.
5. Küsimuste ülevaatamine õpilaste ja ekspertide tagasiside põhjal.
6. Testi läbiviimine suure hulga õpilaste seas. Tulemuste korratavuse kontrollimine erinevate kursuste ja institutsioonide vahel. Vastuste jaotuste kontrollimine. Erinevate statistiliste meetodite kasutamine hindamise usaldusväärsuse tagamiseks.
7. Testi parandamine/arendamine saadud tagasiside põhjal (Madsen et al., 2022).

Mõistmistestid saavad PhysPort-is hinnangu, kui akadeemiliselt kvaliteetse testiga tegu on. Test peab vastama vähemalt mõnele järgmistest tunnustest (Madsen et al., 2022):

- Testi küsimused põhinevad teaduslikel uuringutel, mis käsitlevad õpilaste arusaamist
- Test koostatakse õpilasintervjuusid kasutades
- Test on vaadatud üle ekspertide poolt
- Testi analüüsimiseks on kasutatud sobivat statistilist analüüsi
- Testi on kasutatud mitmes asutuses
- Uuringud on läbi viidud mitme uurimisrühma poolt
- Test on saanud vähemalt ühe eelretsenseeritud publikatsiooni

Kui mõistmistest vastab kõigile seitsmele ülaltoodud tunnusele, saab test kõrgeima, 7. kategooria hinnangu (Madsen et al., 2022).

### 1.3 FCI

Põhikooli riikliku õppekava (2011) alusel on III kooliastmes (7.-9. klass) ette nähtud neli füüsika kursust - valgusõpetus, mehaanika, elektriõpetus ja soojusõpetus. Põhikooli riikliku ainekava lisades leiduva füüsika ainekava alusel toimub jõu mõistete õppimine mehaanika kursuse jooksul, mis on tavapäraselt 8. klassis (PRÕK, 2011). Gümnaasiumis on ette nähtud viis füüsika kursust - Sissejuhatus füüsikasse. Kulgliikumise kinemaatika, Mehaanika, Elektromagnetism, Energia, Mikro- ja megamaailma füüsika (GRÕK, 2011). Gümnaasiumi riikliku ainekava järgi õpitakse Newtoni mehaanikat mehaanika kursusel, mis on tavapäraselt 10. klassis. Kiirendusest räägitakse sissejuhatavalt kinemaatika kursusel (GRÕK, 2011).

Jõu mõisteline test on USA haridusteadlaste poolt koostatud 30-küsimuseline valikvastustega test, mis hindab õpilaste arusaamist Newtoni mehaanikast (Hestenes et al., 1992). Testis küsitavad teemad kattuvad Eesti põhikooli ja gümnaasiumi füüsika õppekava kinemaatika ja mehaanika osaga. FCI küsimustele vastamine ei nõua arvutusi, vaid õpilane peab mõistma, kuidas asjad liiguvad ühe ja mitme jõu mõjul. Küsimustel on viis vastusevarianti. Iga õigesti vastatud küsimus annab ühe punkti.

FCI loomise mõtte oli hinnata ülikooli õpilaste arusaamist Newtoni mehaanikast. FCI on PhysPort mõistmistestide kõige kõrgemal, 7. tasemel. Test on koostatud tudengite avatud vastuste põhjal, avatud vastused on kategoriseeritud enamlevinudateks arusaamisteks, mis tihti peale ei vasta Newtoni mehaanikale ja on seeläbi teaduslikult valed. Kui FCI test oli kokku pandud ja see oli erialaekspertidele ülevaatamiseks antud, arvasid erialaekspertid, et vastusevariantidest on õige vastus liiga lihtsasti ära tuntav ja seeläbi on test liiga kerge (Hestenes et al., 1992). Tegelikult ei vasta õpilased testile hästi. Õpilaste seas on väärarusaamad levinud (Hake, 1998).

FCI-d on on antud õpilastele lahendada enne mehaanika kursust ja sama test on antud õpilastele lahendamiseks ka pärast mehaanika kursust, kus nad õpivad Newtoni mehaanika põhitõdesid ja võiksid seeläbi vastata testile õigesti (Hestenes et al., 1992).

Märkimisväärne on see, et Hallouni ja Hestenesi töö käigus välja töötatud test ei näidanud, et pärast kursuse läbimist oleks tudengite väärarusaamad mehaanikast muutunud. Seega, tegelemata väärarusaamadega kursuse jooksul, saavad tudengid mõistetest valesti aru ja mõistavad loengus antavat infot valesti (Halloun and Hestenes, 1985). Hake (1998) metauurimus näitas, et õpilaste tulemuste paranemine kursuse jooksul on tugevas seoses sellega, kui interaktiivne õpe on ja kuidas kursus on üles ehitatud.

Hake (1998) kasutas suures metatestis FCI eel- ja järeltestide abil kursuste kasumlikkuse hindamiseks järgmist valemit, mida ta nimetas normaliseeritud juurdekasvuks (normalised gain)  $g$ :

$$g = \frac{\text{järeltest}(\%) - \text{eeltest}(\%)}{100\% - \text{eeltest}(\%)}$$

Hake (1998) liigitab  $g$  väärtuste järgi kursuse jooksul toimunud teadmiste juurdekasvu suureks, keskmiseks ja madalaks järgmiselt:

- suur juurdekasv, kui  $g \geq 0,7$
- keskmine juurdekasv, kui  $0,7 > g \geq 0,3$
- madal, kui  $g < 0,3$ .

Traditsiooniliste kursuste (loengud ja seminarid) puhul olid õpilaste järeltestide tulemused väga vähe paremad kui õpilaste eeltestide tulemused. Hake (1998) tulemused aitasid kaasa paljude ülikoolide õppemetoodikate muutmisele interaktiivsemaks (Epstein, 2013).

Jaapanis loodi FCI testist arvutipõhine adaptiivne testimine, mis võimaldas 40-õpilasega klassis testi lühendada 50%–63%, seeläbi testi täpsus vähenes 5–10% (Yasuda et al., 2021). Seeläbi kulub õpilastel testi sooritamiseks vähem aega.

2015. aastal poolitati test kaheks 14-küsimuseliseks alamtestiks (tabel 1), mis hindavad mõlemad võrdselt Newtoni mehaanikast arusaamist (Han et al., 2015). Tänu kahele alamtestile saavad õpilased sooritada enne ja pärast kursuse läbimist erinevaid teste.

Mõned küsimused algsest 30-küsimuselise FCI testist (1,3,16,27,29/30) ei ole kasutusel eel- ja järeltesti loomisel. Need küsimused jäeti kõrvale, et hoida eel- ja järeltest piisavalt lühikesena. Lisaks ei olnud küsimused 1 ja 3 piisavalt eristavad, need olid liiga lihtsad. Küsimuse 15 vastus mõjutas liiga palju küsimusele 16 vastamist, nii jäeti küsimus 16 eel- ja järeltestidest kõrvale. Küsimused 2, 25/26, 28 kasutati eel- ja järeltestis, et hoida testid võrdse raskusega (Han et al., 2015).

Tabel 1: Eel- ja järeltesti konstruktsioonid. Arvud eel- ja järeltestis on algse FCI testi küsimuste järjekorranumbrid. (Han et al., 2015)

Konstruktsioon, mida FCI test käsitleb	Eeltestis kasutatud FCI põhitesti küsimuse number	Järeltestis kasutatud FCI põhitesti küsimuse number
Vabalangemine	2	2
Newtoni III seadus	15, 28	4, 28
Liikumine jõuväljas	13, 17, 26	25, 26, 30
Ringliikumine	5, 6	7, 18
Liikumine gravitatsiooniväljas	12	14
Kinemaatika	19	20
Liikumine mitme jõu mõjul	8–11	21–24

FCI on juba eesti keelde tõlgitud, digitaliseeritud ja viidud veebikeskonda opik.fyysika.ee (Aarna, 2020). Testi küsimused ei ole avalikud, et vältida õpilasi enne testi sooritamist sellega tutvumast. Eesti keelde tõlgitud FCI eeltesti ja järeltesti küsimuste valik FCI põhitestist (Hestenes et al, 1992) erineb natuke Han et al. (2015) jaotusest (tabel 2). Eestikeelses FCI eel- ja järeltesti ja inglisekeelseos eel- ja järeltestis on konstruktsioonid, mida küsitakse, siiski samad.

FCI testi on läbi viidud lisaks USA ka Horvaatias ja seal tehtud testi analüüsi üheparameetrilise IRT mudelit (Planinic et al., 2010). Eestis on FCI testi läbi viidud

Tabel 2: Eel- ja järeltesti konstruktsioonid. Arvud eel- ja järeltestis on algse FCI testi küsimuste järjekorranumbrid.

Konstruktsioon, mida FCI test käsitleb	Eeltestis kasutatud FCI põhitesti küsimuse number	Järeltestis kasutatud FCI põhitesti küsimuse number
Vabalangemine	2	2
Newtoni III seadus	15, 28	4, 28
Liikumine jõuväljas	13, 17, 25	25, 26, 29
Ringliikumine	5, 6	7, 18
Liikumine gravitatsiooniväljas	12	14
Kinemaatika	19	20
Liikumine mitme jõu mõjul	8–11	21–24

põhikooliõpilaste peal (Aarna, 2020), aga seda kasutati õpilaste väärarusaamade kaardistamiseks. FCI testi pole veel kasutatud Eesti gümnaasiumiõpilaste peal kaardistamaks õpilaste Newtoni mehaanikast arusaamiste kasvu, st läbiviidud kursuste kasulikkust. Arusaamiste kasvu on võimalik analüüsida nii IRT kui CTT meetodiga. IRT puhul on oluline ka, kas kasutada IRT ühe-, kahe- või kolmeparameetrilist mudelit. Kõik IRT mudelid ei pruugi suuta leida katseandmetele lähendust (Wallace and Bailey, 2010).

Seega püstitati töös järgnevad uurimisküsimused:

1. Milline IRT mudel on kõige sobilikum Eesti keskkooliõpilaste FCI eel- ja järeltestide analüüsimiseks?
2. Kuidas sobib USA ülikoolide tudengite jaoks välja töötatud FCI test Eesti gümnaasiumiõpilaste Newtoni mehaanikast arusaamade analüüsimiseks.
3. Kasutades analüüsimiseks IRT ja CTT teadmiste juurdekasvu mõõtmise meetodikaid, siis millise Newtoni mehaanika teadmiste juurdekasvu annab Eesti keskkooliõpilastele kinemaatika ja mehaanika kursuse õpetamine?

## 2 Metoodika

Töös püstitatud uurimisküsimustele vastamiseks viidi õpilaste peal läbi FCI eeltest, seejärel läbisid õpilased kinemaatika ja mehaanika kursused ja viidi õpilaste peal läbi FCI järeltest. Testide tulemusi analüüsiti IRT ühe-, kahe- ja kolmeparameetrilise mudeliga, et vaadata, kas katseandmetele (testide tulemustele) on võimalik IRT mudelite graafikuid lähendada. Eel- ja järeltestide tulemusi analüüsiti Hake'i normaliseeritud juurdekasvu valemiga, CTT juurdekasvuvalemiga ja IRT juurdekasvuvalemiga.

### 2.1 Valim

FCI eel- ja järeltestides osalevad õpilased valiti kombineeritud valimiga (mugavus- ja kriteeriumvalim (Rämmer, 2014)). Valimi kriteeriumiks oli, et testi läbijad peavad olema õpilased, kes õpivad gümnaasiumis ja on läbimas kinemaatika ja mehaanika kursust, seega sobisid 10. klasside õpilased. Mugavusvalimina otsustati viia testid läbi 10. klasside peal, keda autorid õpetasid. Kaasati ka kahe teise eesti õpetaja läbi viidud testide andmed, sest IRT analüüsi tarbeks oleks ainult autorite õpetatud õpilastest väheks jäänud (Plannic et al, 2010). Autorite õpetatud õpilasi oli 103, teiste õpetajate õpilastega lisandus sõltuvalt ee- ja järeltestist 385 ja 325 õpilast.. Õpilased olid kolmest erinevast koolist. Õpilaste koguarv oli eeltestis 488 ja järeltestis 445. Eeltestis jaotus õpilaste hulk vastavalt Kool 1 - 91 õpilast, Kool 2 - 70 õpilast, Kool 3 - 327 õpilast. Järeltestis jaotus õpilaste hulk vastavalt Kool 1 - 61 õpilast, Kool 2 - 69 õpilast ja Kool 3 - 315 õpilast. Õpilased olid kõik 10. klassist ja eeltesti tegemise ajal olid nad just alustanud füüsika õppimisega (kinemaatika kursuse alguses), järeltesti tegemise ajal olid nad just mehaanika kursuse lõpetanud.

### 2.2 Andmekogumine

Töös kasutati lisaks esimese autori läbi viidud testidele ka teiste õpetajate poolt läbi viidud testide tulemusi, kuna IRT meetodil analüüsimiseks on oluline suur vastajate arv (Planinic et al., 2010). Teste viidi läbi aastatel 2021-2023. Koolis 3 viidi teste läbi 2021.-2022. õppeaastal, koolis 1 2020.-2021. ja 2022.-2023. õppeaastal ja koolis 2 viis testid läbi esimene autor õppeaastal 2022.-2023.

Õpilaste vastused olid nimelised, kuna õpilased pidid fyysika.ee testikeskkonda sisse logima oma Google'i kontoga. Kuna mõnel õpilasel ei olnud võimalik oma Google'i kontoga sisse logida, kasutas õpilane testi läbiviiva õpetaja Google' kontot. Pärast testi tegemist said õpilased oma tulemust näha ainult punktidenä. Nad ei saanud vaadata, milline küsimus neil valesti läks. Teiste õpilaste tulemusi nad ei näinud (aga mõned õpilased jagasid omavahel oma testide tulemusi). Õpilased ei saanud testi tegemise eest ühtegi hinnet. Õpilastele öeldi suuliselt, et tegu on poolanonüümse testiga, aga nad võiksid anda endast parima ja küsimusele tõsiselt vastata. Õpilastelt ja lastevanematelt testi läbiviimiseks nõusolekut ei küsitud.

Õpilastele öeldi, et nad teevad seda testi õpetaja magistritöö tarbeks. Nii eel- kui järeltesti 14 küsimusele vastamiseks oli opik.fyysika.ee keskkonnas antud 20 minutit.

Tegu oli poolanonüümse testiga, kuna testi tulemuste analüüsimisel olid õpilaste tulemused analüüsijatele näha nimeliselt, aga nimed eemaldati analüüsi käigus. Testi tegemine ei kahjustanud õpilast füüsiliselt ja testi eest hinde mittepanemine vähendas õpilaste vaimset ärevust, seega võib öelda, et eetikanõuetega oli uurimus mingil määral kooskõlas (Eetikakeskus, sa). Õpilaste ja õpilaste vanemate käest kirjalikku nõusolekut ei küsitud.

## 2.3 Andmeanalüüs

FCI eeltestid ja järeltestid viidi läbi kontrollitud olukorras internetikeskkonnas opik.fyysika.ee ja vastused salvestati MS Exceli faili. Eeltestide ja järeltestide tulemusi analüüsisiti IRT meetodiga kasutades Pluralsighti loodud Pythoni koodi (Mashburn, 2021). Pluralsighti koodi modifitseeriti, nii et programm väljastas nii eeltesti kui järeltesti kõikide küsimuste IRT ühe-, kahe- ja kolmeparameetrilise mudeli lähendused ja parameetrid. Programmi modifitseeriti ka nii, et see väljastaks iga küsimuse kohta vastava logistilise regressiooni graafiku koos katsepunktide ja katsepunktide 95%-se usaldusvahemikuga.

Koodi (Mashburn, 2021) kasutati IRT meetodi üheparameetrilise, kaheparameetrilise ja kolmaparameetrilise funktsioonide kõverate (inglise keeles Item Characteristic Curve ehk ICC) loomiseks, mida lähendati kõrgeima tõenäosuse hinnanguga (inglise keeles Maximum likelihood estimation (MLE)) andmepunktidele. MLE ei suuda korrektselt määrata kõigile küsimustele õigesti vastanute ja mitte ühelegi küsimusele õigesti vastanute võimekust (Wallace and Bailey, 2010).

Iga küsimuse kohta leiti, millise tõenäosusega vastab antud küsimusele alagrupp õigesti. Tõenäosus saadakse kõigi sellele küsimusele õigesti vastanute arvu jagamisel õpilaste koguarvuga. Tõenäosusest sai üks andmepunkt graafikul. Mida rohkem inimesi on alagrupis, seda täpsem on tõenäosus.

Iga küsimuse kohta leiti ka küsimusele õigesti vastamise tõenäosuse usaldusvahemik. Usaldusvahemik näitab, mis piirkonda langevad 95% tulemustest ehk aritmeetiline keskmine  $\pm$  kaks standardhälvet. Kui kõik alagrupis olevad õpilased vastasid konkreetsele küsimusele õigesti/valesti on antud alagrupi usaldusvahemik null. Selle põhjuseks võib olla alagrupi väike valim.

Piisavalt suure valimi puhul on suurema punktisumma saanud õpilaste (suurema võimekusega) tõenäosus konkreetsele küsimusele õigesti vastata suurem kui väiksema punktisumma (võimekusega) õpilaste puhul. Kõvera lähendamisel kasutati kõiki alagruppe. Kui programm suudab kõveraid katsepunktidele lähendada, saame hinnata FCI eel- ja järeltesti küsimuste keerukust ja eristatavust (Wallace and Bailey, 2010).

### 3 Tulemused

Uurimisküsimus 1: Milline IRT mudel on sobilikum Eesti keskkooliõpilaste FCI eel- ja järeltestide analüüsimiseks?

Katsepunktidele lähendatud IRT mudelid nii eel- kui ka järeltestis on välja toodud lisades (Lisa 1 - Lisa 20). Lähendatud mudelite parameetrid a, b ja c on toodud tabelites 3, 4, 5 ja 6.

Küsimuste eristatavus varieerus, seega üheparameetiline mudel lähendas mõned küsimused halvemini kui kaheparameetiline mudel. Kolmeparameetrilise mudeli puhul ei suudetud kõiki küsimusi lähendada ning lähendatud küsimuste puhul oli huupi pakkumise parameeter nullilähedane, millest võib järeldada, et kolmeparameetiline mudel ei suuda katsepunkte täpsemalt lähendada (tabelid 4 ja 6).

IRT mudeli lähendamise kvaliteetsust kirjeldab tabelites välja toodud  $\chi^2$ . Mida väiksem on  $\chi^2$ , seda paremini on mudelit lähendatud katseandmetele (Barbera, 2013). Eeltestis oli kõige madalam keskmine  $\chi^2$ , mille puhul mudelit kõigile küsimustele suudeti lähendada kolmeparameetrilisel mudelil (tabel 4), aga kolmeparameetiline mudel ei suutnud lähendust leida 3., 4., 10. ja 13. küsimuse puhul.

Järeltestis ei suudetud kaheparameetrilist ja kolmeparameetrilist mudelit küsimusele 5 lähendada (tabel 6, lisa 17, lisa 19) ning üheparameetriliste mudelite puhul oli  $\chi^2$  väärtus küsimuse 5 puhul märgatavalt suurem kui teiste küsimuste puhul (tabel 5). Kõige madalam keskmine  $\chi^2$  saadi kolmeparameetrilise mudeliga, aga kolmeparameetrilise mudeliga ei suudetud leida kõigile küsimustele lähendusi (tabel 5).

Eeltestis suudeti kõigi küsimuste puhul katsepunktidele lähendada kolme erineva küsimuse eristatavusega üheparameetrilist mudelit. Kaheparameetrilist mudelit suudeti samuti katsepunktidele lähendada kõigi küsimuste puhul. Kolmeparameetrilise mudeli puhul ei suudetud lähendust leida nelja küsimuse puhul (tabel 4).

Järeltestis suudeti kõigi küsimuste puhul katsepunktidele lähendada kolme erineva küsimuse eristatavusega üheparameetrilist mudelit. Kaheparameetrilist mudelit suudeti katsepunktidele lähendada kolmeteistkümne küsimuse puhul neljateistkümnest. Kolmeparameetrilise mudeli puhul ei suudetud lähendust leida kuue küsimuse puhul (tabel 6).

Vaadates madalaimat  $\chi^2$  ja mudeli küsimuste lähenduste õnnestumise arvu, sobib Eesti gümnaasiumiõpilaste FCI eel- ja järeltesti analüüsimiseks kõige paremini IRT kaheparameetiline mudel.

Tabel 3: Üheparameetriliste eeltestide parameetrid

	$a = 3.4$		$a = 6$		$a = 9$	
	b	$\chi^2$	b	$\chi^2$	b	$\chi^2$
1	0,69	34,43	0,51	13,02	0,44	7,74
2	0,81	24,98	0,59	21,86	0,50	84,11
3	0,82	20,54	0,60	40,93	0,50	101,58
4	0,21	41,80	0,22	28,63	0,22	53,34
5	0,25	48,13	0,24	21,71	0,24	10,33
6	0,35	15,04	0,30	6,32	0,28	22,33
7	0,60	52,63	0,46	25,27	0,40	15,73
8	0,89	18,08	0,64	9,97	0,53	27,55
9	1,11	42,11	0,78	18,86	0,64	15,72
10	0,30	51,04	0,27	45,47	0,26	94,69
11	0,95	20,49	0,68	29,13	0,56	76,11
12	1,19	22,72	0,83	13,76	0,68	19,79
13	0,43	37,54	0,36	29,66	0,32	65,24
14	0,74	25,82	0,54	15,40	0,46	35,13
Keskmine	0,67	32,53	0,5	22,86	0,43	44,96

Märkused. Iga rida tähistab ühte küsimust testis

Tabeli alumisel real on leitud lähenduse igale parameetritele keskmine väärtus

a - küsimuse eristatavus; b - küsimuse keerukus;  $\chi^2$  - lähenduse sobivus

Tabel 4: Kahe- ja kolmeparameetrilise eeltesti parameetrid

	a	b	$\chi^2$	a	b	c	$\chi^2$
	1	9,05	0,44	7,78	8,18	0,43	-0,03
2	5,48	0,62	19,83	3,79	0,61	-0,10	17,57
3	2,24	1,10	19,61	-	-	-	-
4	8,27	0,22	42,10	-	-	-	-
5	10,27	0,24	10,38	6,09	0,02	-1,15	6,61
6	6,19	0,30	6,49	5,29	0,25	-0,14	6,29
7	10,04	0,39	17,25	8,68	0,38	-0,04	14,87
8	6,05	0,63	10,01	5,35	0,64	-0,02	9,64
9	9,54	0,63	17,23	14,26	0,60	0,01	10,36
10	7,38	0,27	58,30	-	-	-	-
11	3,41	0,95	20,47	7,41	0,82	0,06	19,92
12	6,25	0,81	13,70	5,04	0,84	-0,02	13,29
13	6,83	0,34	33,68	-	-	-	-
14	6,78	0,52	16,76	4,47	0,48	-0,14	12,93
Keskmine	6,98	0,53	20,97	6,86	0,51	-0,16	11,91

Märkused. Iga rida tähistab ühte küsimust testis

Tabeli alumisel real on leitud lähenduse igale parameetritele keskmine väärtus

a - küsimuse eristatavus; b - küsimuse keerukus; c - huupi pakkumise parameeter;  $\chi^2$  - lähenduse sobivus

Tabel 5: Üheparameetriliste järeltestide parameetrid

	$a = 4.7$		$a = 6.4$		$a = 7.6$	
	b	$\chi^2$	b	$\chi^2$	b	$\chi^2$
1	0,68	14,58	0,61	15,82	0,58	23,79
2	0,58	15,82	0,53	18,32	0,50	30,68
3	0,27	15,12	0,27	14,35	0,27	20,78
4	0,51	22,27	0,47	15,62	0,45	17,36
5	0,73	45,36	0,66	94,33	0,62	151,50
6	0,51	24,68	0,47	19,88	0,46	25,32
7	0,50	11,88	0,46	14,86	0,44	26,32
8	0,51	18,78	0,47	16,26	0,45	21,61
9	0,49	17,02	0,45	18,10	0,44	30,00
10	0,32	22,26	0,31	27,95	0,31	47,33
11	0,70	13,76	0,63	22,55	0,60	37,51
12	0,87	16,56	0,77	17,72	0,73	26,84
13	0,53	19,67	0,48	9,18	0,46	7,82
14	0,12	20,38	0,15	27,38	0,17	47,89
Keskmine	0,52	19,87	0,48	23,74	0,46	36,77

Märkused. Iga rida tähistab ühte küsimust testis

Tabeli alumisel real on leitud lähenduse igale parameetrile keskmine väärtus

a - küsimuse eristatavus; b - küsimuse keerukus;  $\chi^2$  - lähenduse sobivus

Tabel 6: Kahe- ja kolmeparameetrilise järeltesti parameetrid

	Kaheparameetriline			Kolmeparameetriline			
	a	b	$\chi^2$	a	b	c	$\chi^2$
1	5,70	0,63	13,87	5,24	0,62	-0,02	13,56
2	6,00	0,54	16,27	-	-	-	-
3	6,21	0,27	13,86	-	-	-	-
4	6,88	0,46	15,70	5,56	0,42	-0,11	14,01
5	-	-	-	-	-	-	-
6	6,92	0,47	21,27	5,85	0,44	-0,08	19,14
7	5,61	0,48	11,65	-	-	-	-
8	6,31	0,47	16,10	6,17	0,47	-0,01	16,15
9	6,25	0,46	17,30	5,13	0,41	-0,11	14,39
10	6,42	0,31	28,12	-	-	-	-
11	4,73	0,70	13,80	8,89	0,70	0,09	11,57
12	6,16	0,78	16,78	7,39	0,76	0,02	13,65
13	7,66	0,46	7,87	6,36	0,43	-0,08	5,67
14	6,42	0,15	27,59	-	-	-	-
Keskmine	6,25	0,48	16,94	6,32	0,53	-0,04	13,52

Märkused. Iga rida tähistab ühte küsimust testis

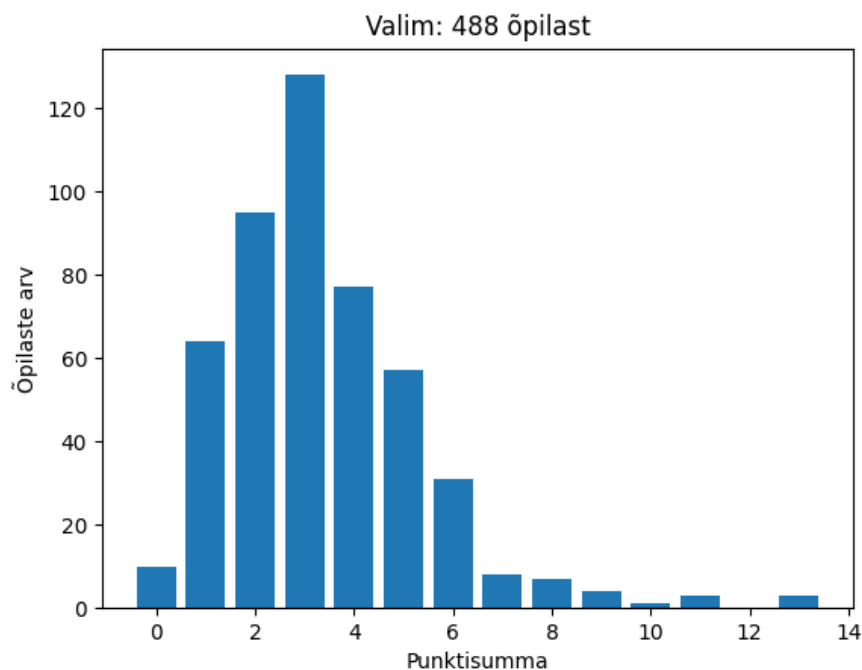
Tabeli alumisel real on leitud lähenduse igale parameetrile keskmine väärtus

a - küsimuse eristatavus; b - küsimuse keerukus; c - huupi pakkumise parameeter;  $\chi^2$  - lähenduse sobivus

Uurimisküsimus 2: “Kuidas sobib USA ülikoolide tudengite jaoks välja töötatud FCI test Eesti gümnaasiumiõpilaste Newtoni mehaanika arusaamade analüüsimiseks.”

Õpilased jaotati testi tulemuste alusel 15 alagruppi (0 punkti kuni 14 punkti) (joonis 5). Saadud testi tulemus jagatud testi kogutulemusega annab testi tulemuse protsentides. Kui mõnes grupis ei olnud ühtegi vastajat, jäeti see andmepunkt vaatluse alt kõrvale. Eeltestidest tekkis 13 alagruppi (0 punkti kuni 11 punkti ja 13 punkti) ja järeltestides 14 alagruppi (0 punkti kuni 13 punkti).

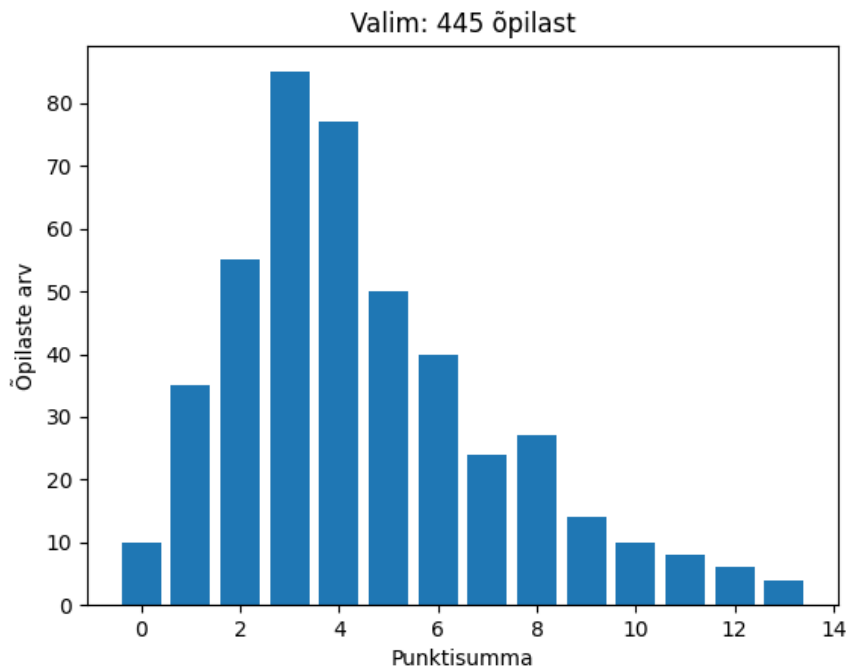
Eeltestis osales 488 õpilast. Õpilaste aritmeetiline keskmine tulemus oli 3,375 punkti (standardhälbe 2,00 punkti) 14 võimalikust. Mediaan oli 3 punkti. Õpilaste keskmine tulemus oli protsentuaalselt 24,1



Joonis 5: Õpilaste jaotus vastavalt saadud eeltesti tulemusele

Järeltestis osales 445 õpilast. Aritmeetiline keskmine oli 4,54 punkti (standardhälbe 2,73 punkti) 14 võimalikust punktist (joonis 6). Mediaan oli 4 punkti. Õpilaste keskmine protsentuaalne tulemus oli 32,4

Seega on Eesti õpilaste tulemused nii eel- kui ka järeltestis madalad.



Joonis 6: Õpilaste jaotus vastavalt saadud järeltesti tulemusele

Uurimisküsimus 3: Kasutades analüüsimiseks IRT ja CTT teadmiste juurdekasvu mõõtmise meetodikaid, siis millise Newtoni mehaanika teadmiste juurdekasvu annab Eesti keskkooliõpilastele kinemaatika ja mehaanika kursuse õpetamine?

Eeltesti puhul saadi keskmiseks testitulemuseks:  $3,375/14 = 24,1$  Järeltesti puhul saadi keskmiseks testitulemuseks:  $4,54/14 = 32,4$

Hake'i (1998) valemi järgi on antud töös Eestis läbiviidud kursuste arusaamade normaliseeritud juurdekasv

$$g = \frac{32,3\% - 24,1\%}{100\% - 24,1\%} = 0,109.$$

Hake'i definitsiooni järgi on antud töös läbiviidud kursuste jooksul õpilaste teadmiste juurdekasv madal.

Selleks, et saaks võrrelda CTT ja IRT järgi leitud arusaamade paranemist logitites, tuleb leida eel- ja järeltesti tulemuste logitite vahe (valemid 1 ja 5)(Wallace and Bailey, 2010):

CTT puhul on õpilaste teadmiste juurdekasv logitites:

$$\ln \frac{32,4\%}{100\% - 32,4\%} - \ln \frac{24,1\%}{100\% - 24,1\%} = 0,412.$$

Et leida õpilaste teadmiste juurdekasv IRT abil, on vaja leida keskmise õpilase eeltestile ja järeltestile õigesti vastamise tõenäosused. Eeltesti õigesti vastamise tõenäosuse leidmiseks leiti kaheparameetrilise IRT mudeli keskmine küsimuse eristatavus  $a = 6,98$  ja keskmine küsimuse keerukus  $b = 0,53$  (tabel 4). Vastavalt IRT kaheparameetrilisele mudelile (valem 3) on

eeltestile keskmiselt õigesti vastamise tõenäosus  $P = 0,0763\%$ .

Keskmise õpilase järeltestile õigesti vastamise tõenäosuse leidmiseks leiti kaheparameetrilise IRT mudeli keskmine küsimuse eristatavus  $a = 6,25$  ja keskmine küsimuse keerukus  $b = 0,48$  (tabel 6). Sellest tulenevalt on järeltestile keskmiselt õigesti vastamise tõenäosus  $P = 0,672\%$ .

IRT analüüsi puhul võetakse teadmiste juurdekasvu arvutamiseks logitites testi tulemuseks küsimusele õigesti vastamise tõenäosus:

$$\ln \frac{0,672\%}{100\% - 0,672\%} - \ln \frac{0,0763\%}{100\% - 0,0763\%} = 2,18 \text{ logitit.}$$

## 4 Arutelu

Uurimisküsimuse “Milline IRT mudel on kõige sobilikum Eesti keskkooliõpilaste FCI eel- ja järeltestide analüüsimiseks?” vastuseks leiti, et nii eeltestis, kui ka järeltestis andis parimaid tulemusi kaheparameetiline IRT mudel, kus varieeriti küsimuste raskuse parameetrit  $b$  ja küsimuste eristatavust  $a$ . Üheparameetrilist mudelit suudeti lähendada mõnede küsimuste halvemini kui kaheparameetrilist mudelit, sest küsimuse eristatavus  $a$  jäi mudeli jooksul muutumatuks. Kolmeparameetiline mudel (tabel 4, tabel 6) ei suutnud kõiki küsimusi lähendada ning lähendatud küsimuste puhul oli “huupi” pakkumise parameeter  $c$  nullilähedane. Kolmeparameetrilist mudelit ei suudeta katsepunktidele täpsemalt lähendada kui kaheparameetrilist mudelit. See, et parameeter  $c$  on kolmeparameetrilise IRT mudeli analüüsi puhul nullilähedane, tähendab seda, et küsimuste valikuvariandid on hästi valitud, õpilased ei vasta küsimustele “huupi”(Wallace and Bailey, 2010) ja kolmeparameetiline IRT analüüs antud küsimustike puhul pole praeguse valimi juures vajalik.

IRT meetodiga saab täpsemaid tulemusi, kui testis üle 50% saanud õpilaste arv kasvaks. Antud valimi puhul oli mõne alagrupi suuruseks üksikud õpilased. Sellest on põhustatud katsepunktide suur hajumine (Lisa 1-20).

Seega kõige paremini sobitus FCI eel- ja järeltesti vastustega IRT kaheparameetiline mudel, millel oli kõige väiksem keskmine  $\chi^2$  ja mis suutis võrreldes kolmeparameetrilise mudeliga leida lähenduskõveraid rohkematele küsimustele.

Uurimisküsimus “Kuidas sobib USA ülikoolide tudengite jaoks välja töötatud FCI test Eesti gümnaasiumiõpilaste mehaanika arusaamade analüüsimiseks” vastamiseks saab ühelt poolt väita, et test sobib Eesti õpilastele halvasti, kuna nii eel- kui järeltesti keskmine tulemus oli tugevalt alla poole punktide (vastavalt 24,1% ja 32,4%) ja seetõttu võib väita, et test ei ole antud valimile kohane (Krull, 2018).

Samas on Eesti õpilaste eeltestide tulemused võrreldavad USA (Hake, 1998) ja Horvaatia (Planinic et al., 2010) koolide testitulemustega. Eesti õpilaste järeltestide tulemused näitasid väga nõrka teadmiste kasvu, mis viitab sellele, et Eesti keskkoolides läbiviidavad mehaanika kursused pole piisavalt interaktiivsed või ei keskendu põhimõistetest arusaamisele.

Kuna Eestis läbi viidud testide madalad tulemused on võrreldavad testide tulemustega USA ülikoolides (Hake, 1998) ja Horvaatias (Plannic, 2010), pole vahet, et FCI on välja töötatud USA ülikoolitudengite jaoks. Eesti 10. klassi õpilased saavad sama (madalaid) tulemusi. Vahe tuleb sisse kui võrrelda teste, mida õpilased teevad pärast mehaanika kursuse läbimist.

Uurimisküsimus “Kasutades analüüsimiseks IRT ja CTT teadmiste juurdekasvu mõõtmise meetodikaid, siis millise Newtoni mehaanika teadmiste juurdekasvu annab Eesti keskkooliõpilastele kinemaatika ja mehaanika kursuse õpetamine?” leidis, et Hake'i normaliseeritud juurdekasv oli 0,110. Hake'i definitsiooni järgi on teadmiste juurdekasv 0,109 madal ja kursus pole suure “kasuteguriga”. USA ülikoolide kursused, mis rakendavad interaktiivseid õppemeetodeid, said järeltestides tunduvalt paremaid tulemusi ja Hake'i

normaliseeritud juurdekasv oli interaktiivsete kursuste puhul “keskmise” või “kõrge”.

Käesolevas töös analüüsitud kursuste madal kasutegur viitab sellele, et läbiviidud kursused polnud piisavalt interaktiivsete õpetamismeetoditega, kursused ei keskendunud piisavalt palju FCI testi stiilis ülesannete arutamisele/lahendamisele või õpilastel oli materjali omandamiseks liiga vähe aega. Mehaanika kursus Eestis koosneb ainult 18-20 75-minutilise tunnist. Hake'i uuringus osalenud USA ülikoolide mehaanikakursuste kestvuse ja mahukuse kohta puudub autoritel ülevaade.

CTT ja IRT kaheparameetrilise mudeli juurdekasv logitites oli vastavalt 0,412 ja 2,18. Juurdekasv logitites tähendab, et sündmuse juhtumise tõenäosus suureneb  $e^x$  korda, kus  $x$  on juurdekasv logitites. See tähendab, et CTT järgi vastas järeldestis keskmine õpilane  $e^{0,412}$  ehk 1,51 korda paremini ja IRT kaheparameetrilise mudeli järgi vastas keskmine õpilane  $e^x$  korda tõenäolisemalt küsimustikule õigesti ehk tõenäosus küsimustikule õigesti vastata kasvas  $e^{2,18}$  korda.

IRT on rohkem mõeldud testi arendamiseks kui olemasoleva testi abil õpilaste tulemuste paranemise uurimiseks. Kiire ja lihtsa analüüsimetodi kursuse kasulikkusest annab juba Hake'i normaliseeritud juurdekasvu meetod.

Edasiarendusena võiks tõlkida ülejäänud Eesti gümnaasiumi füüsika kursuste kohta käivad mõistetestid eesti keelde ning nende kvaliteeti IRT abil uurida. Samuti võiks suure valimiga FCI testimist teostada põhikooli mehaanika kursuse eel ja järel ning ülikoolis bakalaureuse tasemel mehaanika kursuse eel ja järel. Nende testitulemuste IRT meetodiga hindamine annab rohkem infot, millistes haridusastmetes antud test on kasutatav.

Klassikalised õppemeetodid annavad madala võimekuse kasvu (Hake, 1998). Võimekuse kasv on interaktiivsete meetodidega suurem, seega tuleb edasiarendusena välja töötada interaktiivsed õppemeetodid ja nende mõju õpilaste võimekusele FCI abil mõõta. Selle läbi saab luua kursuse, mille puhul õpilaste võimekuse kasv on võimalikult suur. Sama katse saab korrata ka teiste kursuste puhul nii põhikoolis, gümnaasiumis kui ka ülikoolis.

Yasuda et al (2021) lõi interaktiivselt muutuva FCI testi, millega vähenes testi teostamiseks kulunud aeg. Tartu ülikooli akadeemiline test on juba interaktiivselt muutuvaks tehtud (Silm, 2023). Sama asja saab teha ka pikemate teiste mõistetestidega. Seeläbi saavad õpetajad kulutada vähem aega ühe testi tegemisele ja testimist on võimalik teha tihedamini, et saada pidevalt tagasisidet õpilase arengu kohta.

Käesoleva töö puhul ei küsitud kirjalikku nõusolekut ei õpilastelt ega nende vanematelt, et kas õpilased üldse on nõus testis osalema. Edasiste uuringute puhul tuleb seda aga teha, et uuringud oleksid veelgi paremini kooskõlas hea teadustavaga ja vastaksid veelgi rohkem eetikanõuetele (Eetikakeskus, sa).

Antud töö piiranguks on väike valim. Horvaatias teostati 2006/2007 õppeaastal FCI test 1676 gümnaasiumi õpilastega (Planinic et al, 2010) ning Ohio Ülikoolis vahemikus 2003 kuni 2007 eeltestis 2802 ja järeltestis 2729 tudengiga (Wang and Bao, 2010). Järgnevalt tuleks Eestis teostada FCI testimine rohkemates gümnaasiumides rohkemate õpilastega, et saada teha

järeldusi Eesti gümnaasiumites õpetatava mehaanika kursuse kohta. Samuti tasub FCI testi korrata igal aastal uute õpilastega, et saada kursuse kasulikkuse muutuste kohta ülevaadet läbi aja.

## **Tänuõnad**

Täname oma juhendajaid, perekondi, lähedasi, sõpru, kursusekaaslasid, töökaaslasid, koorikaaslasid ja lemmikloomi meie õpingutele kaasaegselt ja rasketel hetkedel toetamast.

## Autorsuse kinnitus

Valle: kogusin eel- ja järeltesti tulemused kokku ja kirjutasin Pythoni koodi toorandmete analüüsimiseks. Pythoni koodis kasutasin teke pandas, numpy, math, scipy, matplotlib ja irt.parameter.estimation. Toorandmetest eemaldasid üleliigsed andmeväljad nagu nimi, meiliaadress, kuupäev, vastusevariant jne. Sorteerisin ja grupeerisin andmed, et neid saaks kasutada edasises andmetöötles. Analüüsi käigus lõin töö tarbeks joonised ja tabelid ning lähendasin ühe-, kahe- ja kolmeparameetrisi mudeleid katsepunktidele. Mudelite lähendamisel pidin algparameetreid varieerima, et parim lähendus saavutada. Töö tekstilise osa kirjutasime Google Docsis, kuid vormistamise tegin dokumentide ettevalmistussüsteemiga LaTeX (<https://www.latex-project.org/>). LaTeXi eelised Wordi ees on valemite, sümbolite ja mitte-ladina tähtede mugavam vormistamine.

Agnes: teostasid õpilastega eel- ja järeltesti. Otsisin allikaid, kontrollisin Valle tehtud mudeleid ja andsin soovitusi analüüsiks. Sõnastasin uurimisküsimused. Kirjutasin teooriateatükke, kirjutasin Valle genereeritud joonistele ja tabelitele kirjeldusi. Kirjutasin sissejuhatuse ja kokkuvõtte. Arvutasin Hake'i juurdekasvu, arvutasin juurdekasvud logitites nii IRT kui CTT meetodil. Teostasid viimase LaTeXi viimistluse.

Töö tekstilise osa kirjutamisel oli mõlemal autoril võrdväärne panus. Lugesime teemaga seotuid teadusartikleid, mille põhjal kirjutasime teooria peatüki. Järgnevalt kirjutasime meetodi, tulemuste ja arutelu peatükid. Viimase osana kirjutasime sissejuhatuse ja eesti- ja võõrkeelsed kokkuvõtted.

Kinnitame, et oleme koostanud ise käesoleva lõputöö ning toonud korrektselt välja teiste autorite ja toetajate panuse. Töö on koostatud lähtudes Tartu Ülikooli haridusteaduste instituudi lõputöö nõuetest ning on kooskõlas heade akadeemiliste tavadega.

Valle Morel  
/allkirjastatud digitaalselt/  
digitaalselt/

25. mai 2023. a.

Agnes Vask  
/allkirjastatud

25. mai 2023. a.

## Kirjandus

- Aarna, S. (2020). Elva gümnaasiumi kaheksandate ja üheksandate klasside õpilaste väärarusaamad seoses newtoni mehaanikast arusaamisega. Publitseerimata magistritöö.
- Barbera, J. (2013). A psychometric analysis of the chemical concepts inventory. *Journal of Chemical Education*, 90(5):546–553.
- Bretz, S. L. and Linenberger, K. J. (2012). Development of the enzyme–substrate interactions concept inventory. *Biochemistry and Molecular Biology Education*, 40(4):229–233.
- Broder, D. L., McColgan, M. W., and Finn, R. A. (2015). Emca - electricity and magnetism conceptual assessment.
- Eetikakeskus (s.a.). Teadustöö tegemine. <https://www.eetika.ee/et/2-teadustoo-tegemineisikud>.
- Epstein, J. (1997). Basic skills diagnostic test (bsdt).
- Epstein, J. (2013). The calculus concept inventory-measurement of the effect of teaching methodology in mathematics. *Notices of the American Mathematical Society*, 60(8):1018–1027.
- Garvin-Doxas, K. and Klymkowsky, M. (2007). Building the biology concept inventory. *Unpublished manuscript that can be found at <http://bioliteracy.net/CABS>, 202007.*
- Gipps, C. (1994). Developments in educational assessment: what makes a good test? *Assessment in Education: Principles, Policy & Practice*, 1(3):283–292.
- GRÕK (2011). Riigi teataja i 2011, 2. <https://www.riigiteataja.ee/akt/123042021011?leiaKehtiv>.
- Hake, R. R. (1998). Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses. *American journal of Physics*, 66(1):64–74.
- Halloun, I. A. and Hestenes, D. (1985). The initial knowledge state of college physics students. *American journal of Physics*, 53(11):1043–1055.
- Han, J., Bao, L., Chen, L., Cai, T., Pi, Y., Zhou, S., Tu, Y., and Koenig, K. (2015). Dividing the force concept inventory into two equivalent half-length tests. *Physical Review Special Topics-Physics Education Research*, 11(1):010112.
- Hestenes, D., Wells, M., and Swackhamer, G. (1992). Force concept inventory. *The physics teacher*, 30(3):141–158.

- Humphreys, E. (2015). First steps of putting research into practice: Utilizing concept inventories to identify biochemistry misconceptions and the development of a guided inquiry activity to correct the identified misconceptions.
- Jacobi, A., Martin, J., Mitchell, J., and Newell, T. (2003). A concept inventory for heat transfer. In *33rd Annual Frontiers in Education, 2003. FIE 2003.*, volume 1, pages T3D–T3D. IEEE.
- Krull, E. (2018). Pedagoogilise psühholoogia käsiraamat. 3.
- Madsen, A., McKagan, S. B., and Sayre, E. C. (2022). How are research-based assessment instruments developed and validated?
- Mashburn, D. (2021). Irt parameter estimation routines. [https://github.com/pluralsight/irt\\_parameter\\_estimation](https://github.com/pluralsight/irt_parameter_estimation).
- McColgan, M. W., Finn, R. A., Broder, D. L., and Hassel, G. E. (2017). Assessing students' conceptual knowledge of electricity and magnetism. *Physical Review Physics Education Research*, 13(2):020121.
- McKagan, S., Perkins, K., and Wieman, C. (2008). Quantum mechanics conceptual survey.
- Midkiff, K. C., Litzinger, T. A., and Evans, D. (2001). Development of engineering thermodynamics concept inventory instruments. In *31st Annual frontiers in education conference. Impact on engineering and science education. Conference proceedings (Cat. No. 01CH37193)*, volume 2, pages F2A–F23vol. IEEE Computer Society.
- Mulford, D. R. and Robinson, W. R. (2002). An inventory for alternate conceptions among first-semester general chemistry students. *Journal of chemical education*, 79(6):739.
- Planinic, M., Ivanjek, L., and Susac, A. (2010). Rasch model based analysis of the force concept inventory. *Phys. Rev. ST Phys. Educ. Res.*, 6:010103.
- PRÕK (2011). Riigi teataja i 2011, 1. <https://www.riigiteataja.ee/akt/129082014020?leiaKehtiv>.
- Rasch, G. (1980). Probabilistic models for some intelligence and attainment tests. eessõna ja järelsõna b.d. wright (laiendatud väljaanne). Esmakordselt avaldatud 1960.
- Rats, L. (2023). Riigieksamid muutuvad järk-järgult elektrooniliseks. <https://www.err.ee/1608958882/riigieksamid-muutuvad-jark-jargult-elektrooniliseks>.
- Shirazi, S. (2017). Student experience of school science. *International Journal of Science Education*, 39(14):1891–1912.
- Sijtsma, K. (2009). On the use, the misuse, and the very limited usefulness of cronbach's alpha. *psychometrika*, 74:107–120.

Silm, G. (2023). Kuidas valmistuda tartu Ülikooli akadeemiliseks testiks?

Smith, M. K., Wood, W. B., and Knight, J. K. (2008). The genetics concept assessment: a new concept inventory for gauging student understanding of genetics. *CBE—Life Sciences Education*, 7(4):422–430.

Tooding, L.-M. (2020). Cronbachi kordaja. <https://samm.ut.ee/cronbachi-kordaja>.

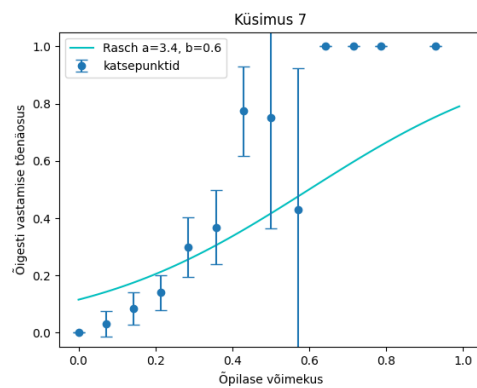
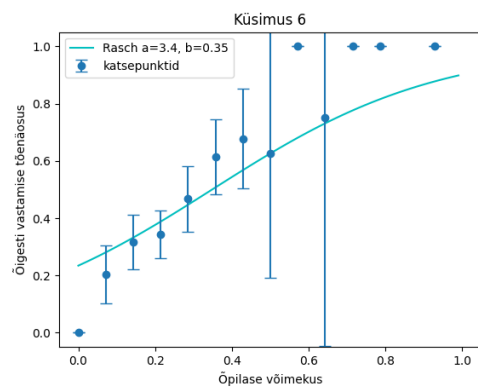
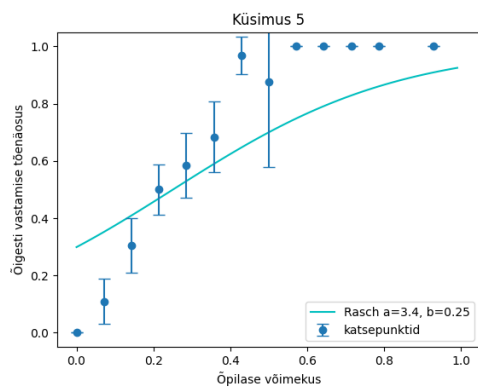
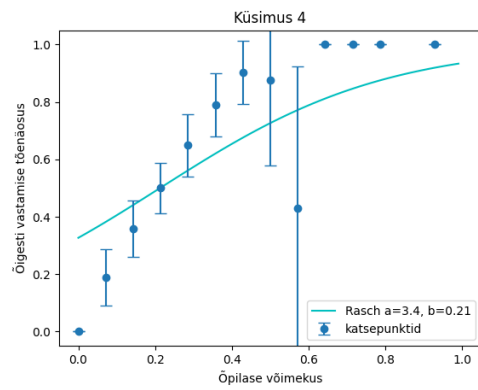
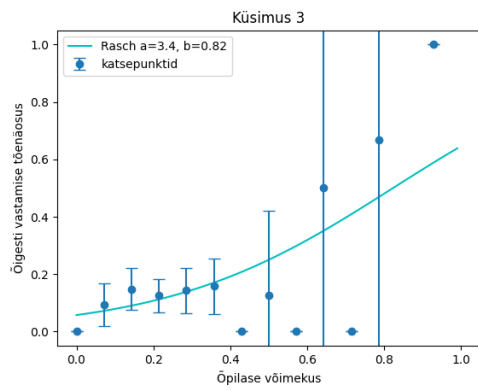
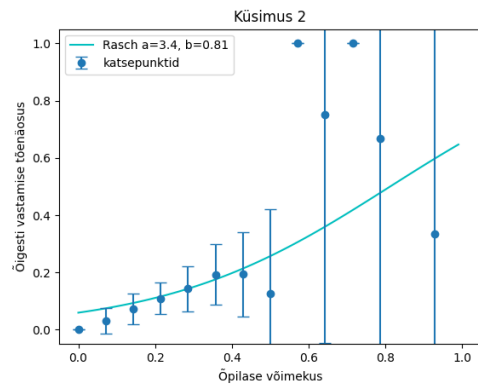
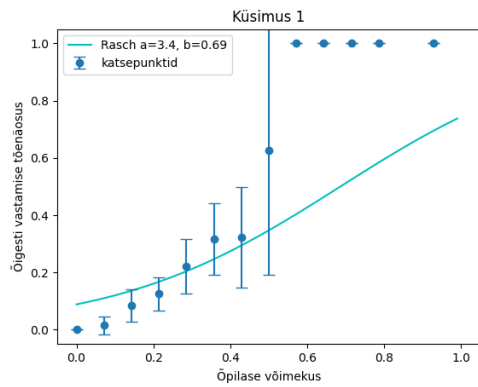
Wallace, C.Š. and Bailey, J. M. (2010). Do concept inventories actually measure anything. *Astronomy Education Review*, 9(1):010116.

Wang, J. and Bao, L. (2010). Analyzing force concept inventory with item response theory. *American Journal of Physics*, 78(10):1064–1070.

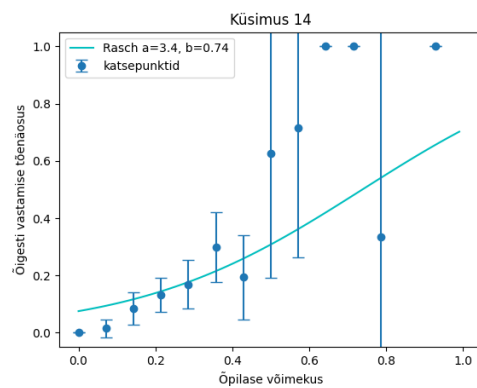
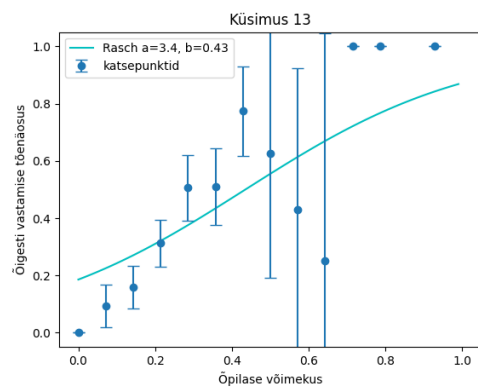
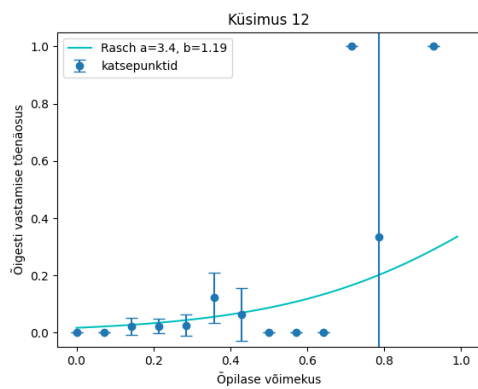
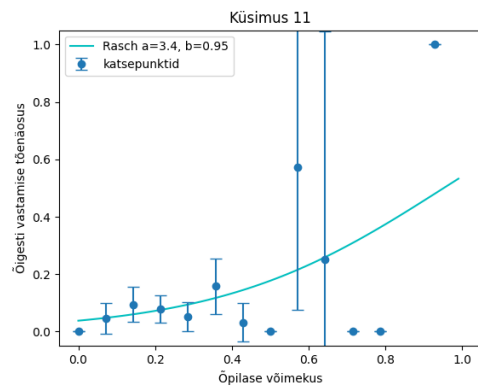
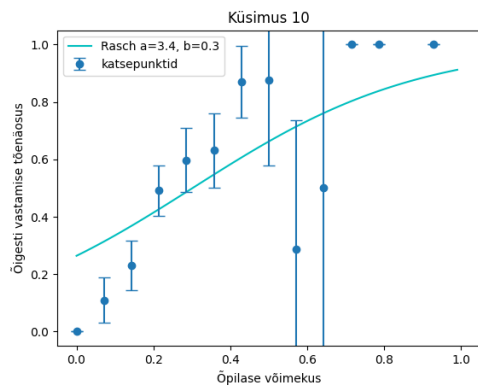
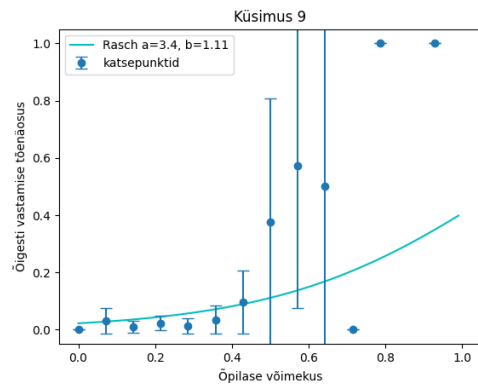
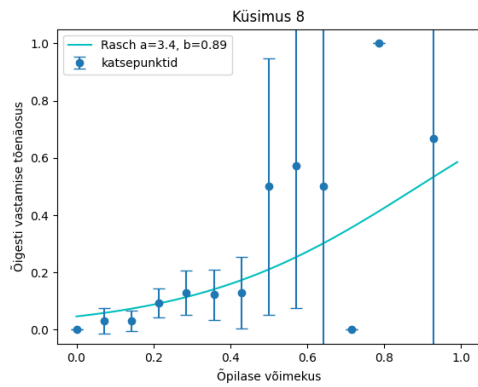
Yasuda, J.-i., Mae, N., Hull, M. M., and Taniguchi, M.-a. (2021). Optimizing the length of computerized adaptive testing for the force concept inventory. *Phys. Rev. Phys. Educ. Res.*, 17:010115.

Õpiedukustestid (s.a.). <https://www.tlu.ee/opmat/ka/opiobjekt/oppevahendid/piedukustestid.html>.

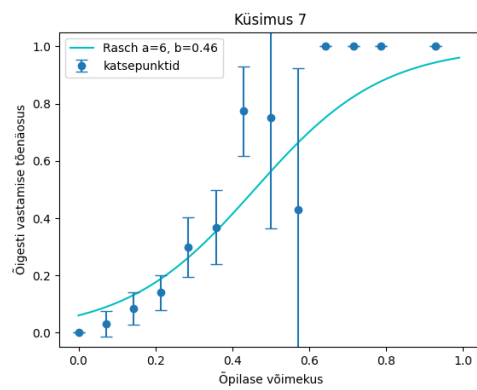
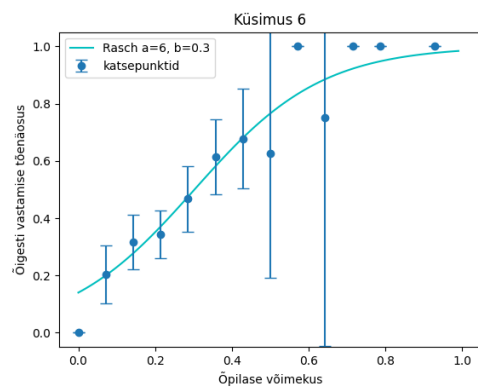
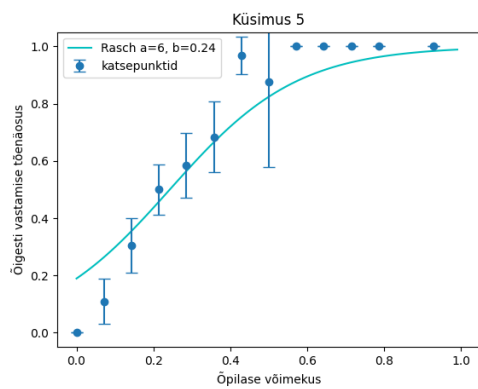
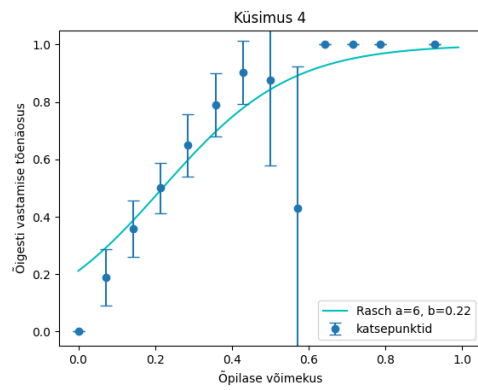
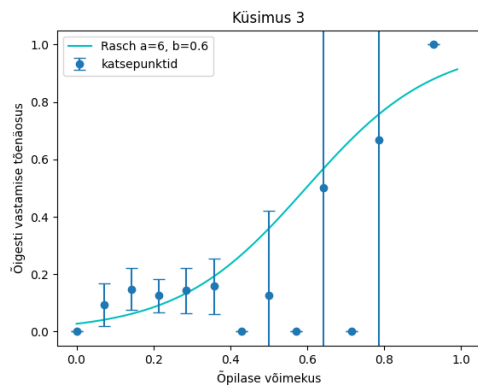
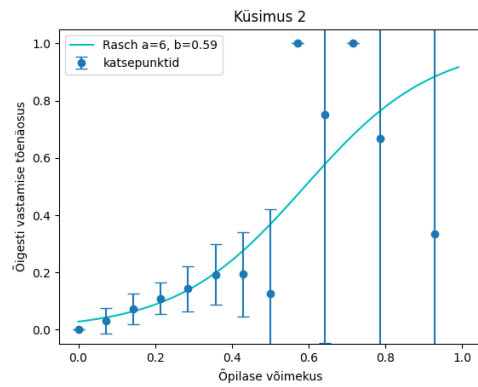
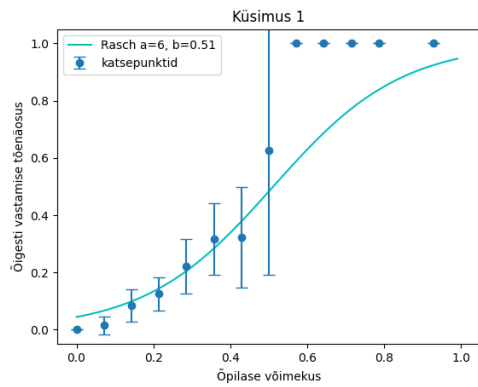
# Lisa 1. 1PL mudel eeltest $a=3.4$ küsimused 1–7



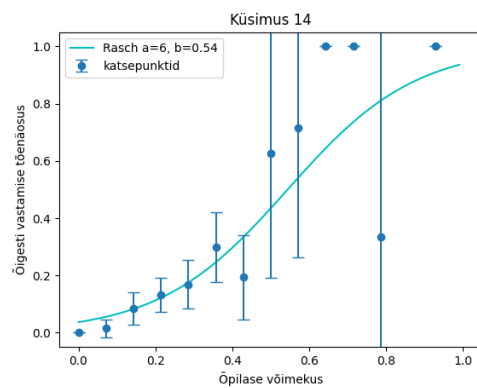
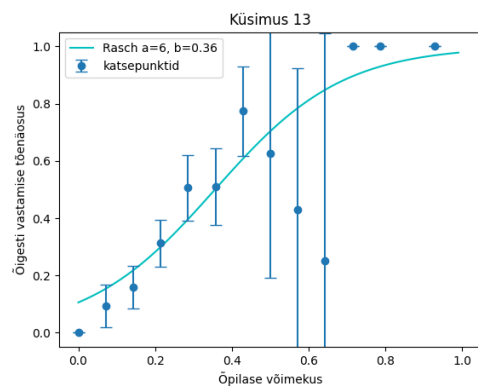
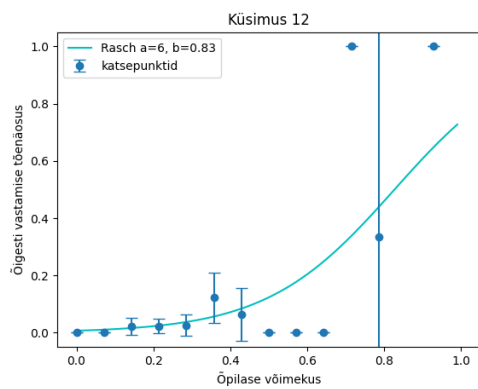
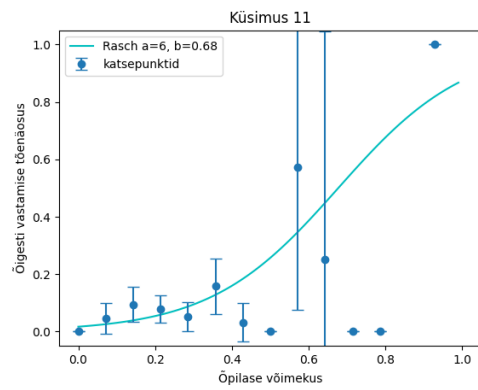
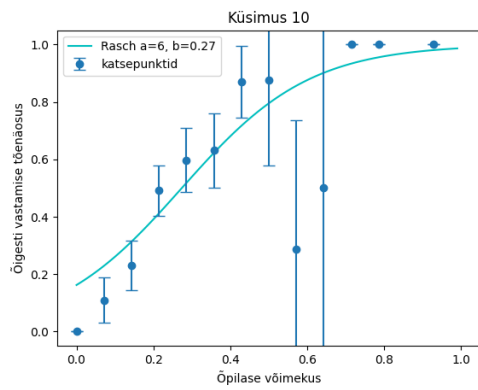
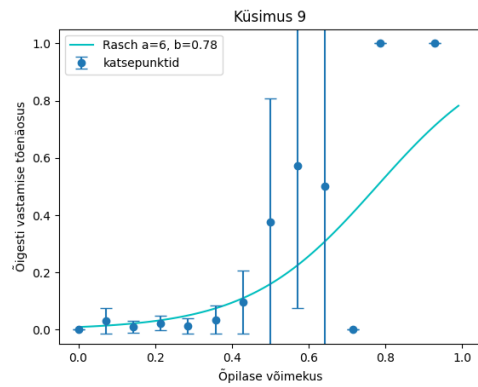
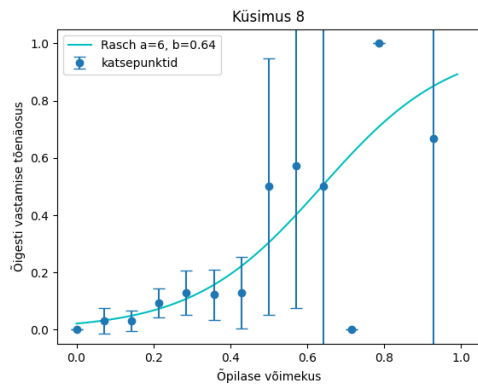
## Lisa 2. 1PL mudel eeltest $a=3.4$ küsimused 8–14



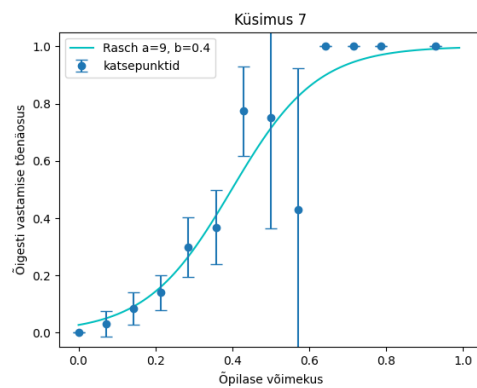
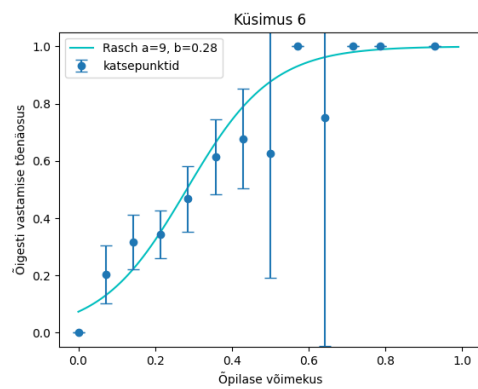
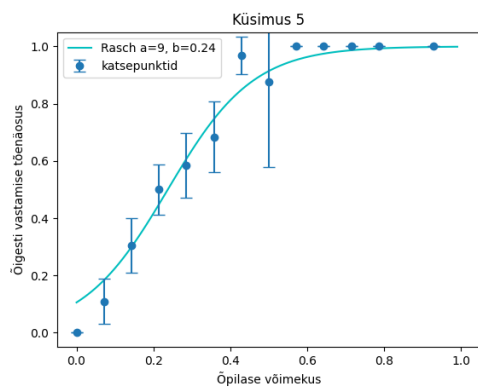
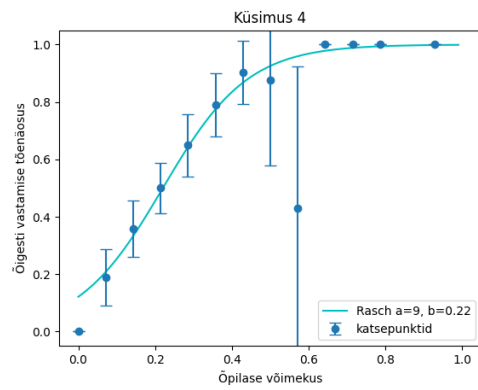
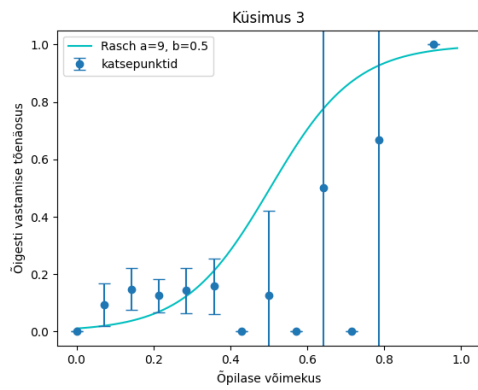
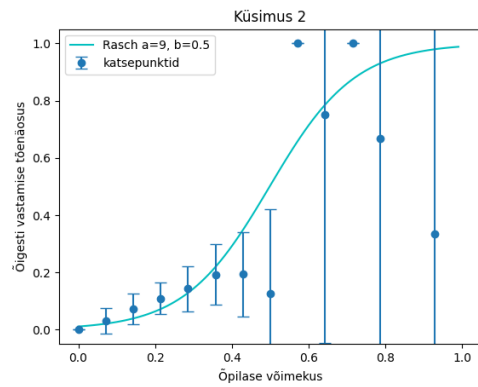
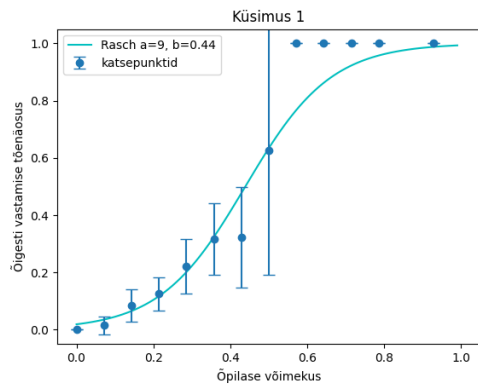
# Lisa 3. 1PL mudel eeltest $a=6$ küsimused 1–7



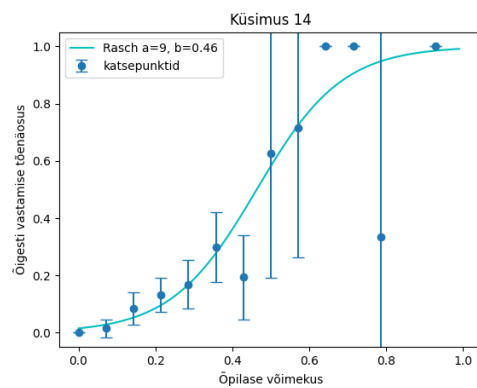
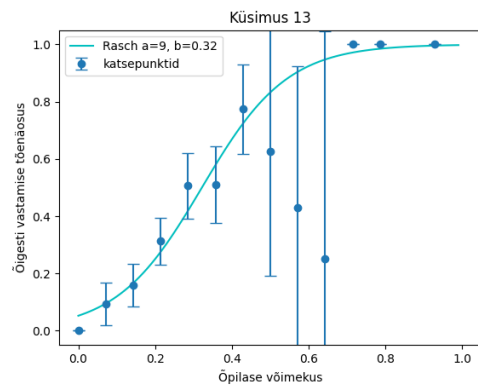
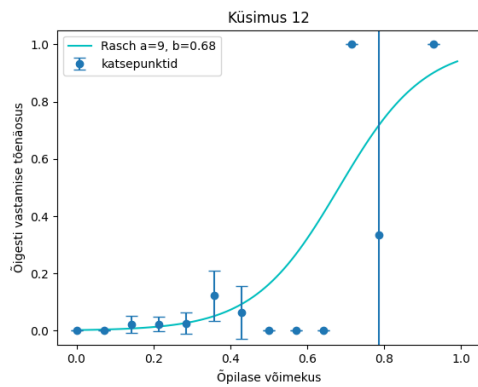
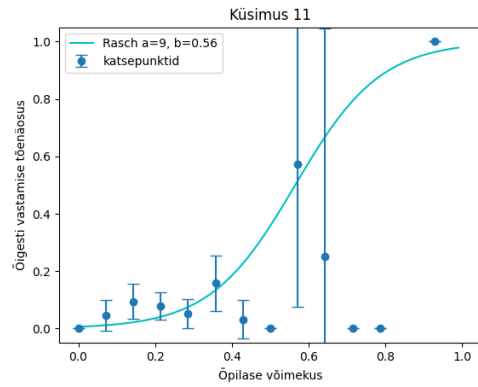
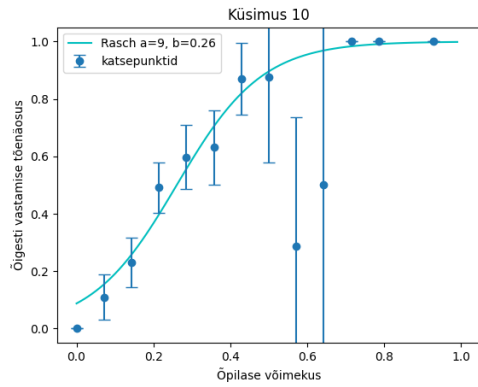
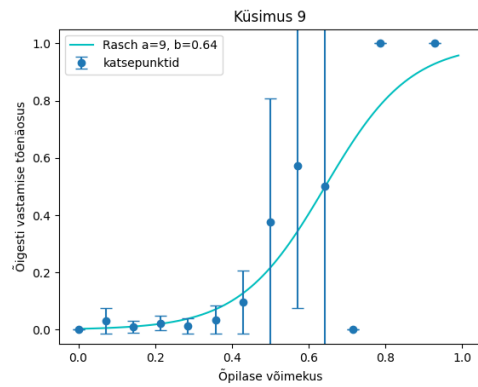
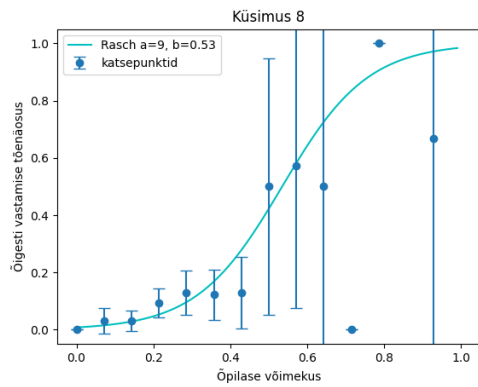
# Lisa 4. 1PL mudel eeltest $a=6$ küsimused 8–14



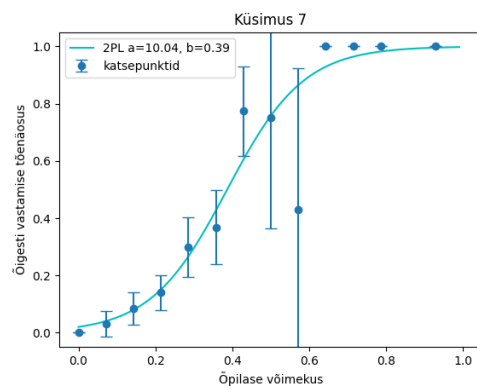
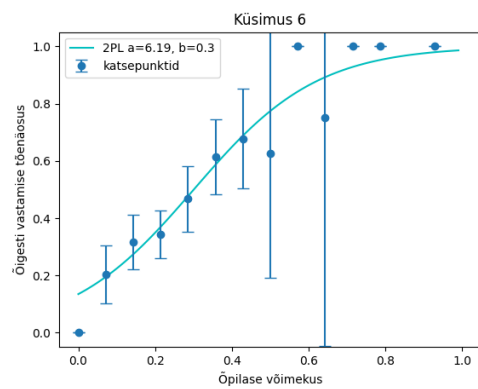
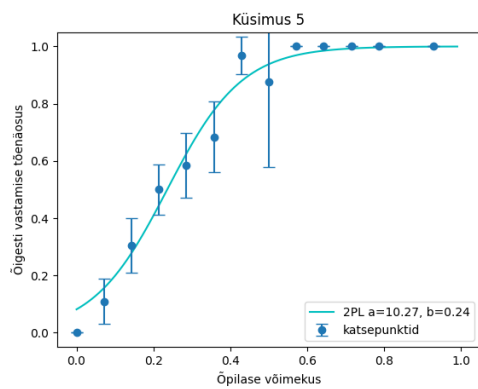
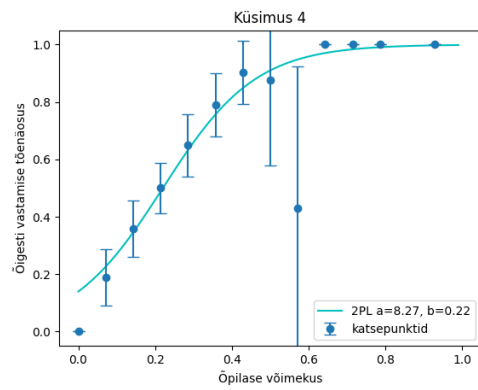
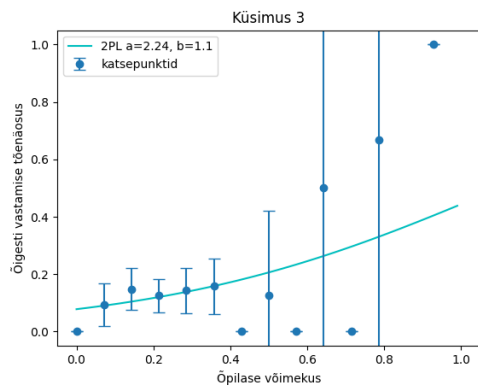
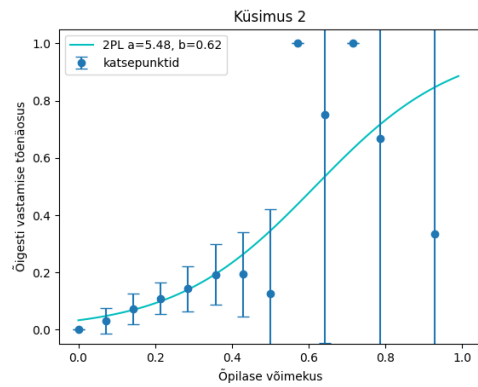
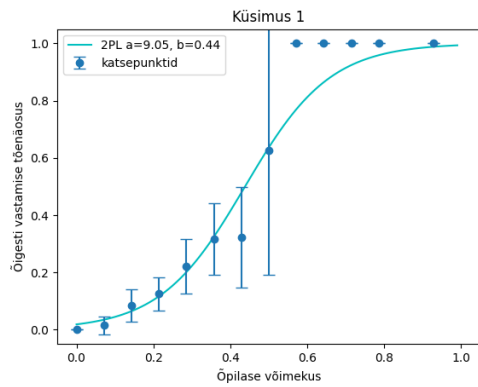
# Lisa 5. 1PL mudel eeltest $a=9$ küsimused 1–7



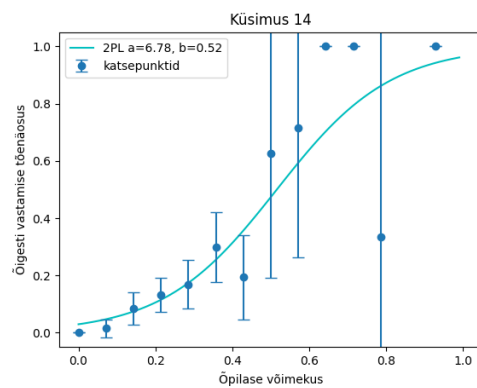
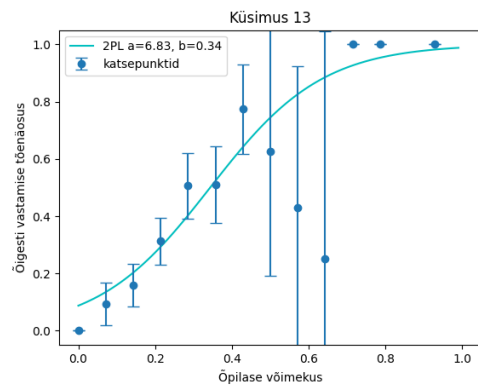
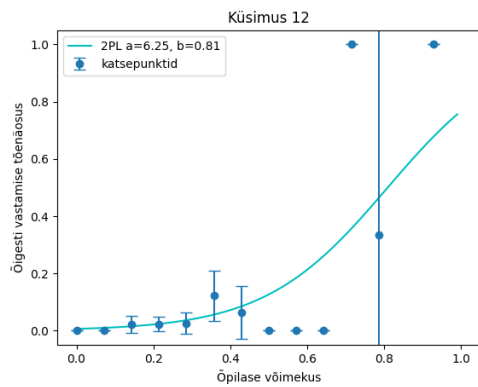
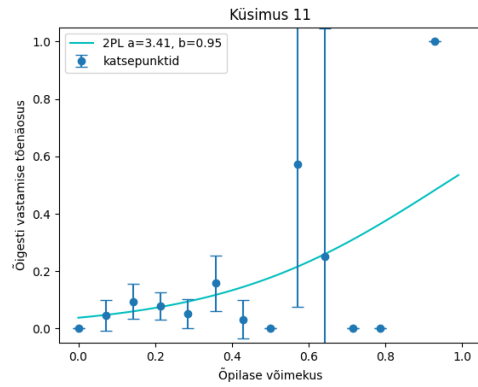
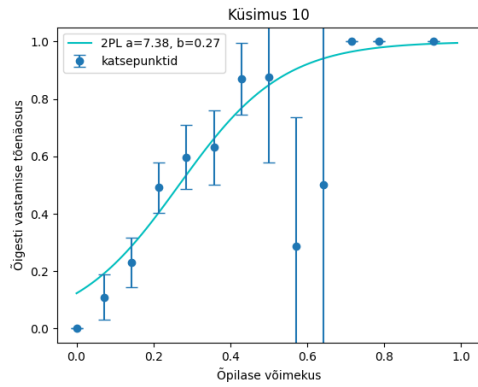
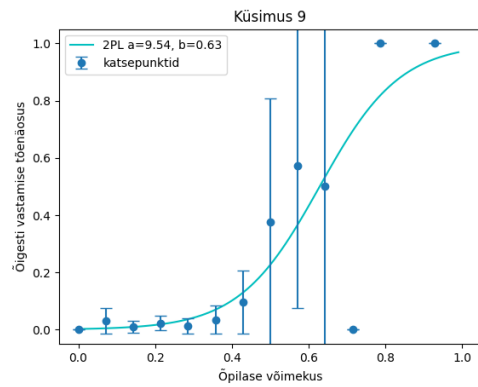
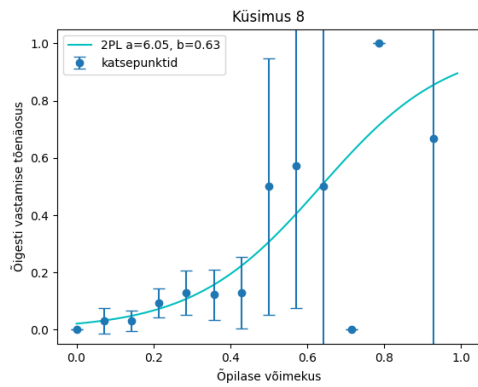
# Lisa 6. 1PL mudel eeltest $a=9$ küsimused 8–14



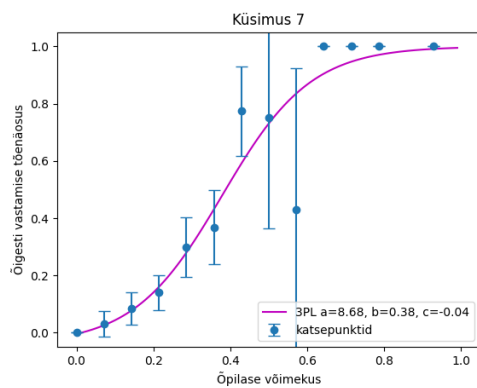
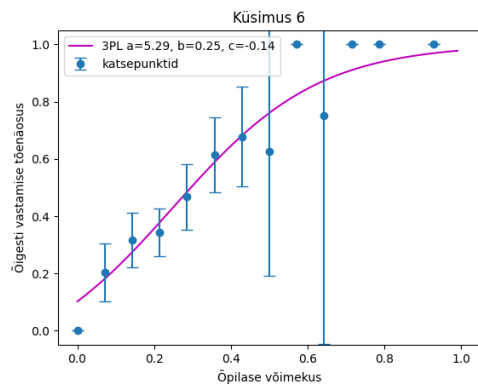
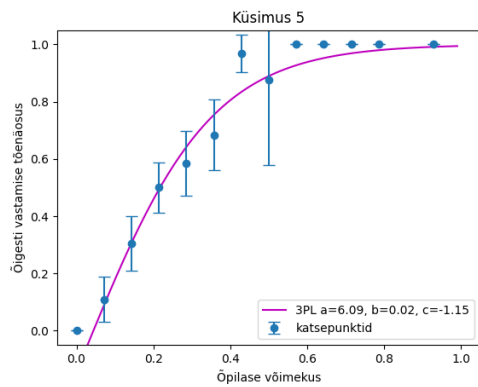
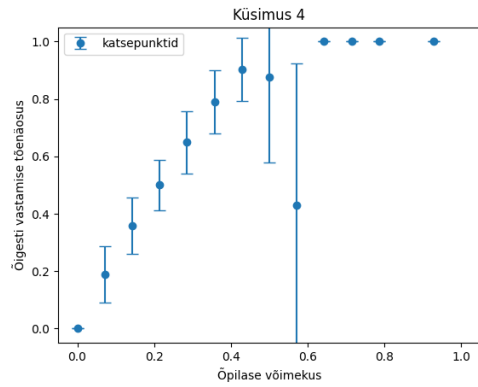
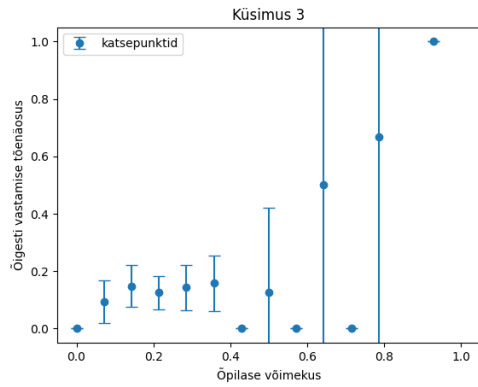
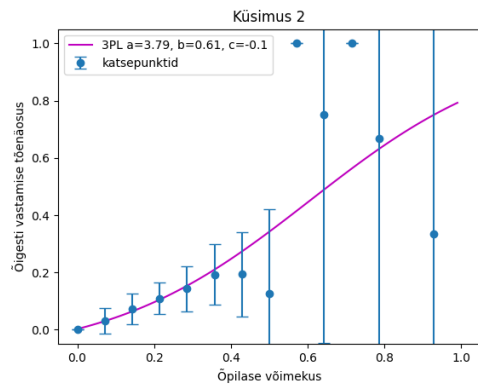
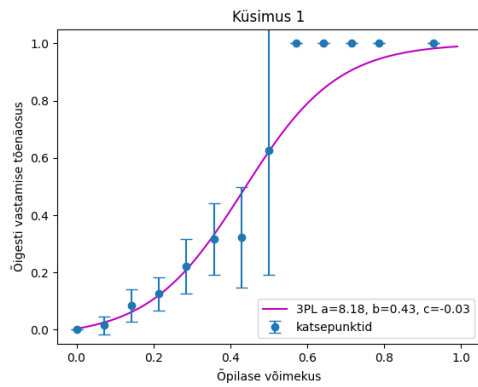
# Lisa 7. 2PL mudel eeltest küsimused 1–7



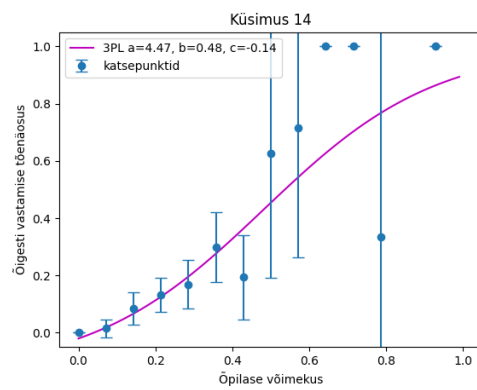
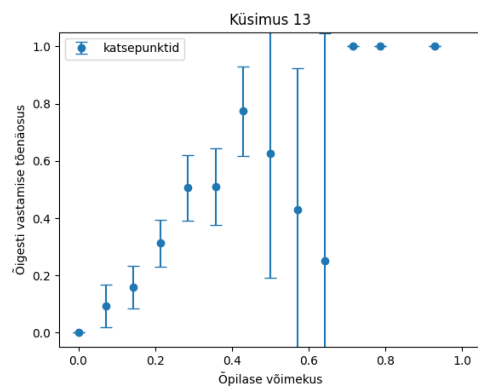
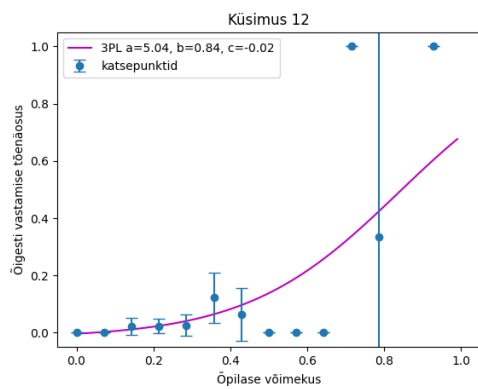
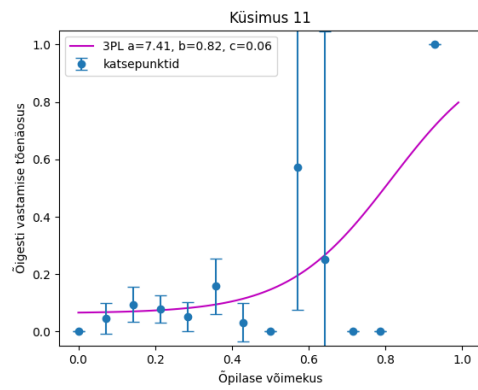
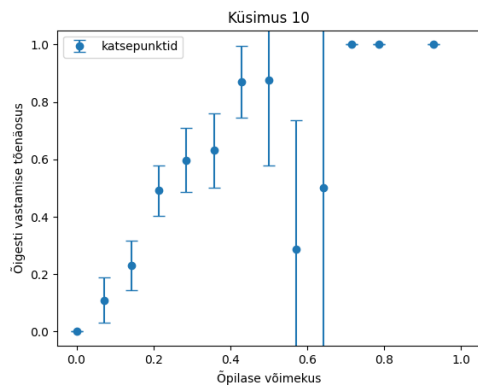
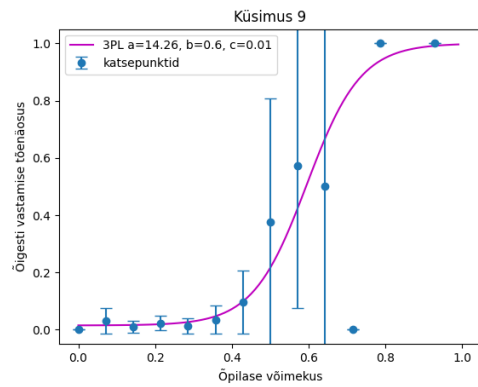
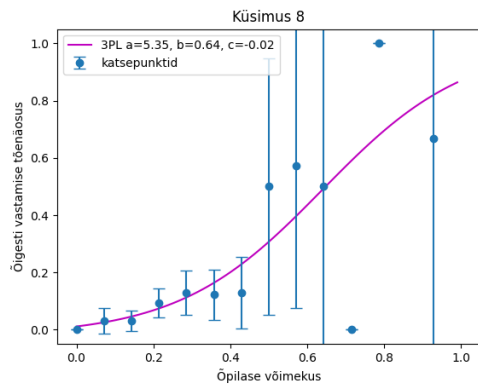
# Lisa 8. 2PL mudel eeltest küsimused 8–14



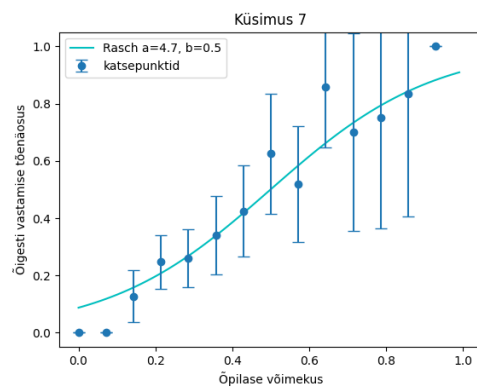
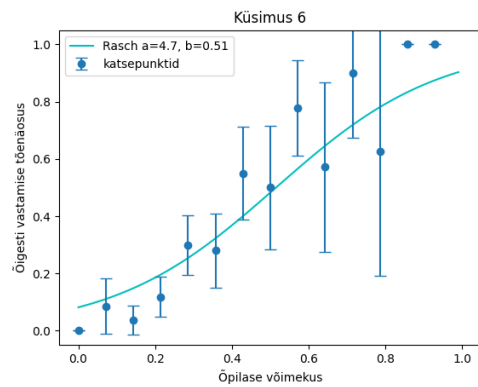
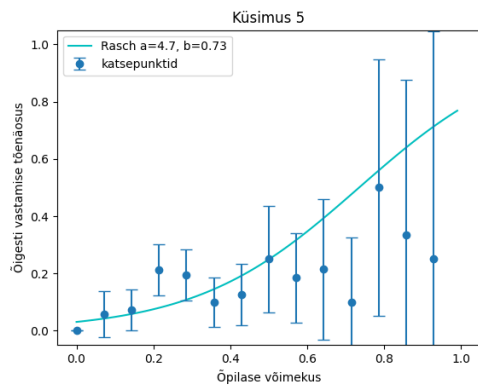
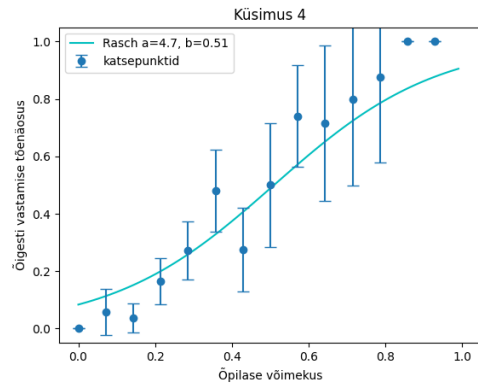
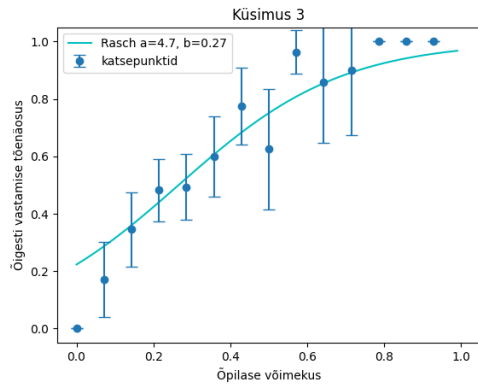
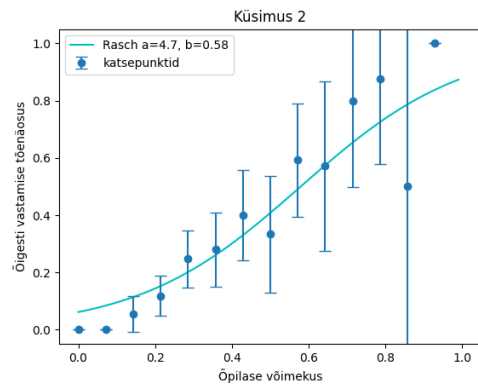
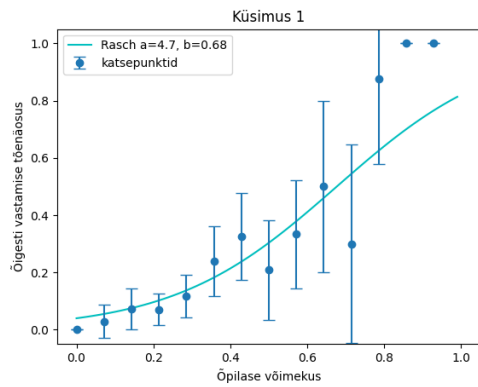
# Lisa 9. 3PL mudel eeltest küsimused 1–7



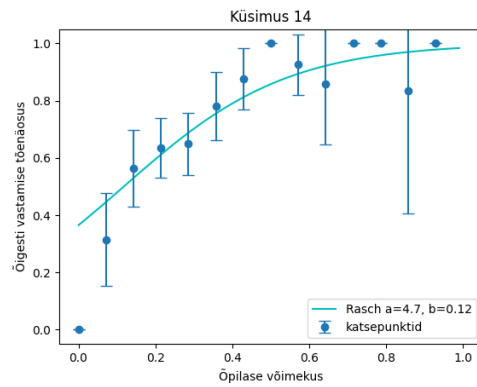
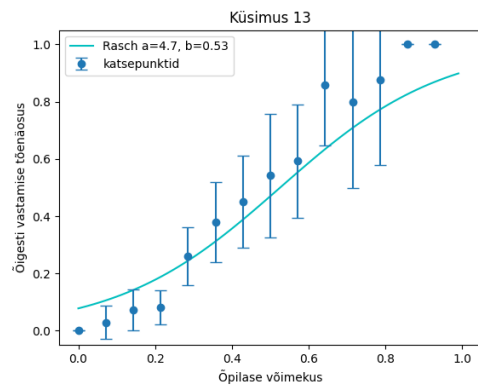
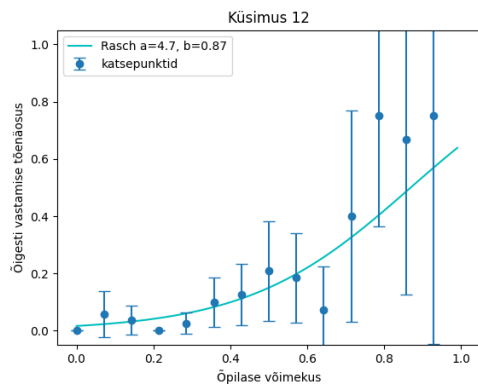
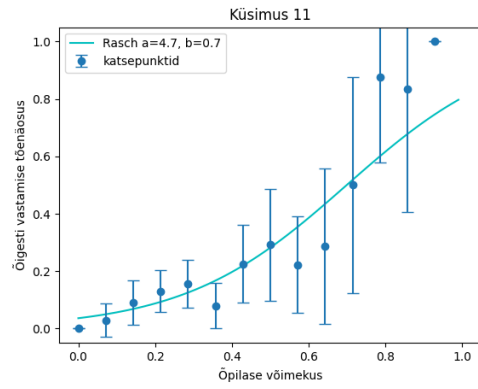
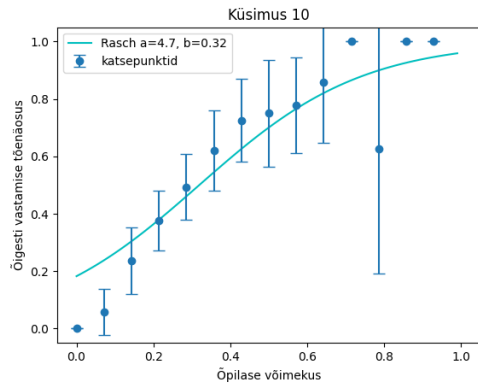
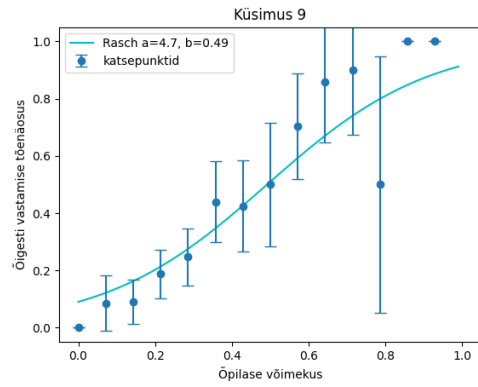
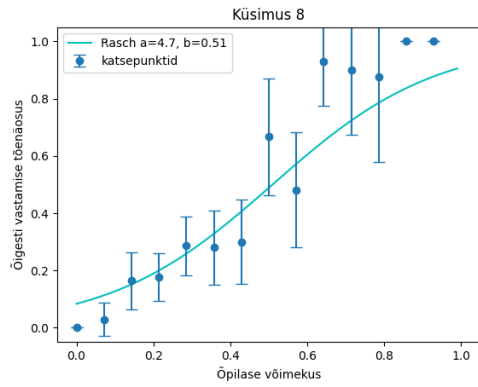
# Lisa 10. 3PL mudel eeltest küsimused 8–14



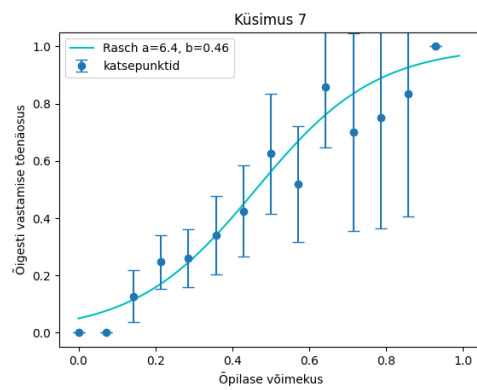
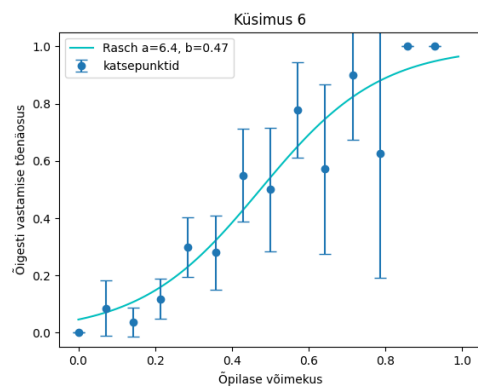
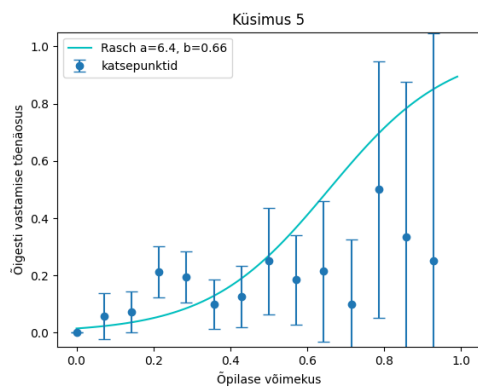
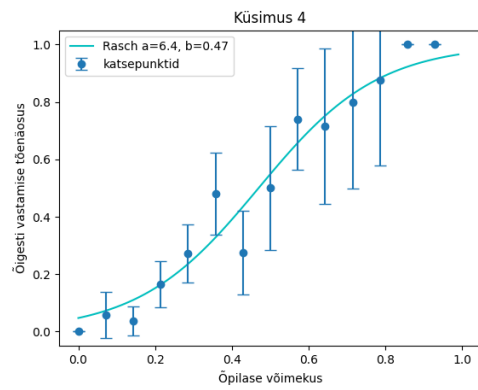
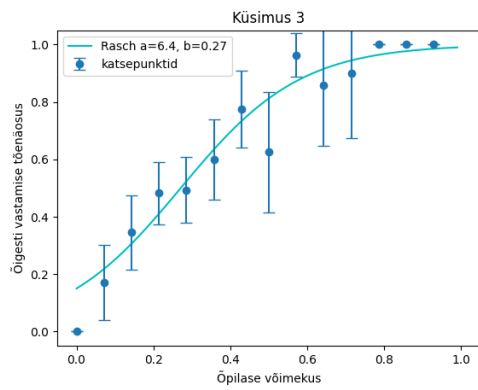
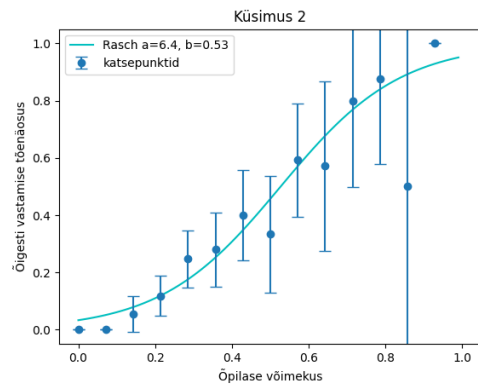
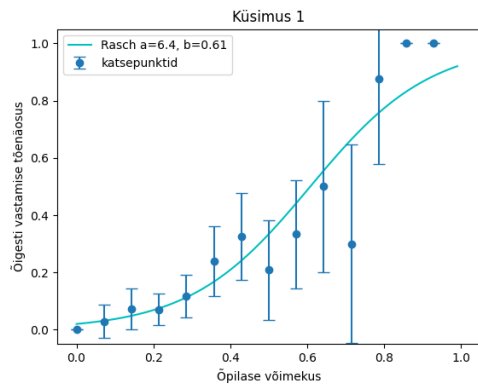
# Lisa 11. 1PL mudel järeltest $a=4.7$ küsimused 1–7



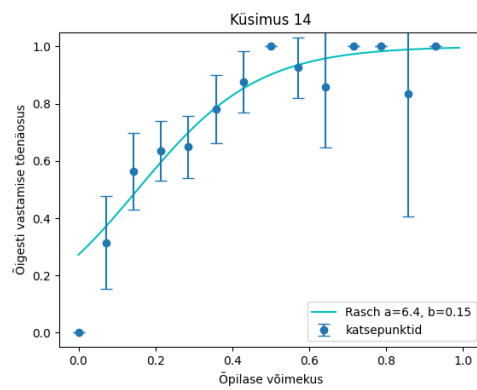
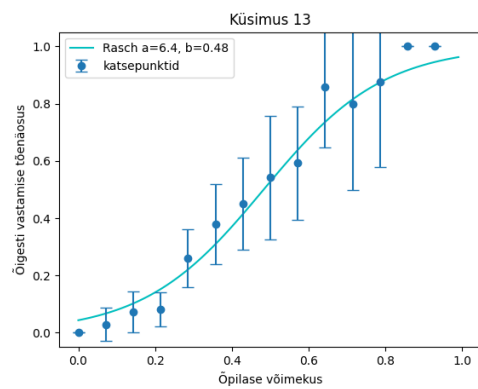
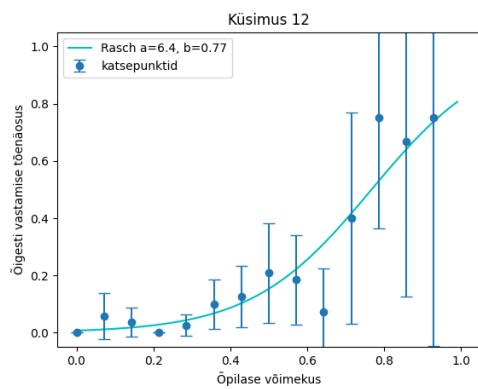
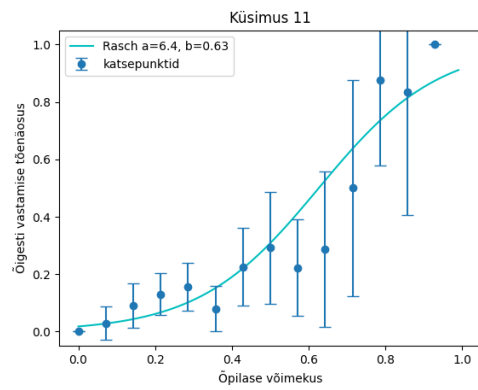
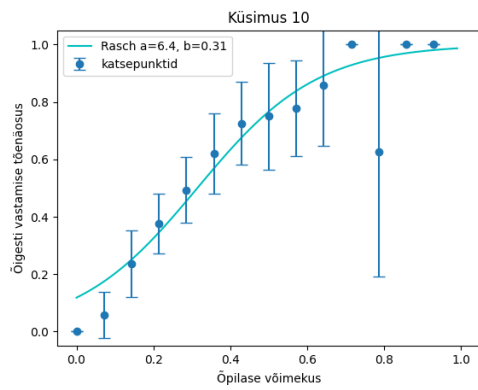
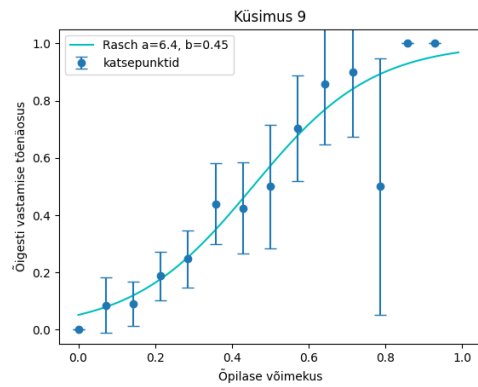
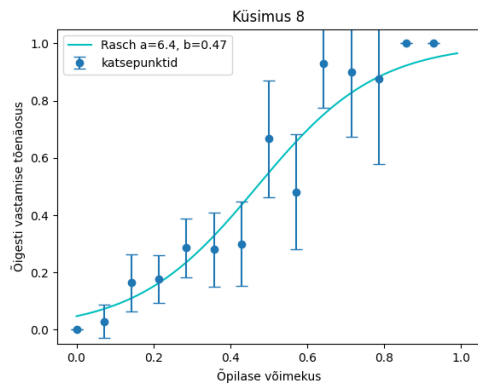
# Lisa 12. 1PL mudel järeltest $a=4.7$ küsimused 8–14



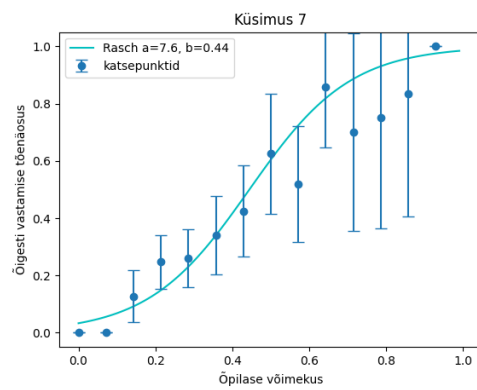
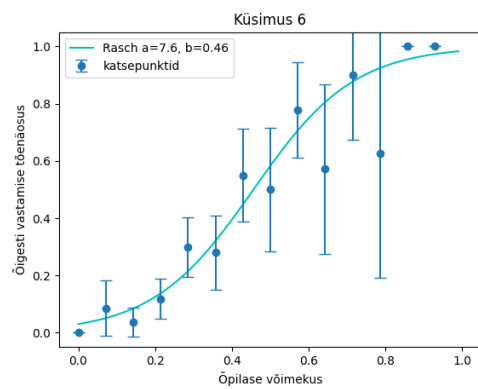
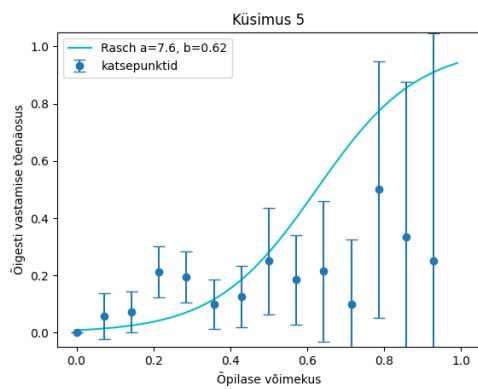
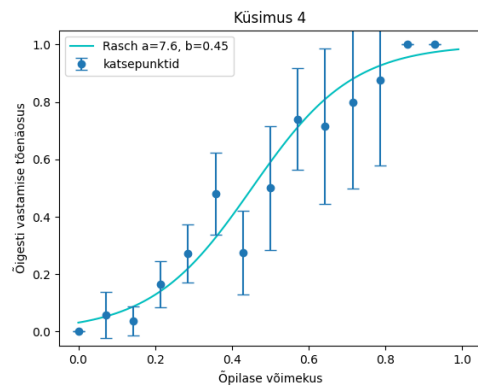
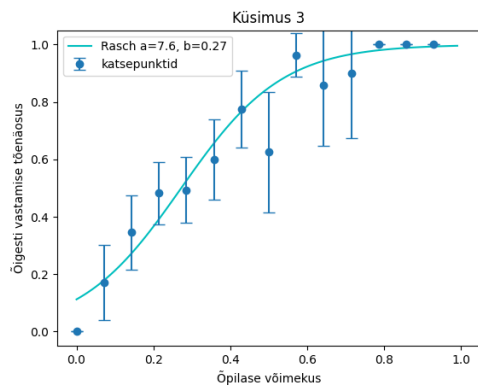
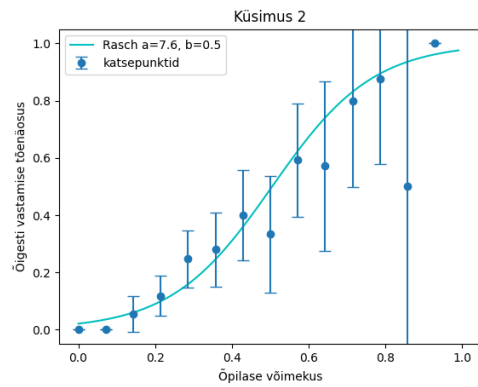
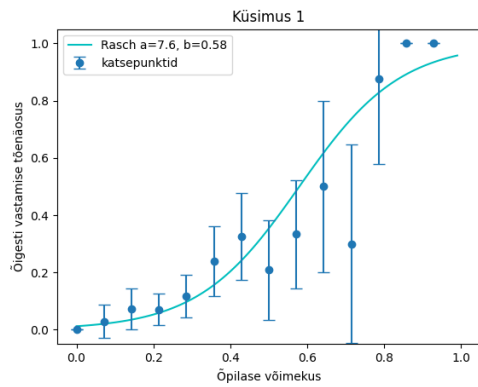
# Lisa 13. 1PL mudel järeltest $a=6.4$ küsimused 1–7



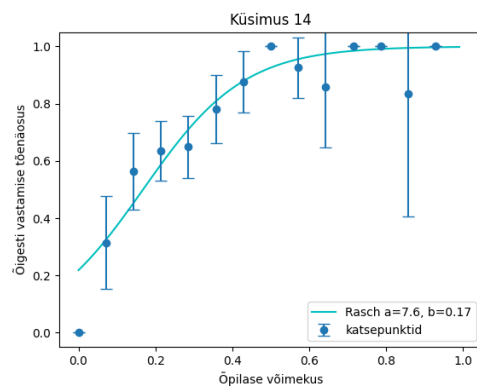
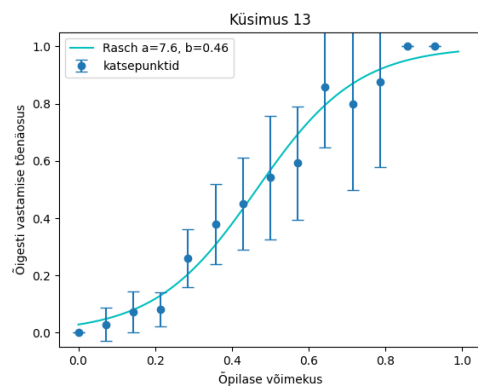
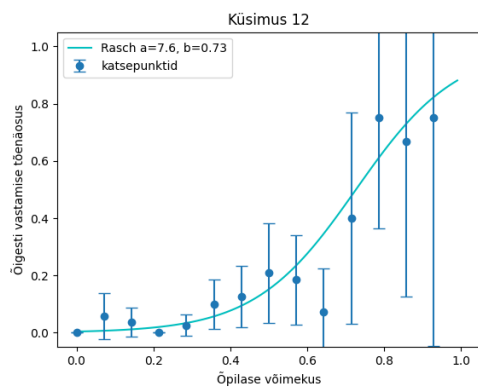
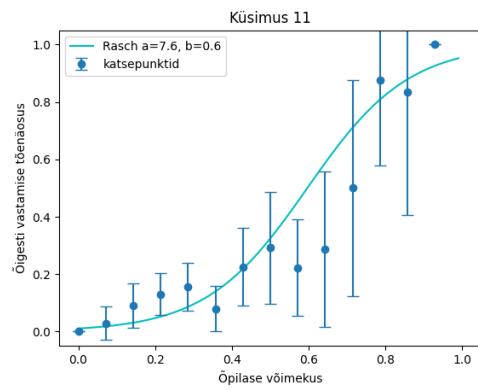
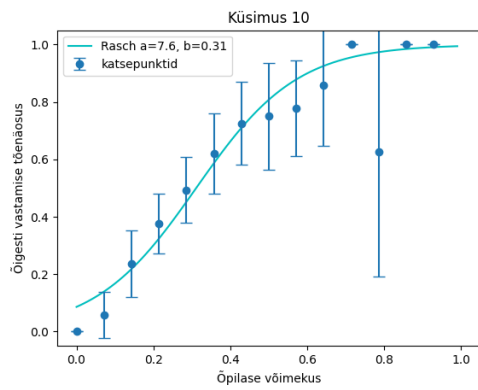
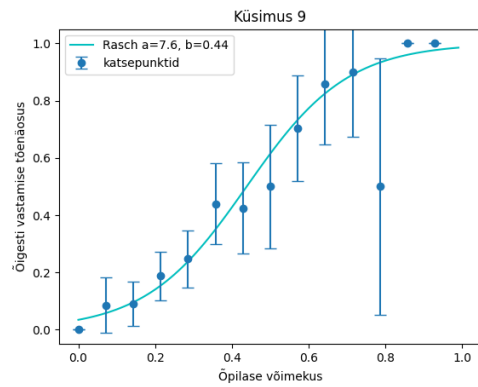
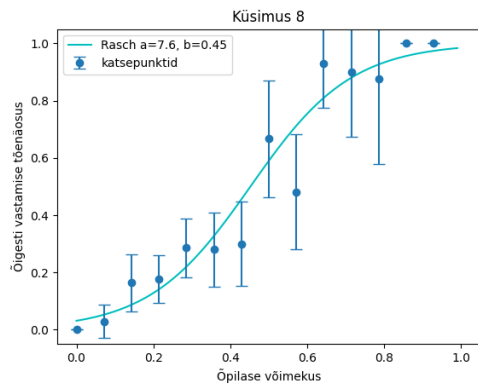
# Lisa 14. 1PL mudel järeltest $a=6.4$ küsimused 8–14



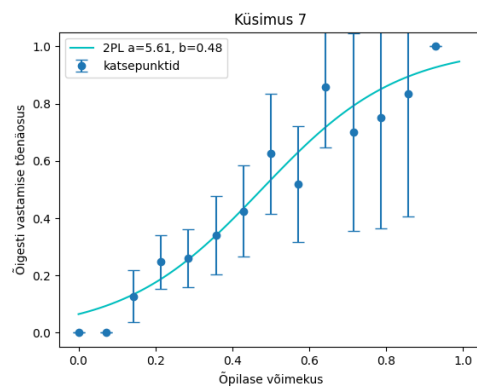
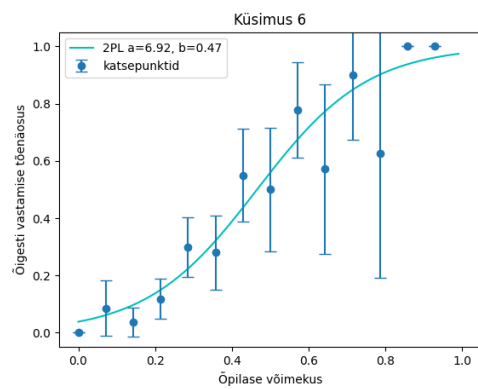
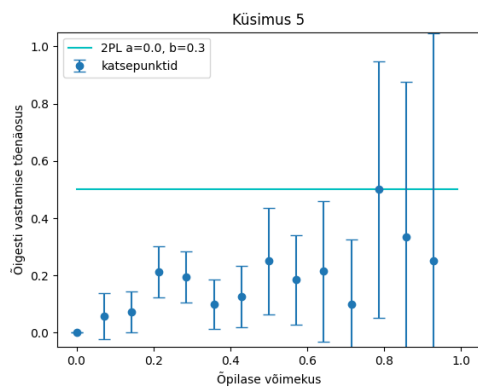
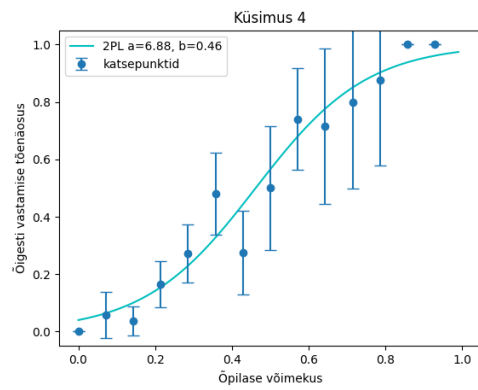
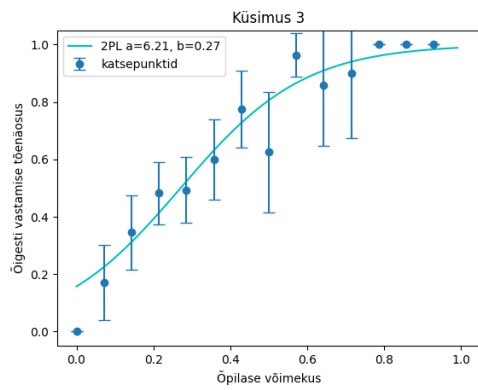
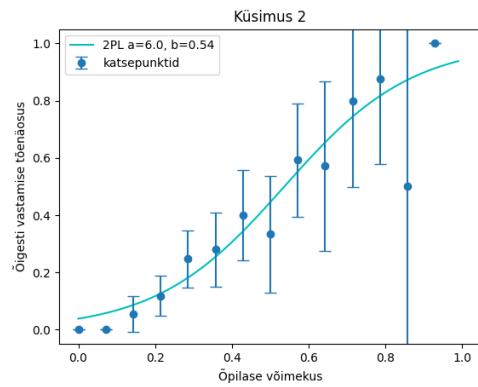
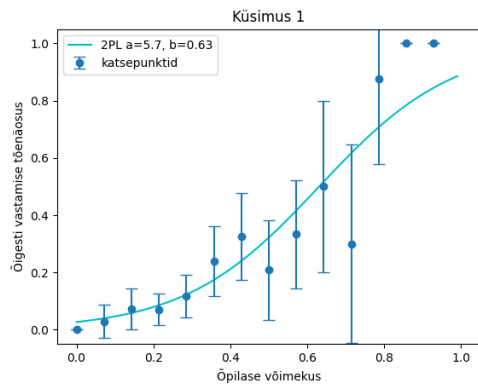
# Lisa 15. 1PL mudel järeltest $a=7.6$ küsimused 1–7



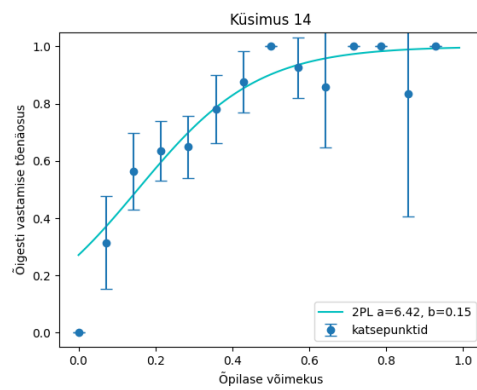
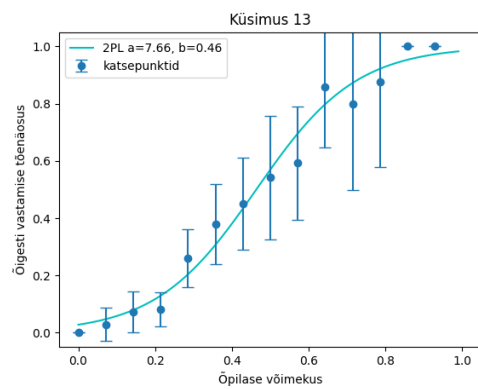
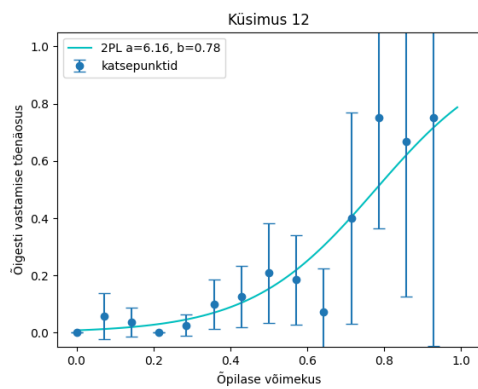
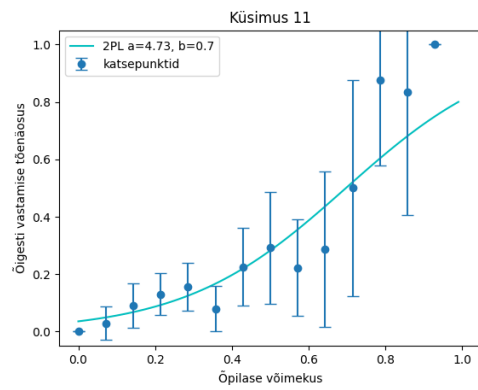
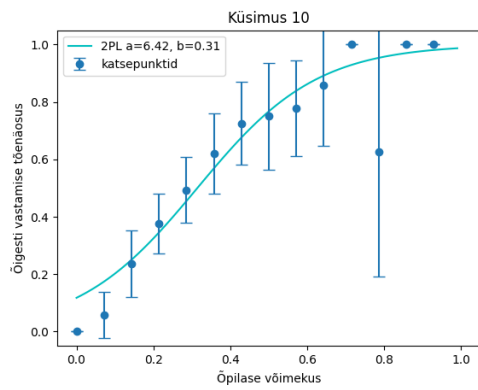
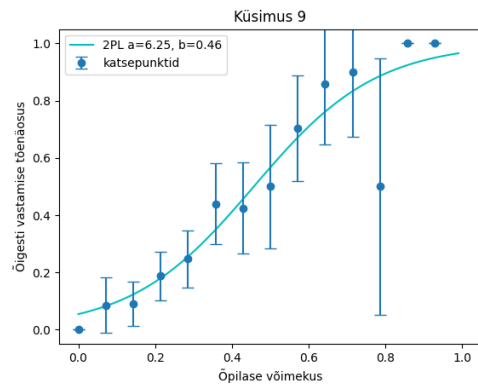
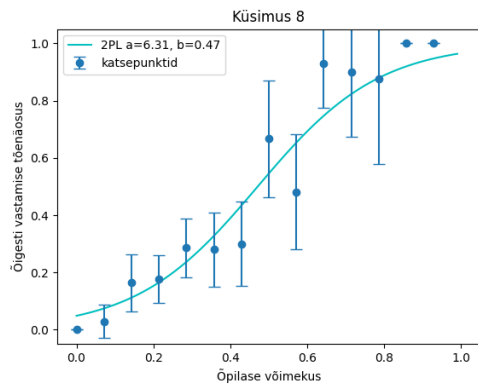
# Lisa 16. 1PL mudel järeltest $a=7.6$ küsimused 8–14



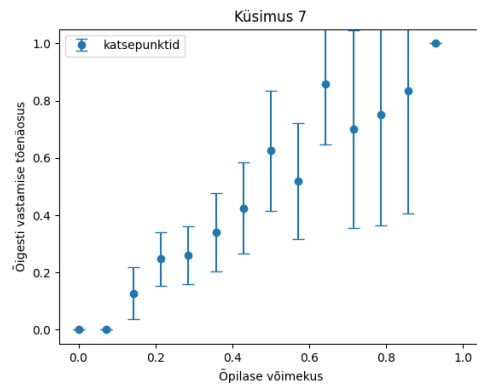
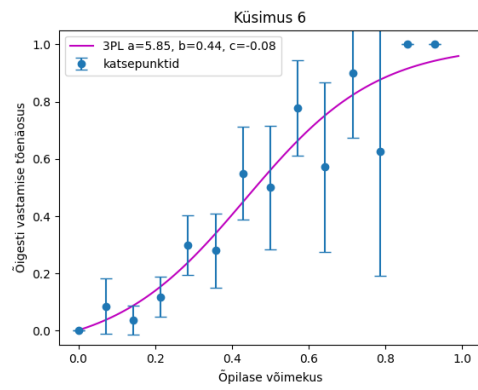
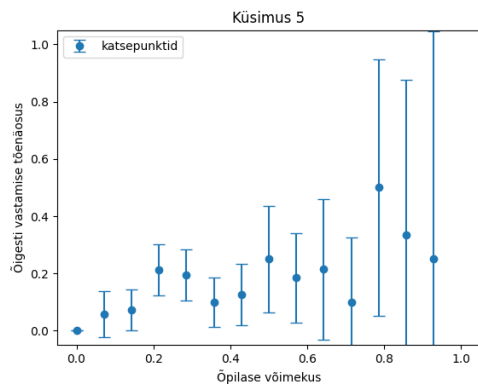
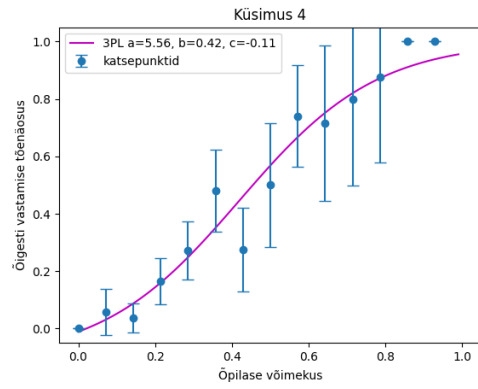
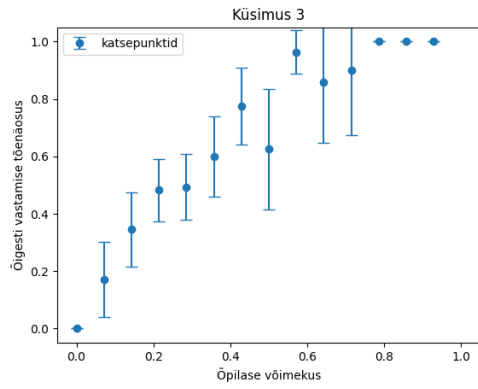
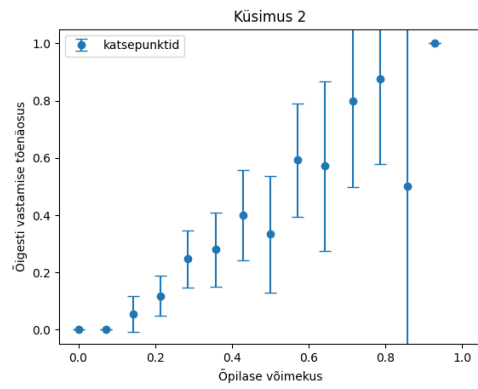
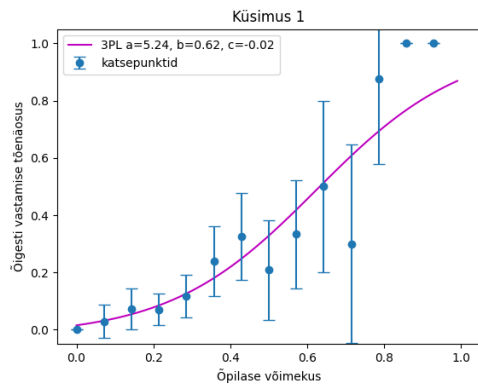
# Lisa 17. 2PL mudel järeldest küsimused 1–7



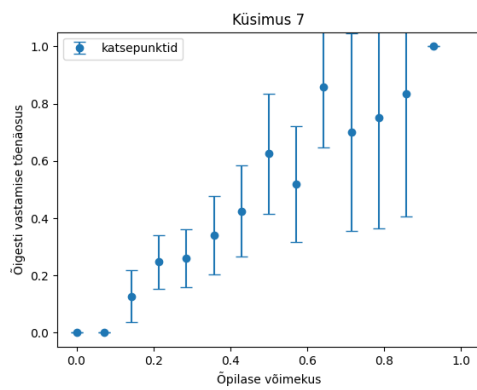
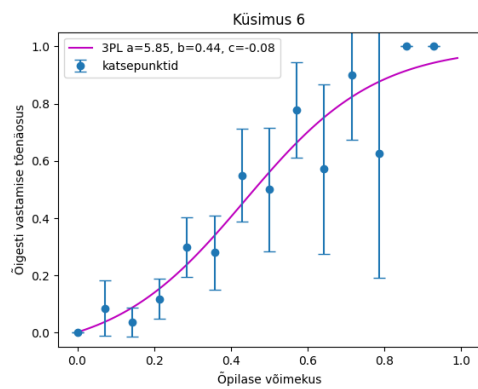
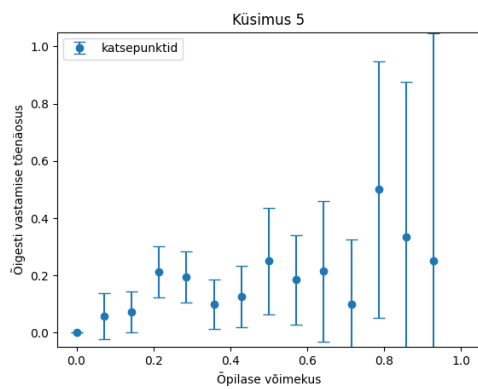
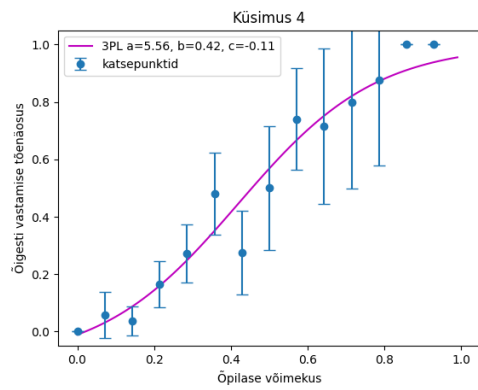
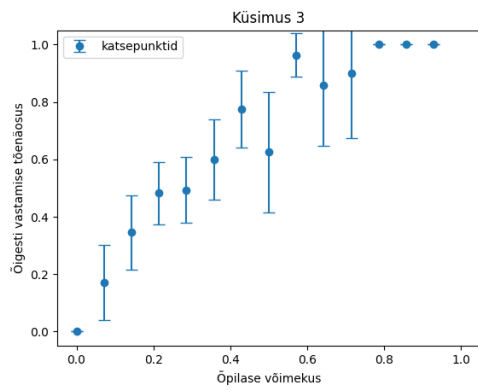
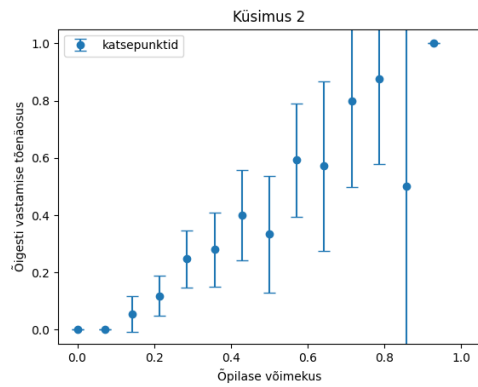
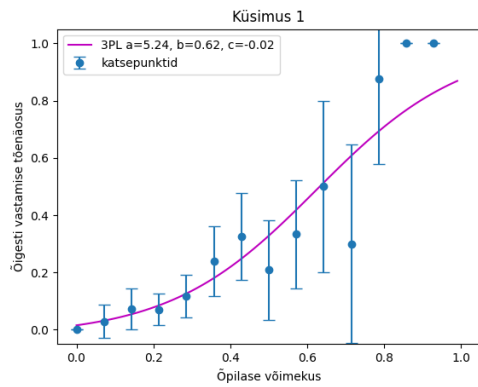
# Lisa 18. 2PL mudel järeltest küsimused 8–14



# Lisa 19. 3PL mudel järeltest küsimused 1–7



# Lisa 20. 3PL mudel järeltest küsimused 8–14



# **Lihtlitsents lõputöö reprodutseerimiseks ja lõputöö üldsusele kättesaadavaks tegemiseks**

Meie, Valle Morel ja Agnes Vask,

1. anname Tartu Ülikoolile tasuta loa (lihtlitsentsi) enda loodud teose

## **Üksikvastuste teooria rakendamine jõu mõistelise testi analüüsimisel,**

mille juhendajad on vanemteadur Kaido Reivelt ja lektor Ingrid Koni, reprodutseerimiseks eesmärgiga seda säilitada, sealhulgas lisada digitaalarhiivi DSpace kuni autoriõiguse kehtivuse lõppemiseni.

2. Anname Tartu Ülikoolile loa teha punktis 1 nimetatud teos üldsusele kättesaadavaks Tartu Ülikooli veebikeskkonna, sealhulgas digitaalarhiivi DSpace kaudu Creative Commons'i litsentsiga CC BY NC ND 4.0, mis lubab autoritele viidates teost reprodutseerida, levitada ja üldsusele suunata ning keelab luua tuletatud teost ja kasutada teost ärieesmärgil, kuni autoriõiguse kehtivuse lõppemiseni.
3. Oleme teadlikud, et punktides 1 ja 2 nimetatud õigused jäävad alles ka autoritele.
4. Kinnitame, et lihtlitsentsi andmisega ei riku me teiste isikute intellektuaalomandi ega isikuandmete kaitse õigusaktidest tulenevaid õigusi.

Valle Morel ja Agnes Vask

25. mai 2023. a.