Vidurinio ugdymo mokinių pasiekimų vertinimo ir panaudojimo scenarijų analizė

Vidurinio ugdymo mokinių pasiekimų vertinimo ir panaudojimo scenarijų analizė

NACIONALINĖ ŠVIETIMO AGENTŪRA

Vilnius, 2021

Autoriai:
Indrė Pusevaitė
Gintautas Jakštas,
iki 2020 m. gruodžio 15 d.

2021 m. balandis
Projektas „Bendrojo ugdymo tyrimų, vertinimo ir stebėsenos sistemos plėtra: mokinių pasiekimų vertinimas“
Nr. 09.2.1-ESFA-V-706-02-001/F3-155

Redaktorė Vilma Pavelčikaitė-Bialoglovienė
Viršelis Dalės Dubonienės
Maketavo Valdas Daraškevičius

Turinys

Įvadas 4
1. Tarpinių pasiekimų vertinimo organizavimo modelių kaštai ir efektyvumas 5
1.1. Modeliuojamos kaštų dedamosios 5
1.2. Kaštų dedamųjų įverčiai 6
1.3. Scenarijų kaštų įverčiai 7
„Modelis 1.“: kai mokinių dalies pasiekimų vertinimas vyksta 2–3 / 3–4 kartus viename ugdymo koncentre elektroniniu būdu ir mokinių pateikti atsakymai vertinami automatiškai, kai kaupiami ir saugojami 1–2 / 3–4 / 12 mokomųjų dalykų įvertinimai. 7
„Modelis 2.“: kai mokinių dalies pasiekimų vertinimas vyksta 2–3 / 3–4 kartus viename ugdymo koncentre elektroniniu būdu ir visų mokinių pateikti atsakymai vertinami elektroninėje aplinkoje, vertintojui priskiriant įvertinimus, kai kaupiami ir saugojami 1–2 / 3–4 / 12 mokomųjų dalykų įvertinimai. 8
„Modelis 3.“: kai mokinių dalies pasiekimų vertinimas vyksta 2–3 / 3–4 kartus viename ugdymo koncentre elektroniniu būdu ir atrinktų vertinimo moderavimui (nuo 10 iki 30 proc.) mokinių pateikti atsakymai vertinami elektroninėje aplinkoje, vertintojui priskiriant įvertinimus, kai kaupiami ir saugojami 1–2 / 3–4 / 12 mokomųjų dalykų įvertinimai. 8
„Modelis 4.“: kai mokinių dalies pasiekimų vertinimas vyksta 2–3 / 3–4 kartus viename ugdymo koncentre žodžiu arba vykdant projektinę veiklą ir visų mokinių darbai vertinami elektroninėje aplinkoje, vertintojui priskiriant įvertinimus, kai kaupiami ir saugojami 1–2 / 3–4 / 12 mokomųjų dalykų įvertinimai..... 9
„Modelis 5.“: kai mokinių dalies pasiekimų vertinimas vyksta 2–3 / 3–4 kartus viename ugdymo koncentre žodžiu arba vykdant projektinę veiklą ir atrinktų vertinimo moderavimui (nuo 10 iki 30 proc.) mokinių pateikti darbai vertinami elektroninėje aplinkoje, vertintojui priskiriant įvertinimus, kai kaupiami ir saugojami 1–2 / 3–4 / 12 mokomųjų dalykų įvertinimai..... 10
„Modelis 6.“: kai mokinių dalies pasiekimų vertinimas vyksta 2–3 / 3–4 kartus viename ugdymo koncentre mišriu būdu (elektroniniu būdu / žodžiu arba vykdant projektinę veiklą ir vertinami automatiškai / elektroninėje aplinkoje, vertintojui priskiriant įvertinimą ir atrinktų vertinimo moderavimui (nuo 10 iki 30 proc.) mokinių pateikti darbai vertinami elektroninėje aplinkoje, vertintojui priskiriant įvertinimus, kai kaupiami ir saugojami 1–2 / 3–4 / 12 mokomųjų dalykų įvertinimai. 11
2. Tarpinių pasiekimų patikrinimų poveikis mokinių pasiekimams 12
2.1. Ankstesni tyrimai ir analitinis modelis 12
2.2. Duomenys ir metodai 13
2.3. Rezultatai 14
Matematikos rezultatai 16
Lietuvių kalbos ir literatūros rezultatai 17
2.4. Vertinimo apribojimai ir gairės tolesniems tyrimams 18
3. Tarpinių pasiekimų vertinimų panaudojimo modeliai 20
3.1. Panaudojimo modelių vertinimo kriterijai 20
3.2. Panaudojimo modelių vertinimo rezultatai 21
1 modelis: nacionalinių tarpinių patikrinimų rezultatų panaudojimas 22
2 modelis: VBE perlaikymas 26
3 modelis: metinių pažymių panaudojimas 28
4. Išvados ir rekomendacijos 30
Literatūra 33
Priedai 34
Mokinių ir mokyklų charakteristikų įtaka pirmojo tarpinio matematikos pasiekimų patikrinimo rezultatams 34
Mokinių ir mokyklų charakteristikų įtaka tikimybei dalyvauti trijuose tarpiniuose matematikos rezultatų patikrinimuose (patekti į tikslinę grupę) 35
Pirmosios matematikos patikrinimų bangos poveikio antrosios bangos pasiekimams vertinimo rezultatai 36

Įvadas

Šios analizės **tikslas** – išanalizuoti alternatyvius tarpinių mokinių pasiekimų patikrinimų organizavimo ir jų rezultatų (ne)panaudojimo baigiamųjų įvertinimų formavimui (sertifikavimui) modelius bei pateikti rekomendacijas dėl tinkamiausio modelio taikymo. Sąvoka „baigiamieji įvertinimai“ vartojama plačiai – tai įvertinimai, kuriais brandos atestate yra arba gali būti apibendrinami mokinių akademiniai pasiekimai vidurinio ugdymo proceso pabaigoje. Priklausomai nuo aptariamo modelio, ši sąvoka apima ir ataskaitos rengimo metu Lietuvoje taikomus įvertinimus (kaip valstybinių brandos egzaminų rezultatai), ir galimus išvestinius įverčius.

Analizės rezultatais siekiama prisidėti prie Nacionalinės švietimo agentūros įgyvendinamo projekto „Bendrojo ugdymo tyrimų, vertinimo ir stebėsenos sistemos plėtra: mokinių pasiekimų vertinimas“ įgyvendinimo. Šio projekto tikslas – tobulinti nacionalinę mokinių pasiekimų vertinimo sistemą, sudarant sąlygas kaupiamojo vertinimo diegimui bei individualios mokinių pasiekimų pažangos stebėsenai.

Analizės **uždaviniai**:

1) įvertinti alternatyvių tarpinių pasiekimų patikrinimų *organizavimo* scenarijų kaštus bei efektyvumą,

2) įvertinti tarpinių pasiekimų patikrinimų *poveikį mokinių pasiekimams*,

3) įvertinti alternatyvių tarpinių pasiekimų patikrinimų (ne)panaudojimo modelių naudas bei rizikas,

4) remiantis analizės rezultatais pateikti rekomendacijas dėl tinkamiausio tarpinių patikrinimų organizavimo ir jų rezultatų panaudojimo modelio.

Analizės struktūrą sudaro keturi skyriai. **Pirmame skyriuje** analizuojami tarpinių pasiekimų vertinimo organizavimo modelių kaštai ir efektyvumas. Pirmiausia pristatomos prielaidos apie vertinimo kaštų dedamąsias ir jų įverčius. Tuomet pateikiami šešių pasirinktų pasiekimų vertinimo modelių įverčiai. Atsižvelgiant į galimų mokinių pasiekimų vertinimo modelių gausą ir galimą kaštų dedamųjų bei įverčių kaitą, sukurta universali modelio kaštų skaičiuoklė. **Antrame skyriuje** vertinamas tarpinių pasiekimų patikrinimų poveikis mokinių pasiekimams. Analizė atlikta remiantis 2020 m. rugsėjo–2021 m. kovo mėnesiais vykusio tarpinių pasiekimų vertinimo sistemos išbandymo duomenimis. **Trečias skyrius** skirtas tarpinių patikrinimų rezultatų (ne)panaudojimo modelių analizei. Analizuojami trys modeliai:

1) nacionalinių tarpinių patikrinimų[1] rezultatų panaudojimo,

2) metinių pažymių panaudojimo,

3) VBE perlaikymo.

Šiame skyriuje aptariamos su kiekvienu modeliu siejamos naudos (lyginant su *status quo*), prielaidos ir priemonės naudai gauti, rizikos (lyginant su *status quo*), rizikų vertinimo ir valdymo būdai. Ketvirtame skyriuje pateikiamos analizės rezultatais grįstos išvados ir rekomendacijos.

[1] Patikrinimų sistema Lietuvoje išbandyta 2020 m. rugsėjo–2021 m. kovo mėnesiais.

1. Tarpinių pasiekimų vertinimo organizavimo modelių kaštai ir efektyvumas

Šiame skyriuje vertinami skirtingų mokinių pasiekimų vertinimo modelių kaštai ir efektyvumas. Pirmiausia pristatomos kaštų dedamosios ir jų įverčiai, apskaičiuoti remiantis 2020 m. lapkričio 16 d. duomenimis. Atsižvelgiant į galimų mokinių pasiekimų vertinimo modelių gausą ir galimą kaštų dedamųjų bei įverčių kaitą, sukurta universali modelio kaštų skaičiuoklė „Excel“ programoje.

1.1. Modeliuojamos kaštų dedamosios

1. Kai testuojama elektroniniu būdu, testavimo kaštai apima:
    a. e. testavimo aplinka (paslaugų sutartis, metinė paslaugų kaina);
    b. techninis palaikymas testavimų metu (paslaugų sutartis, kaina už 1 testavimą);
    c. testavimas (paslaugų sutartis, priklauso nuo testuojamų mokinių skaičiaus);
    d. testavimų administravimas Lietuvoje (vienas testavimas);
    e. užduočių turinio parengimas vienam vertinimui elektroniniu būdu;
    f. rezultatų parengimas ir perdavimas mokykloms;
    g. galimi, bet neprivalomi kaštai (galimybė pasirinkti):
        I. vieno mokinio darbo vertinimas, kai vertinimas atliekamas elektroninėje aplinkoje;
        II. moderatoriaus vieno mokinio darbo patikrinimas, kai vertinimas atliekamas elektroninėje aplinkoje;
        III. rezultatų parengimas ir perdavimas mokykloms, kai vertinama elektroninėje aplinkoje vertintojui priskiriant balą;
        IV. testavimo rezultatų saugojimo kaštai, kai testuojama elektroninėje aplinkoje.
2. Kai testuojama žodžiu, testavimo kaštai apima:
    a. užduočių turinio parengimas vienam vertinimui žodžiu;
    b. testavimo rezultatų ir atlikčių saugojimo kaštai, kai vertinama žodžiu arba vykdant projektinę veiklą;
    c. galimi, bet neprivalomi kaštai (galimybė pasirinkti):
        I. vieno atsiskaitymo žodžiu vertinimas;
        II. vieno atsiskaitymo žodžiu administravimas mokykloje;
        III. jeigu vertinama elektroninėje aplinkoje, tai apima visus elektroninės aplinkos kaštus kaip ir testuojant elektroniniu būdu.
3. Kai vertinama projektinė veikla, vertinimo kaštai apima:
    a. rezultatų ir atlikčių saugojimo kaštai, kai vertinama žodžiu arba vykdant projektinę veiklą;
    b. galimi, bet neprivalomi kaštai (galimybė pasirinkti):
        I. vadovavimas vieno projektinio darbo rengimui mokykloje;
        II. mokinių ir mokytojų konsultavimas nacionaliniu lygmeniu;
        III. vieno projektinio darbo vertinimas;
        IV. jeigu vertinama elektroninėje aplinkoje, tai apima visus elektroninės aplinkos kaštus kaip ir testuojant elektroniniu būdu.
4. Kai vertinama mišriu būdu, vertinimo kaštai apima visus testavimo elektroniniu būdu kaštus ir vertinimo žodžiu kaštus.
5. E. testavimo administravimo mokykloje kaštai nėra laikytini vertinimo kaštų dalimi.[2]

[2] Sprendimas priimtas NŠA siūlymu.

1.2. Kaštų dedamųjų įverčiai

1 lentelė. Atsižvelgiant į 2020 m. lapkričio 16 d. turimą informaciją, skaičiuoklėje naudojamos kaštų reikšmės

	Kaštų kategorija	Kaina vienam mokiniui	Pastovi kainos dalis vienam testavimui, nepriklausomai nuo testuojamų mokinių skaičiaus	Pastovi kainos dalis (nepriklauso nuo testuojamų mokinių skaičiaus ir testavimų kiekio)
Elektroninės sistemos kaštai	E. testavimo aplinka (paslaugų sutartis programinei įrangai)	- €	- €	30,748 €
	Techninis palaikymas testavimų metu (paslaugų sutartis, kaina už 1 testavimą)	- €	7,800 €	- €
	Testavimas (paslaugų sutartis), kaina, kai testuojama iki 200.000 mokinių	1.37 €	- €	- €
	Testavimas (paslaugų sutartis), kaina, kai testuojama nuo 200.000 iki 300.000 mokinių	1.17 €	- €	- €
	Testavimas (paslaugų sutartis), kaina, kai testuojama daugiau nei 300.000 mokinių	1.00 €	- €	- €
	Testavimų administravimas Lietuvoje (vienas testavimas)	- €	13,105 €	- €
Užduotys	Užduočių turinio parengimas vienam vertinimui **elektroniniu** būdu	- €	20,969 €	- €
	Užduočių turinio parengimas vienam vertinimui **žodžiu**	- €	20,969 €	- €
Darbų vertinimas	Vieno mokinio darbo vertinimas, kai vertinimas atliekamas **elektroninėje aplinkoje**	6.81 €	- €	- €
	Moderatoriaus vieno mokinio darbo patikrinimas, kai vertinimas atliekamas **elektroninėje aplinkoje**	6.81 €	- €	- €
	Automatinis darbų vertinimo programavimas vienam testavimui	- €	- €	- €
	Vieno **projektinio** darbo vertinimas	20.00 €	- €	- €
	Vieno atsiskaitymo **žodžiu** vertinimas	6.81 €	- €	- €
Rezultatų perdavimas	Rezultatų parengimas ir perdavimas mokykloms, kai vertinama elektroniniu būdu automatiškai	- €	5,242 €	- €
	Rezultatų parengimas ir perdavimas mokykloms, kai vertinama elektroninėje aplinkoje vertintojui priskiriant balą	- €	5,242 €	- €
Rezultatų saugojimas	Testavimo rezultatų saugojimo kaštai, kai testuojama elektroninėje aplinkoje	0.05 €	- €	5,000 €
	Testavimo rezultatų saugojimo kaštai, kai vertinama žodžiu arba vykdant projektinę veiklą	0.05 €	- €	5,000 €
Kiti kaštai	Vadovavimas vieno projektinio darbo rengimui	25.00 €	- €	- €
	Mokinių, rengiančių projektinius darbus, konsultavimas nacionaliniame lygmenyje	50.00 €	- €	- €
	Vieno atsiskaitymo žodžiu administravimas mokyklose	- €	- €	- €

1.3. Scenarijų kaštų įverčiai

„Modelis 1.“: kai mokinių dalies pasiekimų vertinimas vyksta 2–3 / 3–4 kartus viename ugdymo koncentre elektroniniu būdu ir mokinių pateikti atsakymai vertinami automatiškai, kai kaupiami ir saugojami 1–2 / 3–4 / 12 mokomųjų dalykų įvertinimai.

Modelio prielaidos[3]:

- mokinių skaičius – 30.000 / 20.000;
- vertinami visi mokiniai;
- vertinimas vyksta elektroniniu būdu;
- atsakymai vertinami automatiškai;
- visi testavimo rezultatai kaupiami ir saugomi.

Modelio kaštų kategorijos:

- e. testavimo aplinka (paslaugų sutartis, metinė kaina);
- techninis palaikymas testavimų metu (paslaugų sutartis, kaina už 1 testavimą);
- testavimas (paslaugų sutartis), kaina, kai testuojama iki 200.000 mokinių;
- testavimų administravimas Lietuvoje (vienas testavimas);
- užduočių turinio parengimas vienam vertinimui elektroniniu būdu;
- rezultatų parengimas ir perdavimas mokykloms, kai vertinama elektroniniu būdu automatiškai;
- testavimo rezultatų saugojimo kaštai, kai testuojama elektroninėje aplinkoje.

2 lentelė. Kaštai (Eur), kai 30.000 mokinių testuojami 2–4 kartus, 1, 2, 4 arba 12 mokomųjų dalykų.

Mokomųjų dalykų kiekis / Testavimų kiekis	1	2	4	12
2	213,680	390,112	742,976	2,154,432
3	301,896	566,544	1,095,840	3,213,024
4	390,112	742,976	1,448,704	4,271,616

3 lentelė. Kaštai (Eur), kai 20.000 mokinių testuojami 2–4 kartus, 1, 2, 4 arba 12 mokomųjų dalykų.

Mokomųjų dalykų kiekis / Testavimų kiekis	1	2	4	12
2	185,780	334,812	632,876	1,825,132
3	260,296	483,844	930,940	2,719,324
4	334,812	632,876	1,229,004	3,613,516

Vieno mokomojo dalyko vienas testavimas 10 tūkst. mokinių kainuotų 97 tūkst. Eur. (9,7 Eur vienam mokiniui). Mokinių skaičiui išaugus tris kartus, iki 30 tūkst., kaštai išaugtų iki 125,5 tūkst. Eur (4,18 Eur vienam mokiniui). Didinant mokomųjų dalykų skaičių iki 4 ir testavimų kiekį iki 3, kai testuojami 30 tūkst. mokinių, testavimo kaštai išauga beveik iki 1,1 mln. Eur (3,04 Eur už vieno mokinio vieną testavimą). Kadangi testuojant elektroniniu būdu yra sąlyginai dideli pastovūs kaštai (teoriškai net testuojant 1 mokinį, kaštai siektų beveik 83 tūkst. Eur.), kuo daugiau testavimų atliekama, tuo mažesnės vieno testavimo sąnaudos. Šio modelio efektyvumas auga, augant testavimų ir testuojamųjų skaičiui. Atsižvelgiant į tai, jog testai vertinami automatiškai, mokytojams nereikėtų skirti papildomai laiko vertinti mokinių atliktims, o pats testavimas būtų ugdymo proceso dalis, tad intervencija į ugdymo procesą būtų minimali.

[3] Skaičiuoklės lange „Skaičiuoklė (85.2.1)“ parinktos šio modelio prielaidos.

„Modelis 2.“: kai mokinių dalies pasiekimų vertinimas vyksta 2–3 / 3–4 kartus viename ugdymo koncentre elektroniniu būdu ir visų mokinių pateikti atsakymai vertinami elektroninėje aplinkoje, vertintojui priskiriant įvertinimus, kai kaupiami ir saugojami 1–2 / 3–4 / 12 mokomųjų dalykų įvertinimai.

„Modelis 2.“ nuo „Modelio 1.“ skiriasi tuo, kad šiuo atveju atsakymai nėra automatiškai įvertinami, vertintojai priskiria įvertinimus elektroninėje aplinkoje.

Modelio prielaidos:

- mokinių skaičius – 30.000 / 20.000;
- vertinami visi mokiniai;
- vertinama elektroniniu būdu;
- vertinami elektroninėje aplinkoje, vertintojui priskiriant įvertinimus;
- mokinių darbų vertinimas nelaikomas ugdymo proceso dalimi (jei darbų vertinimas būtų laikomas ugdymo proceso dalimi, tai „Modelio 2.“ kaštai nesiskirtų nuo „Modelio 1.“);
- visi testavimo rezultatai kaupiami ir saugomi.

4 lentelė. Kaštai (Eur), kai 30.000 mokinių testuojami 2–4 kartus, 1, 2, 4 arba 12 mokomųjų dalykų.

Mokomųjų dalykų kiekis / Testavimų kiekis	1	2	4	12
2	622,280	1,207,312	2,377,376	7,057,632
3	914,796	1,792,344	3,547,440	10,567,824
4	1,207,312	2,377,376	4,717,504	14,078,016

5 lentelė. Kaštai (Eur), kai 20.000 mokinių testuojami 2–4 kartus, 1, 2, 4 arba 12 mokomųjų dalykų.

Mokomųjų dalykų kiekis / Testavimų kiekis	1	2	4	12
2	458,180	879,612	1,722,476	5,093,932
3	668,896	1,301,044	2,565,340	7,622,524
4	879,612	1,722,476	3,408,204	10,151,116

Kai mokinių atsakymus vertina vertintojai elektroninėje aplinkoje, vieno darbo vertinimo kaštai siekia 6,81 Eur. Nepriklausomai nuo masiškumo, šie kintami kaštai sudaro didžiąją vertinimo kaštų dalį (nuo 59,5 proc. Iki 69,7 proc.). Augant mokytojų, o kartu ir vertintojų, darbo užmokesčiui, vertinimų ir vertinamų mokinių skaičiui, šio modelio kaštai išaugtų dar labiau. Be to, daugėjant vertinimų, būtų reikalingas didesnis vertintojų įtraukimas. Kadangi vertintojais paprastai dirba mokytojai, tai gali turėti neigiamos įtakos ugdymo procesui. Šio modelio rizika yra ir galimos žmogiškosios vertintojų klaidos, subjektyvumas.

„Modelis 3.“: kai mokinių dalies pasiekimų vertinimas vyksta 2–3 / 3–4 kartus viename ugdymo koncentre elektroniniu būdu ir atrinkti vertinimo moderavimui (nuo 10 iki 30 proc.) mokinių pateikti atsakymai vertinami elektroninėje aplinkoje, vertintojui priskiriant įvertinimus, kai kaupiami ir saugojami 1–2 / 3–4 / 12 mokomųjų dalykų įvertinimai.

Modelio prielaidos:

- mokinių skaičius – 30.000 / 20.000;
- vertinami visi mokiniai;
- vertinama elektroniniu būdu;
- vertinami elektroninėje aplinkoje, vertintojui priskiriant įvertinimus, o mokinių darbų vertinimas laikomas ugdymo proceso dalimi;
- 30 proc. darbų peržiūrimi vertinimo moderatorių;
- visi testavimo rezultatai kaupiami ir saugomi.

6 lentelė. Kaštai (Eur), kai 30.000 mokinių testuojami 2–4 kartus, 1, 2, 4 arba 12 mokomųjų dalykų.

Mokomųjų dalykų kiekis / Testavimų kiekis	1	2	4	12
2	336,260	635,272	1,233,296	3,625,392
3	485,766	934,284	1,831,320	5,419,464
4	635,272	1,233,296	2,429,344	7,213,536

7 lentelė. Kaštai (Eur), kai 20.000 mokinių testuojami 2–4 kartus, 1, 2, 4 arba 12 mokomųjų dalykų.

Mokomųjų dalykų kiekis / Testavimų kiekis	1	2	4	12
2	267,500	498,252	959,756	2,805,772
3	382,876	729,004	1,421,260	4,190,284
4	498,252	959,756	1,882,764	5,574,796

„Modelis 3.“ nuo „Modelio 2.“ skiriasi tuo, jog 30 proc. darbų yra peržiūrimi vertinimo moderatorių. Tačiau darbus vertinantiems mokytojams nėra papildomai apmokama, tai laikoma ugdymo proceso dalimi. Vertinimo moderatorių kaštai priklauso nuo mokinių skaičiaus: vieno darbo moderavimas – 6,81 Eur. Šis modelis, lyginant su „Modeliu 1.“, suvaldo riziką dėl galimų vertintojų klaidų. Ilgainiui mažėjant netikslių vertinimų skaičiui, peržiūrimų darbų dalis, o kartu ir šio modelio kaštai, galėtų mažėti. Pervertinamų darbų dalį taip pat būtų galima diferencijuoti, atsižvelgiant į istorinį skirtingų mokytojų (arba mokyklų) darbų vertinimo tikslumą.

„Modelis 4.“: kai mokinių dalies pasiekimų vertinimas vyksta 2–3 / 3–4 kartus viename ugdymo koncentre žodžiu arba vykdant projektinę veiklą ir visų mokinių darbai vertinami elektroninėje aplinkoje, vertintojui priskiriant įvertinimus, kai kaupiami ir saugojami 1–2 / 3–4 / 12 mokomųjų dalykų įvertinimai.

Modelio prielaidos:

- mokinių skaičius – 30.000;
- vertinami visi mokiniai;
- vertinama projektinė veikla (jeigu vertinama atsiskaitant žodžiu, tai kaštai nesiskirs nuo „Modelio 1.“);
- vertinami elektroninėje aplinkoje, vertintojui priskiriant įvertinimus, o mokinių darbų vertinimas laikomas ugdymo proceso dalimi;
- įtraukiamas vadovavimas projektinio darbo rengimui (vadovavimas vieno projektinio darbo rengimui – 25 Eur);
- įtraukiamas konsultavimas nacionaliniu mastu, rengiant projektinį darbą (konsultuojamas kiekvienas mokinys vieną kartą, nepriklausomai nuo to, kiek projektų rengia, – 50 Eur);
- visi testavimo rezultatai kaupiami ir saugomi.

8 lentelė. Kaštai (Eur), kai 30.000 mokinių testuojami 2–4 kartus, 1, 2, 4 arba 12 mokomųjų dalykų, konsultavimas nacionaliniu mastu nėra įtraukiamas.

Mokomųjų dalykų kiekis / Testavimų kiekis	1	2	4	12
2	921,742	1,056,236	1,325,224	2,401,176
3	988,989	1,190,730	1,594,212	3,208,140
4	1,056,236	1,325,224	1,863,200	4,015,104

9 lentelė. Kaštai (Eur), kai 30.000 mokinių testuojami 2–4 kartus, 1, 2, 4 arba 12 mokomųjų dalykų, įtrauktas konsultavimas nacionaliniu mastu.

Mokomųjų dalykų kiekis / Testavimų kiekis	1	2	4	12
2	2,421,742	2,556,236	2,825,224	3,901,176
3	2,488,989	2,690,730	3,094,212	4,708,140
4	2,556,236	2,825,224	3,363,200	5,515,104

„Modelis 4.“ yra vienas iš imliausių finansiniams kaštams. Vertinimas žodžiu sutrikdytų ugdymo procesą, nes vertinimas vyktų individualiai, priešingai nei testuojant elektroninėje aplinkoje, vienu metu negalėtų būti testuojami visi mokiniai. Siekiant, kad vertinamas būtų vykdomas elektroninėje aplinkoje, reikalingi kaštai elektroninei sistemai palaikyti. Augant mokinių skaičiui ir mokytojų darbo užmokesčiui, šio testavimo kaštai išaugtų dar labiau.

„Modelis 5.“: kai mokinių dalies pasiekimų vertinimas vyksta 2–3 / 3–4 kartus viename ugdymo koncentre žodžiu arba vykdant projektinę veiklą ir atrinktų vertinimo moderavimui (nuo 10 iki 30 proc.) mokinių pateikti darbai vertinami elektroninėje aplinkoje, vertintojui priskiriant įvertinimus, kai kaupiami ir saugojami 1–2 / 3–4 / 12 mokomųjų dalykų įvertinimai.

Modelio prielaidos[4]:

- mokinių skaičius – 30.000;
- vertinami visi mokiniai;
- vertinama projektinė veikla;
- vertinami elektroninėje aplinkoje, vertintojui priskiriant įvertinimus, o mokinių darbų vertinimas laikomas ugdymo proceso dalimi;
- įtraukiamas vadovavimas projektinio darbo rengimui (vadovavimas vieno projektinio darbo rengimui – 25 Eur);
- 30 proc. darbų peržiūrimi vertinimo moderatorių;
- įtraukiamas konsultavimas nacionaliniu mastu, rengiant projektinį darbą (konsultuojamas kiekvienas mokinys vieną kartą, nepriklausomai nuo to, kiek projektų rengia, – 50 Eur);
- visi testavimo rezultatai kaupiami ir saugomi.

10 lentelė. Kaštai (Eur), kai 30.000 mokinių testuojami 2–4 kartus, 1, 2, 4 arba 12 mokomųjų dalykų, konsultavimas nacionaliniu mastu nėra įtraukiamas.

Mokomųjų dalykų kiekis / Testavimų kiekis	1	2	4	12
2	1,044,322	1,301,396	1,815,544	3,872,136
3	1,172,859	1,558,470	2,329,692	5,414,580
4	1,301,396	1,815,544	2,843,840	6,957,024

11 lentelė. Kaštai (Eur), kai 30.000 mokinių testuojami 2–4 kartus, 1, 2, 4 arba 12 mokomųjų dalykų, įtrauktas konsultavimas nacionaliniu mastu.

Mokomųjų dalykų kiekis / Testavimų kiekis	1	2	4	12
2	2,544,322	2,801,396	3,315,544	5,372,136
3	2,672,859	3,058,470	3,829,692	6,914,580
4	2,801,396	3,315,544	4,343,840	8,457,024

„Modelis 5.“, lyginant su „Modeliu 4.“, skiriasi tuo, jog papildomai reikalingi kaštai mokinių konsultavimui.

[4] Skaičiuoklės lange „Skaičiuoklė (85.2.5)“ pasirinktos šio modelio prielaidos, suteikiama galimybė keisti pasirinkimus.

„Modelis 6.“: kai mokinių dalies pasiekimų vertinimas vyksta 2–3 / 3–4 kartus viename ugdymo koncentre mišriu būdu (elektroniniu būdu / žodžiu arba vykdant projektinę veiklą ir vertinami automatiškai / elektroninėje aplinkoje, vertintojui priskiriant įvertinimą ir atrinktų vertinimo moderavimui (nuo 10 iki 30 proc.) mokinių pateikti darbai vertinami elektroninėje aplinkoje, vertintojui priskiriant įvertinimus, kai kaupiami ir saugojami 1–2 / 3–4 / 12 mokomųjų dalykų įvertinimai.

Modelio prielaidos:

- mokinių skaičius – 30.000 / 20.000;
- vertinami visi mokiniai;
- 30 proc. darbų peržiūrimi vertinimo moderatorių;
- atsiskaitoma žodžiu ir vertinama elektroninėje aplinkoje;
- vertinami elektroninėje aplinkoje, vertintojui priskiriant įvertinimus, o mokinių darbų vertinimas laikomas ugdymo proceso dalimi (pateikiama ir alternatyva, kai vertinimas nėra laikomas ugdymo proceso dalimi);
- visi vertinimo rezultatai kaupiami ir saugomi.

12 lentelė. Kaštai (Eur), kai 30.000 mokinių testuojami 2–4 kartus, 1, 2, 4 arba 12 mokomųjų dalykų.

Mokomųjų dalykų kiekis / Testavimų kiekis	1	2	4	12
2	384,698	725,648	1,407,548	4,135,148
3	555,173	1,066,598	2,089,448	6,180,848
4	725,648	1,407,548	2,771,348	8,226,548

13 lentelė. Kaštai (Eur), kai 20.000 mokinių testuojami 2–4 kartus, 1, 2, 4 arba 12 mokomųjų dalykų.

Mokomųjų dalykų kiekis / Testavimų kiekis	1	2	4	12
2	315,438	588,128	1,133,508	3,315,028
3	451,783	860,818	1,678,888	4,951,168
4	588,128	1,133,508	2,224,268	6,587,308

14 lentelė. Kaštai (Eur), kai 30.000 mokinių testuojami 2–4 kartus, 1, 2, 4 arba 12 mokomųjų dalykų, tačiau vertinimas nėra laikomas ugdymo proceso dalimi

Mokomųjų dalykų kiekis / Testavimų kiekis	1	2	4	12
2	997,598	1,747,148	3,246,248	9,242,648
3	1,372,373	2,496,698	4,745,348	13,739,948
4	1,747,148	3,246,248	6,244,448	18,237,248

Vertinimas mišriu būdu yra pats brangiausias ir reikalautų didžiausių ugdymo proceso korekcijų.

Apžvelgus visus modelius, atsižvelgiant į įtaką ugdymo procesui, kaštams, rezultatų patikimumą, rekomenduojama pradėti nuo „Modelio 1“, nes šis modelis turėtų mažiausius poreikius ugdymo proceso korekcijoms, mažiausius kaštus, didžiausią rezultatų objektyvumą. Siekiant plėsti tarpiškai testuojamų dalykų skaičių, šis modelis taptų vis efektyvesnis kaštų prasme. Sėkmingai įdiegus šį modelį, rekomenduojame įsivertinti galimybę pereiti prie mišraus modelio tam, jog vertinimas galėtų apimti platesnį vertinamų kompetencijų kiekį, prijungiant vertinimą žodžiu ir projektinių darbų rengimą.

2. Tarpinių pasiekimų patikrinimų poveikis mokinių pasiekimams

Šiame skyriuje vertinamas 2020 m. rugsėjo–2021 m. kovo mėnesiais vykdytų bandomųjų lietuvių kalbos ir literatūros ir matematikos tarpinių patikrinimų poveikis mokinių pasiekimams. Pirmiausia pristatomas analitinis modelis – ankstesni tyrimai, iš kurių kildinamos skyriuje tikrinamos hipotezės. Tuomet aptariami poveikio vertinimui naudojami duomenys ir metodai. Galiausiai pristatomi vertinimo rezultatai ir klausimai tolesniems tyrimams.

2.1. Ankstesni tyrimai ir analitinis modelis

Įvairūs tyrimai rodo, kad tarpinis vertinimas turi teigiamą poveikį mokinių pasiekimams, ypač – kai po jo teikiamas grįžtamasis ryšys. Poveikio mechanizmai skirstomi į tiesioginius ir netiesioginius (angl. *mediation effects*). Tiesioginis poveikis – poveikis, kuris pasireiškia dėl paties testo laikymo proceso. Pavyzdžiui, pastebima, kad poreikis prisiminti ir panaudoti žinias testo laikymo metu gerina ilgalaikį informacijos įsiminimą.[5] Netiesioginis poveikis pasireiškia, nes vertinimas keičia mokinių ir mokytojų elgesį ugdymo procese. Pavyzdžiui, nustatyta, kad dažnesnis tarpinis vertinimas didina susikaupimą pamokose[6], skatina nuosekliai mokytis viso kurso metu.[7] Kitas svarbus netiesioginio poveikio mechanizmas – testų rezultatai informuoja mokinį ir mokytoją apie esamą žinių ir gebėjimų lygį, o tai leidžia tobulinti ugdymo(si) procesą. Pavyzdžiui, gauti duomenys gali būti naudojami formuojamajam vertinimui vykdyti, padėti nukreipti mokymo(si) pastangas į silpnąsias mokinių pasiekimų sritis, parinkti mokinio gebėjimų lygį geriau atitinkantį ugdymo turinį[8]. Dėl šių ir kitų priežasčių vyresnių klasių mokiniai viso ugdymo proceso metu yra vertinami mokytojų.

Galima tikėtis, kad dažnesnis tarpinis nacionalinis pasiekimų patikrinimas turėtų papildomą teigiamą poveikį mokinių egzaminų rezultatams vidurinio ugdymo pakopos pabaigoje. Tam yra dvi priežastys. Pirma, pažymiai, gaunami ugdymo procese, moksleivius ir jų šeimas apie jų pasiekimus nacionaliniu lygiu dažnai informuoja netiksliai. Viena vertus, taip gali nutikti todėl, kad veikdami tik savo mokyklos ribose ir vadovaudamiesi šiuo metu esamais dalykų pasiekimų aprašais, mokytojai ne visuomet yra pajėgūs objektyviai įvertinti mokinių pasiekimus nacionaliniu mastu. Kita vertus, jie gali būti nesuinteresuoti vertinti objektyviai (pavyzdžiui, dėl poreikio konkuruoti su kitais mokytojais ir mokyklomis).[9] Netiksli informacija riboja jos panaudojamumą ugdymo(si) proceso tobulinimui, o drauge – įtaką mokinių dalykų turinio žinioms ir gebėjimams. Antra, egzaminų rezultatai priklauso ne tik nuo dalyko žinių, tačiau ir nuo bendrųjų testo laikymui reikalingų gebėjimų, kaip gebėjimas suprasti testo struktūrą ir tikslą. Todėl pastebima, kad moksleiviai, kurie bandydavo laikyti tam tikro tipo egzaminus, juos laikydami pasiekia aukštesnių rezultatų.[10] Kitaip tariant, jei nacionaliniai tarpiniai patikrinimai ir baigiamieji egzaminai pasižymi panašiais tikslais ir struktūra, tarpinis pasiekimų patikrinimas gali ne tik padidinti mokinių žinias ir gebėjimus, tačiau ir suteikti geresnių įgūdžių juos pademonstruoti vėlesnių egzaminų metu. Remiantis šiais argumentais, keliama pirmoji hipotezė.

H1. Vidutiniškai vertinimas turi teigiamą poveikį matuojamiems mokinių rezultatams 12 klasėje.

Ankstesni tyrimai taip pat rodo, kad testų poveikis gali priklausyti nuo jų turinio charakteristikų, testą laikančių moksleivių savybių ir jų tarpusavio sąveikos. Vis dėlto šių tyrimų išvados prieštaringos. Dalyje tyrimų pastebima, sudėtingi nacionaliniai patikrinimai gali demotyvuoti labiau pažeidžiamus moksleivius – kilusius iš žemų pajamų šeimų, priklausančius tautinėms mažumoms, prasčiau akademiškai pasirengusius – ir ilgainiui turėti neigiamą poveikį jų pasiekimams.[11] Kita vertus, yra priežasčių manyti, kad labiau pažeidžiamiems moksleiviams nacionaliniai tarpiniai patikrinimai gali būti ypač naudingi. Pavyzdžiui, jie galėtų iš dalies kompensuoti galimybių ugdytis už mokyklos ribų stoką. Pastebima, kad Lietuvoje iš aukščiausio socialinio-ekonominio statuso šeimų kilę dvyliktokai korepetitorių paslaugomis naudojosi beveik tris kartus dažniau nei kilusieji

[5] Pavyzdžiui, Roediger III, H.L. and Karpicke, J.D., 2006. The power of testing memory: Basic research and implications for educational practice. *Perspectives on psychological science*, *1*(3), pp.181-210.

[6] Pavyzdžiui, Szpunar, K.K., Moulton, S.T. and Schacter, D.L., 2013. Mind wandering and education: from the classroom to online learning. *Frontiers in psychology*, *4*, p.495.

[7] Pavyzdžiui, Fitch, Mildred L., Arthur J. Drucker, and J. A. Norton Jr. „Frequent testing as a motivating factor in large lecture classes." *Journal of Educational Psychology* 42, no. 1 (1951): 1.

[8] Pavyzdžiui, Baumert, J. and Demmrich, A., 2001. Test motivation in the assessment of student skills: The effects of incentives on motivation and performance. *European Journal of Psychology of Education*, *16*(3), p.441.

[9] Pavyzdžiui, Wikström, C. and Wikström, M., 2005. Grade inflation and school competition: an empirical analysis based on the Swedish upper secondary schools. *Economics of education Review*, *24*(3), pp.309-322

[10] Pavyzdžiui, Kulik, J.A., Kulik, C.L.C. and Bangert, R.L., 1984. Effects of practice on aptitude and achievement test scores. *American Educational Research Journal*, *21*(2), pp.435-447.

[11] Pavyzdžiui, Ou, Dongshu. „To leave or not to leave? A regression discontinuity analysis of the impact of failing the high school exit exam." *Economics of Education Review* 29, no. 2 (2010): 171-186

iš žemiausio statuso šeimų[12]. Pagal naudojimosi korepetitorių paslaugomis masto atotrūkį tarp žemiausio ir aukščiausio SES šeimų – Lietuva užima vieną žemiausių vietų tarp PISA tyrime dalyvavusių šalių: 52 iš 63.[13] Pagrindinė korepetitorių paslaugų naudojimo priežastis – siekis geriau pasirengti egzaminams. Atsižvelgiant į tai, keliamos 2–5 hipotezės.

H2. Vertinimo poveikis priklauso nuo testų turinio (variantų).

H3. Tarpiniai patikrinimai turi didesnį teigiamą poveikį moksleiviams, kilusiems iš žemų pajamų šeimų.

H4. Tarpiniai patikrinimai turi didesnį teigiamą poveikį moksleiviams, iki intervencijos sukaupusiems žemesnes kompetencijas.

H5. Tarpiniai patikrinimai turi didesnį teigiamą poveikį moksleiviams, priklausantiems kalbinėms mažumoms.

2.2. Duomenys ir metodai

Duomenys

Tarpinių patikrinimų poveikio vertinimui naudojami 2020 m. rugsėjo–lapkričio mėnesiais vykdytų bandomųjų lietuvių kalbos ir literatūros testų (dviejų bangų) ir 2020 m. rugsėjo–2021 m. kovo mėnesiais vykdytų matematikos testų (3 bangų) rezultatai.

Pirmojoje matematikos gebėjimų vertinimo bangoje dalyvavo ir testus atliko 2332 moksleiviai, o antrojoje – 1741. Pirmose dviejose testavimo bangose galėjo dalyvauti tik tie moksleiviai, kurių mokyklos buvo pakviestos dalyvauti tyrime remiantis kvotine atranka. Trečiojoje tyrimo bangoje, sudarius galimybes testavime dalyvauti visoms norą pareiškusioms mokykloms, testą laikė 5004 moksleiviai. Pirmojoje lietuvių kalbos ir literatūros gebėjimų vertinimo bangoje dalyvavo ir testus pabaigė 1770 moksleivių, o antrojoje – 1806. Užduočių nebaigusių mokinių rezultatai neanalizuojami.

Vertinimo poveikis matuojamas išskiriant tikslines ir kontrolines grupes. Tikslines grupes sudaro moksleiviai, kurie dalyvavo visose vertinimo bangose, o kontrolinę – prisijungusieji tik su paskutine banga (lietuvių kalbos ir literatūros atveju su antrąja, o matematikos – su trečiąja). Jeigu vertinimas turi teigiamos įtakos mokinių pasiekimams, tai kitoms sąlygoms nekintant tikslinės, daugiau kartų tarpiniuose vertinimuose dalyvavusių mokinių grupės pasiekimai turėtų būti aukštesni.

Vertinant matematikos testų poveikį pasiekimams tikslinę grupę sudaro 1234 moksleiviai, o kontrolinę – 3038 (prieš panašiausių atvejų atranką). Vertinant lietuvių kalbos ir literatūros testų poveikį pasiekimams, tikslinę grupę sudaro 1175 moksleiviai, o kontrolinę 631.

Kintamieji

Testų turinio poveikis matuojamas pasitelkiant ankstesnės vertinimo bangos metu laikyto testo varianto požymį. Pirmosios ir antrosios tarpinių matematikos pasiekimų patikrinimo bangos metu moksleiviai atliko 3 variantų testus, o pirmosios lietuvių kalbos ir literatūros patikrinimų bangos metu – 2 variantų.

Mokinių rezultatai tarpiniuose patikrinimuose matuojami surinktų taškų procentine dalimi nuo visų galimų taškų. Siekiant įsitikinti, kad matavimų rezultatų neiškraipo testų sudėtingumo skirtumai, modeliai sudaryti su nestandartizuotais ir standartizuotais pasiekimų kintamaisiais. Standartizavimas reiškia, kad kiekvieno testo varianto rezultatų skirstinys transformuojamas į Z-normalųjį skirstinį taip, kad visų tam tikro tipo testą laikiusių moksleivių rezultatų vidurkis būtų lygus nuliui, o standartinis nuokrypis – vienetui. Po atliktos transformacijos sudarytų modelių rezultatai nesiskiria, todėl dėl paprastumo pristatomi tik modelių su nestandartizuotais kintamaisiais rezultatai.

Mokinių pasiekimai iki intervencijos sudarant modelius matuojami PUPP rezultatais.[14] Siekiant subalansuoti kontrolinę ir tikslinę grupes, lietuvių kalbos ir literatūros atveju pasiekimai iki intervencijos papildomai matuojami 11 klasės metiniais pažymiais. Metiniai pažymiai konvertuoti, atsižvelgiant į mokomojo dalyko lygį (A arba B) pagal taisykles, taikomas apskaičiuojant priėmimo į aukštąsias mokyklas konkursinius balus.

Mokinių šeimos finansinis statusas nustatomas atsižvelgiant į tai, ar 12 klasėje moksleiviai gavo finansinę arba kitokią socialinę paramą (pvz., nemokamus pietus). Socialinė parama skiriama mokiniams, kurių šeimos nariams tenkančios vidutinės pajamos yra mažesnės nei 1,5 valstybės remiamų pajamų (toliau – VRP, šiuo metu 192 Eur), o kai kuriais savivaldybės nustatytais atvejais – 2 VRP (šiuo metu 256 Eur).

[12] STRATA.

[13] STRATA, *Kaip skirtingo SES moksleiviams sukurti lygias galimybes įgyti aukštąjį išsilavinimą?*, 2020

[14] Taip pat išbandyta alternatyva kompetencijas iki intervencijos matuoti indeksu, sudarytu atsižvelgiant į PUPP pasiekimus 10 klasėje ir metinius pažymius 11 klasėje. Indekso konstravimui taikytas principinių komponenčių analizės metodas. Sprendimo atsisakyta dėl silpnos metinių pažymių koreliacijos su PUPP bei testavimo rezultatais.

Mokinių priklausymas kalbinės mažumos grupei nustatoma pagal tai, kokia kalba moksleivis mokėsi 12 klasėje. Vis dėlto šis požymis neleidžia tiksliai nustatyti mokinio (jo šeimos) gimtosios kalbos.

Metodai

Pagrindiniai metodiniai iššūkiai siekiant įvertinti tarpinių patikrinimų poveikį kyla dėl to, kad sprendimas dalyvauti patikrinime ir prisijungimo laikas (prisijungimas su pirmąja arba antrąja banga) nėra atsitiktiniai (endogeniškumo problema). Tikėtina, kad į testavimą anksčiau, su pirmąja banga, įsitrauks mokyklos ir moksleiviai, kurie dėl savo savybių aukštesnių rezultatų būtų pasiekę bet kuriuo atveju. Tyrime ši metodinė problema sprendžiama derinant panašiausių atvejų analizės (toliau – PSM, angl. *propensity score matching*) ir atvirkštinių intervencijos tikimybei svorių priskyrimo (toliau – IPW, angl. *inverse probability weighting*) metodus.

Paruošiamajame žingsnyje, taikant PSM, siekiama įvertinti moksleivių tikimybes dalyvauti abiejuose, o ne tik antrajame, vertinimuose. Tikimybė įsitraukti į pirmąją vertinimo bangą panaudojama siekiant sudaryti tikslinę ir kontrolinę (arba kontrolines) grupes, kurios nesiskiria pagal kitas aktualias – turinčias įtakos tikimybei dalyvauti vertinime ir pasiekimams – savybes. Savybės, kurių aktualumas tikrinamas: mokyklos vietovės tipas (miestas arba kaimas), mokyklos ir moksleivio mokomoji kalba, akademinių pasiekimų lygis iki intervencijos, finansinis statusas, moksleivių lytis. Antruoju žingsniu vertinamas tarpinių pasiekimų patikrinimų poveikis: tikslinės ir kontrolinės grupės akademinių pasiekimų skirtumai vidurinio ugdymo pakopos pabaigoje.

Poveikio vertinimui taikomas vidutinio intervencijos poveikio (angl. *average treatment effect*) įvertis. Tai statistika, skirta įvertinti tikėtiną intervencijos (pasiekimų vertinimo) poveikį visai moksleivių populiacijai, o ne tik eksperimente (viename iš tarpinių pasiekimų vertinimų) dalyvavusiems moksleiviams.

2.3. Rezultatai

Abiejuose tarpiniuose pasiekimų patikrinimuose dalyvavę moksleiviai (tikslinė grupė) pagal daugelį savybių skyrėsi nuo dalyvavusių tik viename patikrinime (kontrolinės grupės). Pavyzdžiui, tikslinė matematikos patikrinimo grupė pasižymėjo aukštesniais PUPP rezultatais ir metiniais pažymiais. Be to, mokėsi mokyklose, kurių mokiniai istoriškai nacionaliniuose egzaminuose pasirodydavo geriau, t. y. pasiekdavo aukštesnių VBE rezultatų. Šios savybės ne tik reikšmingai didino tikimybę dalyvauti abiejose patikrinimų bangose, tačiau ir turėjo teigiamos įtakos tarpinių patikrinimų rezultatams. Pavyzdžiui, vienu balu aukštesni matematikos PUPP pasiekimai tikimybę dalyvauti visuose trijuose tarpiniuose matematikos žinių patikrinimuose (o ne viename jų) didino vidutiniškai 8 procentais. Tuo pat metu 1 balu aukštesni matematikos PUPP pasiekimai pirmojo tarpinio matematikos pasiekimų patikrinimo rezultatus didino 4,4 proc. punktais.

Pritaikius PSM ir IPW metodus suformuotos tikslinė ir kontrolinė tarpinių pasiekimų patikrinimų dalyvių grupės, kurių moksleiviai pagal poveikio vertinimui aktualias charakteristikas tarpusavyje statistiškai reikšmingai nesiskiria. Poveikio vertinimui aktualiomis charakteristikomis laikomos charakteristikos, kurios turi tiesioginės įtakos ir tikimybei patekti į tikslinę intervencijos grupę (dalyvauti visuose tarpiniuose patikrinimuose), ir pasiekimų patikrinimų rezultatams[15].

[15] Mokinių ir mokyklų charakteristikų įtakos tikimybei dalyvauti trijuose tarpiniuose matematikos rezultatų patikrinimuose (patekti į tikslinę grupę) ir patikrinimo rezultatams modeliai pateikiami pirmame ir antrame prieduose.

13 lentelė. Matematikos tarpinių pasiekimų vertinimo dalyvių charakteristikos

	Grupių statistikos po panašiausių atvejų atrankos ir svėrimo			(4) Kontrolinė grupė prieš panašiausių atvejų atranką ir svėrimą
	(1) Tikslinė grupė	(2) Kontrolinė grupė	(3) Vidurkių skirtumo statistinis reikšmingumas, KS testo P reikšmė	
Gimnazijų moksleivių dalis	99 %	99 %	1	99 %
Vidutinė finansinę paramą gaunančių moksleivių dalis mokykloje	14 %	15 %	0.232	13 %
Mieste esančių mokyklų dalis	70 %	70 %	1	79 %
Mokyklų, kuriose mokymas vyksta tik lietuvių kalba, dalis	88 %	93 %	0.78	97 %
Vidutiniai mokyklos absolventų lietuvių kalbos ir literatūros VBE rezultatai ankstesniais metais	41.362	41.222	0.558	40.43
Vidutiniai mokyklos absolventų matematikos VBE rezultatai ankstesniais metais	28.112	27.363	0.518	26.138
Matematikos PUPP pasiekimų vidurkis	7.141	6.922	0.318	6.588
Lietuvių kalbos ir literatūros PUPP pasiekimų vidurkis	8.055	7.974	0.329	7.897
Matematikos balų vidurkis 11 kl. (konvertuotas atsižvelgiant į mokymo lygį)	4.397	4.293	0.422	4.133
Moteriškos lyties moksleivių dalis	54 %	56 %	0.671	58 %
Finansinę paramą gaunančių moksleivių dalis	13 %	15 %	1	13 %
Ne lietuvių kalba besimokančių moksleivių dalis	12 %	7 %	0.837	3 %

14 lentelė. Lietuvių kalbos ir literatūros tarpinių pasiekimų vertinimo dalyvių charakteristikos

	Grupių statistikos po panašiausių atvejų atrankos ir svėrimo			(4) Kontrolinė grupė prieš panašiausių atvejų atranką ir svėrimą
	(1) Tikslinė grupė	(2) Kontrolinė grupė	(3) Vidurkių skirtumo statistinis reikšmingumas, KS testo P reikšmė	
Gimnazijų moksleivių dalis	98 %	99 %	1	100 %
Vidutinė finansinę paramą gaunančių moksleivių dalis mokykloje	15 %	16 %	0.021	14 %
Mieste esančių mokyklų dalis	64 %	66 %	0.995	72 %
Mokyklų, kuriose mokymas vyksta tik lietuvių kalba, dalis	88 %	95 %	0.088	99 %
Vidutiniai mokyklos absolventų lietuvių kalbos ir literatūros VBE rezultatai ankstesniais metais	41.038	41.257	0.243	41.905
Vidutiniai mokyklos absolventų matematikos VBE rezultatai ankstesniais metais	27.477	27.568	0.241	27.228
Matematikos PUPP pasiekimų vidurkis	6.647	6.612	0.371	6.567
Lietuvių kalbos PUPP pasiekimų vidurkis	7.647	7.671	0.945	7.713
Lietuvių kalbos balų vidurkis 11 kl. (konvertuotas atsižvelgiant į mokymo lygį)	4.387	4.368	0.377	4.377
Moteriškos lyties moksleivių dalis	56 %	58 %	0.426	60 %
Finansinę paramą gaunančių moksleivių dalis	14 %	13 %	1	11 %
Ne lietuvių kalba besimokančių moksleivių dalis	11 %	5 %	0.23	1 %

Matematikos rezultatai

Vidutinis dviejų tarpinių matematikos žinių patikrinimų poveikis mokinių rezultatams siekė -1,8 procentinio punkto ir buvo statistiškai nereikšmingas. Poveikio mastas statistiškai reikšmingai nesiskyrė, nepriklausomai nuo antroje bangoje laikyto testo varianto (15 lentelė). Statistiškai reikšmingos įtakos pasiekimams neturėjo ir dviejose ankstesnėse pasiekimų vertinimo bangose laikytų testų variantų kombinacija. **Taigi galima teigti, kad bandomojo tarpinių patikrinimų priemonės išbandymo metu bandytų testų turinys neturėjo ilgalaikės (stabilios) įtakos patikrinimų poveikio mastui.**

Vis dėlto tarpinių patikrinimų priemonės išbandymo metu pastebėta silpnesnių testų turinio svarbos signalų. Pirma, pasiekimai antrosios patikrinimų bangos metu statistiškai reikšmingai priklausė nuo pirmosios bangos metu laikyto testo varianto[16]. Moksleiviams, tarpinio patikrinimo metu laikiusiems 1 ir 3 testo variantus, poveikis buvo teigiamas ir statistiškai reikšmingas. Tarpinis patikrinimas jų pasiekimus padidino atitinkamai 3,4 ir 4,3 proc. punkto. Tuo metu moksleiviams, laikiusiems 2 testo variantą, tarpinis patikrinimas neturėjo statistiškai reikšmingo poveikio. Tai potencialiai galima aiškinti tuo, jog šis testas moksleiviams iškėlė ypatingą iššūkį. Šį testo variantą laikiusiems moksleiviams teko per tą patį laiką išspręsti daugiausiai užduočių. Jį laikę moksleiviai vidutiniškai surinko mažiausiai taškų, o surinkti visų galimų taškų nepavyko nė vienam moksleiviui (16 lentelė). Remiantis ankstesniais tyrimais, nacionaliniai patikrinimai, kurie reikalauja reikšmingai aukštesnių, nei moksleivių turimi, gebėjimų, gali paskatinti nusivilti savo jėgomis, demotyvuoti ir ilgainiui turėti neigiamą poveikį mokinių pasiekimams. **Antra, poveikis moksleiviams, kurie abiejų patikrinimų bangų metu laikė pasiekimų ugdymui palankiausius[17] testų variantus, buvo santykinai arti statistinio reikšmingumo (p<0,14) ribos ir siekė 1,8.** Taigi ateityje, siekiant išlaikyti ir tobulinti tarpinių patikrinimų sistemą, jų turinio tobulinimas turėtų išlikti tarp prioritetinių uždavinių.

Tarpinių patikrinimų poveikio mastas skirtingais pasiekimais ir socialinėmis-demografinėmis charakteristikomis iki intervencijos pasižymėjusiems moksleiviams nesiskyrė. Tyrimo imtyje moksleiviams, kurių PUPP įvertinimai buvo vienu balu žemesni, tarpinių pasiekimų patikrinimo poveikis buvo mažesnis 0,58 proc. punkto, o priklausiusiems kalbinėms mažumoms – 2,3 proc. punkto. Vis dėlto šie skirtumai buvo statistiškai nereikšmingi. Taigi, kitoms sąlygoms nekintant, tikėtina, kad tarpinius pasiekimų patikrinimus vykdant nacionaliniu mastu jų poveikis skirtingų moksleivių grupių pasiekimams būtų panašus.

15 lentelė. Tarpinio matematikos pasiekimų vertinimo įtaka mokinių rezultatams: apibendrinimas

Tikrinama hipotezė	Rodiklis		Intervencijos poveikis
H1. **Vidutiniškai** vertinimas turi teigiamą poveikį matuojamiems mokinių rezultatams 12 klasėje	Vidutinis poveikis		-1.8
H2. Vertinimo poveikis priklauso nuo **testų turinio**	Poveikis moksleiviams, **antroje** bangoje laikiusiems testo variantą (lyginant su kontroline grupe):	Testas 1 (J)	-2.4*
		Testas 2 (K)	-1.62
		Testas 3 (V)	-1.40
	1 testo varianto poveikis, lyginant su 2 (beta koeficientų skirtumas)		-0.73
	1 testo varianto poveikis, lyginant su 3 (beta koeficientų skirtumas)		-0.95
	2 testo varianto poveikis, lyginant su 3 (beta koeficientų skirtumas)		-0.22
H3. Tarpiniai patikrinimai turi didesnį teigiamą poveikį **moksleiviams, kilusiems iš žemų pajamų šeimų**	Poveikio skirtumas žemų pajamų šeimų moksleiviams (lyginant su likusiais)		* Duomenys vertinimo metu neprieinami

[16] Pirmosios matematikos patikrinimų bangos poveikio antrosios bangos pasiekimams vertinimo rezultatai pateikiami trečiame priede.
[17] Vertinant pagal poveikio beta koeficientus, pateikiamus 15 lentelėje ir trečiame priede.

Tikrinama hipotezė	Rodiklis		Intervencijos poveikis
H4. Tarpiniai patikrinimai turi didesnį teigiamą poveikį **moksleiviams, iki intervencijos sukaupusiems žemesnes kompetencijas**	Poveikio skirtumas iki intervencijos aukštesnes kompetencijas (1 PUPP balu) sukaupusiems moksleiviams		-0.58
H5. Tarpiniai patikrinimai turi didesnį teigiamą poveikį **moksleiviams, priklausantiems kalbinėms mažumoms**	Poveikio skirtumas ne lietuvių kalba besimokantiems (tautinių mažumų) moksleiviams (lyginant su likusiais)		-2.3

P reikšmės: . - <0.1, * - <0,05,** - <0,01, ***-<0,001. P reikšmė yra tikimybė, kad tikroji statistika (tuo atveju, kai H0 teisinga) yra ne mažesnis už stebimą įvertį.

Duomenys: ŠVIS ir NŠA; skaičiavo autoriai.

16 lentelė. Pirmosios matematikos tarpinių pasiekimų vertinimo bangos rezultatai

Testo variantas	Surinktų taškų dalies vidurkis, proc.	Surinktų taškų sumos vidurkis	Maksimali surinktų taškų suma	Maksimali galimų surinkti taškų suma	Mokinių skaičius
1	43.4	4.3	10	10	780
2	36.6	7.7	19	21	755
3	37.9	5.3	14	14	797

Duomenys: ŠVIS ir NŠA; skaičiavo autoriai.

Lietuvių kalbos ir literatūros rezultatai

Vidutinis tarpinio lietuvių kalbos ir literatūros partikrinimo poveikis mokinių rezultatams siekė -1 proc. punktą ir buvo statistiškai nereikšmingas. Moksleiviams, pirmosios patikrinimų bangos metu laikiusiems pirmą testo variantą, galėjo pasireikšti nedidelis, tačiau statistiškai reikšmingas neigiamas poveikis: jų rezultatai buvo 2,3 proc. punkto žemesni nei pirmojoje patikrinimų bangoje nedalyvavusių moksleivių. Vis dėlto skirtumas tarp abiejų variantų testų poveikių siekia vos 0,9 proc. punkto ir yra statistiškai nereikšmingas. Taigi negalima teigti, kad lietuvių kalbos ir literatūros atveju poveikis reikšmingai priklausė nuo testo varianto (17 lentelė).

Tarpinių pasiekimų vertinimas galėjo būti 4,2 proc. punkto naudingesnis moksleiviams, kilusiems iš žemų pajamų šeimų. Šis skirtumas buvo arti statistinio reikšmingumo ribos (p<0,1). Poveikio skirtumo kryptis moksleiviams, priklausantiems kalbinėms mažumoms, taip pat teigiama, tačiau skirtumas statistiškai nereikšmingas.

Remiantis tik teksto suvokimo užduočių rezultatais, gautos analogiškos išvados: tarpinis lietuvių kalbos ir literatūros pasiekimų patikrinimas neturėjo reikšmingo poveikio mokinių teksto suvokimo rezultatams, o poveikio mastas statistiškai reikšmingai nesiskyrė, nepriklausomai nuo laikyto testo varianto ir mokinių savybių. Poveikio teksto suvokimo rezultatams vertinimas atliktas siekiant įsitikinti, kad bendrų poveikio įverčių neiškraipo matavimo problemos, kylančios dėl didelės antrojo patikrinimo metu rašymo užduočių neatlikusių (0 taškų surinkusių) mokinių dalies (31 proc.).

Viena galimų žemesnio lietuvių kalbos ir literatūros patikrinimų poveikio priežasčių – Lietuvių kalbos ir literatūros elektroninės užduotys tiesiogiai nėra susijusios su lietuvių kalbos ir literatūros egzamino programos reikalavimais. Tai galėjo lemti mažesnį motyvacinį patikrinimų poveikį.

17 lentelė. Tarpinio lietuvių kalbos ir literatūros pasiekimų vertinimo įtaka mokinių rezultatams: apibendrinimas

Tikrinama hipotezė	Rodiklis		Poveikio įvertis
H1. **Vidutiniškai** vertinimas turi teigiamą poveikį matuojamiems mokinių rezultatams 12 klasėje	Vidutinis vertinimo poveikis		-1
H2. Vertinimo poveikis priklauso nuo **testų turinio**	Poveikis moksleiviams, **antroje** bangoje laikiusiems testo variantą (lyginant su kontroline grupe)	Nr. 1	-2.3*
		Nr. 2	-1.4
	1 testo varianto poveikis, lyginant su 2 (beta koeficientų skirtumas)		+0.9
H3. Tarpiniai patikrinimai turi didesnį teigiamą poveikį **moksleiviams, kilusiems iš žemų pajamų šeimų**	Poveikio skirtumas žemesnių pajamų šeimų moksleiviams		*Duomenys vertinimo metu neprieinami
H4. Tarpiniai patikrinimai turi didesnį teigiamą poveikį **moksleiviams, iki intervencijos sukaupusiems žemesnes kompetencijas**	Poveikio skirtumas iki intervencijos aukštesnes kompetencijas (1 PUPP balu) sukaupusiems moksleiviams		-0.15
H5. Tarpiniai patikrinimai turi didesnį teigiamą poveikį **moksleiviams, priklausantiems kalbinėms mažumoms**	Poveikio skirtumas ne lietuvių kalba besimokantiems (tautinių mažumų) moksleiviams		+4.2

P reikšmės: . - <0.1, * - <0,05,** - <0,01, ***-<0,001. P reikšmė yra tikimybė, kad tikroji statistika (tuo atveju, kai H0 teisinga) yra ne mažesnis už stebimą įvertį.

Duomenys: ŠVIS ir NŠA; skaičiavo autoriai.

Apibendrinant, bandomojo tarpinių patikrinimų išbandymo metu 2020 m. rugsėjo–2021 m. kovo mėnesiais ši intervencija neturėjo stabilios ir reikšmingos įtakos mokinių pasiekimams. Siekiant nustatyti, kokios aplinkybės galėjo turėti įtakos intervencijos sėkmei, reikalingas papildomas kokybinis tyrimas. Pavyzdžiui, gali būti, kad nacionaliniu mastu organizuojami tarpiniai pasiekimų patikrinimai ne papildė, o pakeitė kitas vyresniųjų klasių mokinių ugdymo procese įprastas priemones ir praktikas (kaip mokinių pasiekimų tarpinis tikrinimas klasėse, pasitelkiant ankstesnių metų VBE užduotis, klase.eduka.lt, egzaminatorius.lt ir panašias priemones).

2020 m. rugsėjo–2021 m. kovo mėnesiais vykdyti tarpinių pasiekimų patikrinimų rezultatai nesuteikia pagrindo tokią pačią intervenciją tęsti ir plėsti mokinių pasiekimų didinimo tikslais. Vis dėlto šios vertinimo išvados neturėtų būti laikomos galutinėmis. Vertinimo apribojimai ir gairės tolesniems tyrimams pristatomi kitame poskyryje.

2.4. Vertinimo apribojimai ir gairės tolesniems tyrimams

1. **Galima tikėtis, kad tarpinių patikrinimų rezultatus integravus į baigiamuosius įverčius ir naudojant priėmimui į aukštąsias mokyklas poveikis būtų didesnis.** Remiantis anksčiau atliktais tyrimais, mokytis labiau motyvuoja vertinimai, kurie yra susieti su kitais mokiniui aktualiais „apdovanojimais“, kaip galimybės studijuoti aukštojoje mokykloje[18]. Vertinimo metu tarpiniai rezultatai neturėjo tiesioginių „kietųjų“ padarinių galutiniams mokinių vertinimams ar stojimo galimybėms.

2. **Poveikio vertinimo metu tikslinė grupė buvo laikiusi vienu nacionaliniu lietuvių kalbos ir literatūros testu ir dviem matematikos testais daugiau nei kontrolinė**. Tuo metu ankstesni tyrimai rodo, kad testų laikymo gebėjimai auga drauge su praktikos kiekiu. Todėl poveikio vertinimą prasminga kartoti po VBE sesijos, kai bus galima matuoti maksimalaus tarpinių patikrinimų skaičiaus poveikį.

3. **Santykinai maža tyrimo imtis.** Tyrimo metu prieinama moksleivių imtis leidžia gana patikimai patikrinti vidutinį tarpinių patikrinimų poveikį mokinių pasiekimams. **Vis dėlto imtis per maža, siekiant patikimai**

[18] OECD, *Synergies for better learning: an international perspective on evaluation and assessment*, 2013

įvertinti poveikio skirtumus skirtingoms moksleivių grupėms ir analizuoti sudėtingesnius, ne tiesinius, ryšius. Pavyzdžiui, patikrinti hipotezę, kad tarpiniai patikrinimai turi didžiausią poveikį iki intervencijos moksleiviams, kurių akademiniai pasiekimai iki intervencijos buvo vidutiniški: nei ypač aukšti, nei žemi. Todėl poveikio vertinimą prasminga kartoti po VBE sesijos, kai tyrimui bus galima panaudoti visos VBE laikančių moksleivių populiacijos duomenis.

4. **Tikrinamos žinios skyrėsi tarp patikrinimo bangų.** Pavyzdžiui, vienoje bangoje galėjo būti tikrinamos trigonometrijos, o kitoje – tiesinės algebros žinios. Šie patikrinimų turinio skirtumai ribojo galimybes dėl ankstesnių pasiekimų patikrinimų pagilintas dalykines žinias ir gebėjimus pademonstruoti vėlesniuose patikrinimuose. Todėl poveikio vertinimą prasminga kartoti po VBE sesijos – galima tikėtis, kad platus VBE tikrinamų žinių ir gebėjimų spektras sudarys palankesnes sąlygas pademonstruoti platesnį tarpinių patikrinimų metu įgytų kompetencijų spektrą.

5. **Tikėtina, kad įtakos stebimiems rezultatams turėjo nuotolinis darbas dėl COVID-19: mokymasis namuose ir nuotolinis patikrinimų laikymas.** Pavyzdžiui, jei nuotolinis egzaminų laikymas reikšmingai padidino nusirašinėjimo dažnį, drauge mažėjo pasiekimų įverčių patikimumas ir galimybės jais remiantis daryti prasmingas išvadas apie tarpinių patikrinimų poveikį mokinių žinioms ir gebėjimams. Todėl poveikio vertinimą prasminga pakartoti ateityje, ugdymo procesui normalizavusis.

3. Tarpinių pasiekimų vertinimų panaudojimo modeliai

Didelių statymų (angl. *high stakes*) egzaminų sistemoms – sistemoms, kuriose sertifikuojant vidurinio ugdymo pasiekimus ir priimant į aukštąsias mokyklas ypač svarbūs vienkartiniai centralizuoti egzaminai, – būdingos kelios problemos. Tarp jų – didelė atsitiktinumo įtaka vertinimų rezultatams arba žemas patikimumas, menkas vertinimo validumas, ugdymo turinio ir priemonių susiaurėjimas („teaching for the test“).[19] Galima teigti, kad dalis šių problemų būdingos ir Lietuvai.

Šiame skyriuje analizuojami alternatyvūs tarpinių patikrinimų rezultatų panaudojimo baigiamųjų balų formavimui ir sertifikavimui modeliai, kurie galėtų padėti spręsti su *high stakes* egzaminų sistemomis siejamas problemas bei prisidėtų siekiant kitų švietimo politikos tikslų. Analizuojami trys modeliai:

1) nacionalinių tarpinių patikrinimų[20] rezultatų panaudojimo,

2) VBE perlaikymo,

3) metinių pažymių panaudojimo.

3.1 poskyryje aptariami pagrindiniai modelių vertinimo kriterijai: tai validumas, patikimumas, kaštai ir poveikis pagal baigiamajam balui keliamus tikslus (funkcijas), taip pat – galimi šalutiniai padariniai. 3.2 poskyryje plačiau pristatomi modelių bruožai ir pateikiami modelių vertinimo rezultatai.

3.1. Panaudojimo modelių vertinimo kriterijai

Tarpinių vertinimų panaudojimo scenarijai vertinami atsižvelgiant į bendruosius įverčių (jų dedamųjų) kokybės kriterijus (15 lentelė) ir įverčių atitiktį jiems priskiriamoms funkcijoms (16 lentelė).

Vertinant atsižvelgiama į pasiūlytos priemonės suderinamumą su Lietuvos švietimo politikos kontekstu: kitais bendrojo ugdymo organizavimo ir vertinimo sistemų sprendimais.

18 lentelė. Bendrieji įverčių (baigiamųjų balų) kokybės kriterijai

Kriterijus	Komentaras
Validumas	Ar vertinimas gerai atspindi elgesį ir mokymosi rezultatus, kuriuos siekiama pamatuoti? Teoriškai validumas auga didinant testavimo būdų įvairovę, didesnį svorį skiriant mokytojo vertinimui.
Patikimumas	Ar rezultatai tikslūs? Kiek tikėtina, kad pakartojus vertinimą gautume tą patį rezultatą? Patikimumas auga daugėjant testavimų, automatizuojant, centralizuojant. Empiriškai galima tikrinti, kiek tarp skirtingų testų svyruoja moksleivių vieta geriausiųjų eilėje.
Kaštai	Kiek kainuoja vertinimo organizavimas? Vertiname pagal sudarytą kaštų apskaičiavimo modelį.
Poveikis	Ar įvertis atliepia jam priskirtas funkcijas? Galimos baigiamojo įverčio funkcijos pristatomos 2 lentelėje.

Sudaryta remiantis Harlen (2007), Marzano (2000)

19 lentelė. Baigiamojo balo funkcijos, kurių atžvilgiu vertinamas įverčio sudarymo modelio poveikis (tinkamumas)

Funkcija, kurios atžvilgiu vertinamas modelio poveikis	Komentarai
Brandos atestato suteikimas	Administracinė funkcija. Aktualu tik tuo atveju, jei pasiūlytas modelis turi įtakos tam, kas gali gauti brandos atestatą. Pvz., jei tarpinių pasiekimų vertinimas tarnauja kaip slenkstis siekiant laikyti baigiamuosius egzaminus. Klausimų, į kuriuos reikėtų atsakyti, pavyzdys – ar sistema nebus pernelyg griežta / selektyvi?
Priėmimas į aukštąsias mokyklas	Administracinė funkcija. Ar baigiamasis įvertis padeda atrinkti tinkamiausius studentus?
Mokinių motyvavimas mokytis nuosekliai	Motyvavimo funkcija. Motyvavimo lygis teoriškai priklausys nuo to, kiek baigiamieji balai bus taikomi kitoms funkcijoms: brandos atestatų išdavimui, priėmimui į aukštąsias mokyklas. Tarpinių testavimų poveikį mokinių pasiekimams vertinsime empiriškai. Vis dėlto šiuo metu išbandomų testų metu testų rezultatai nebus naudojami įrašant į atestatą ir baigiamojo balo sudarymui. Balą naudojant galutinio įvertinimo formavimui galima tikėtis tokio paties arba didesnio poveikio.

Sudaryta remiantis Harlen (2007), Marzano (2000)

[19] OECD 2013

[20] Patikrinimų sistema Lietuvoje išbandyta 2020 m. rugsėjo–2021 kovo mėnesiais.

20 lentelė. Šalutiniai panaudojimo modelio padariniai

Šalutinis padarinys	Komentarai
Suinteresuotų šalių pasitenkinimas vertinimo sistema	Mokinių ir moksleivių lūkesčius pasiekimų vertinimo vidurinio ugdymo pabaigoje sistemai atspindi NŠA atlikta apklausa[21]. Teiginiui, kad dalykų brandos egzaminai turėtų būti ne vienintelis instrumentas, skirtas vidurinės mokyklos abituriento pasiekimams įvertinti ir baigiamajam balui suformuoti, pritarė 84 proc. moksleivių ir jų tėvų. Teiginiui, kad dalį brandos egzaminų baigiamojo įvertinimo mokiniai turėtų gauti iki baigiamojo egzamino, pritarė 82 proc. respondentų. NŠA apklausos respondentai dažniausiai nurodė, kad iki baigiamojo egzamino mokiniai turėtų gauti 50 proc. brandos egzaminų baigiamojo įvertinimo. Taip manė 40 proc. respondentų. Dar 24 proc. respondentų nurodė, kad iki baigiamojo egzamino mokiniai turėtų gauti 30–40 proc. brandos egzaminų baigiamojo įvertinimo. Rinkdamiesi iš uždaro alternatyvų sąrašo respondentai dažniausiai nurodė, kad iki brandos egzamino sukaupta baigiamojo įvertinimo dalis turėtų susidėti iš elektroninių testų, vykdomų III–IV gimnazijos klasėse ir vertinamų automatiškai (variantui pritarė 49 proc. respondentų), pagrindinio ugdymo pasiekimų patikrinimų (PUPP) lietuvių kalbos ir literatūros, matematikos rezultatų (46 proc.), mokyklos mokytojų (su pagalba) vertinamų darbų (43 proc.). Vis dėlto kiekvienam konkrečiam baigiamojo įvertinimo dėmeniui pritarė mažiau nei pusė respondentų. Variantas, kad dalį baigiamojo įvertinimo galėtų sudaryti mokytojų rašomi pažymiai, uždaruose klausimuose respondentams nepateiktas. Vis dėlto tarp savo pasiūlymą pateikusių respondentų šį variantą siūlė beveik pusė (44 proc.).
Suinteresuotų šalių santykiai	Teigiamą poveikį turi sistemos, kuriose didesnė išorinio vertinimo svarba, nes mokytoją paverčia ne mokinio teisėju, o sąjungininku, siekiant bendro tikslo – pasiruošti išoriniam vertinimui[22].
Aukštojo mokslo prieinamumo lygybė	Teigiamą poveikį turi sistemos, kurios didesnė individualizuoto vidinio vertinimo, pažymių svarba. Pažymių atotrūkiai įprastai tarp žemo ir aukšto SES, miestų ir kaimų moksleivių įprastai būna mažesni nei VBE – dėl pažymių infliacijos išlošia žemesnių pasiekimų moksleiviai.[23] Moksleivius, kuriems nacionaliniu lygiu iškelti standartai per aukšti, išorinis vertinimas demotyvuoja, didina mokyklos nebaigimo dažnį. Efektas ypač stiprus žemų pajamų, tautinių mažumų, žemesnių pasiekimų moksleiviams[24].

3.2. Panaudojimo modelių vertinimo rezultatai

Vertinami tarpinių patikrinimų (ne)panaudojimo scenarijai taip pat išskirti. Vertinant panaudojimo scenarijus bei siūlant jų rizikų valdymo priemones remiamasi mokslinių ir viešosios politikos poveikio vertinimų rezultatais, užsienio šalių vertinimo-sertifikavimo praktikomis, kaštų modeliavimu, suinteresuotų šalių apklausų ir tarpinių pasiekimų organizavimo testavimo rezultatų duomenimis, taip pat kitomis autorių turimomis žiniomis (ekspertiniu vertinimu).

Vertinama pagal kriterijus, įvardytus 3.1 skyriuje. **Naudomis arba rizikomis laikomos tik tos tikėtinos modelių įgyvendinimo pasekmės, kurios autorių vertinimu reikšmingai ir vienareikšmiškai pakeistų esamą padėtį.**

[21] Išori*nio apibendrinamojo vertinimo sistemos kaita: elektroninė mokytojų apklausa, 2019 m. rugsėjis-spalis.* NŠA vykdytoje apklausoje dalyvavo 145 biologijos, 174 istorijos, 285 lietuvių kalbos ir literatūros, 314 matematikos ir 349 užsienio kalbų mokytojai. 2019 m. gruodžio mėn. *Išorinio apibendrinamojo vertinimo sistemos kaita: elektroninė tėvų ir mokinių apklausa, 2019 m. gruodis* (3182 dalyviai).

[22] Pavyzdžiui, Bishop, John. „Drinking from the fountain of knowledge: Student incentive to study and learn–externalities, information problems and peer pressure." *Handbook of the Economics of Education* 2 (2006): 909-944.

[23] Str. apie pažymių infliaciją Švedijoje.

[24] Pavyzdžiui, Ou, Dongshu. „To leave or not to leave? A regression discontinuity analysis of the impact of failing the high school exit exam." *Economics of Education Review* 29, no. 2 (2010): 171-186.

1 modelis: nacionalinių tarpinių patikrinimų rezultatų panaudojimas

a) **Svorių** scenarijus. Pagal šį scenarijų tarpinių patikrinimų rezultatai turėtų tapti baigiamųjų įvertinimų dalimi. Tai galima įgyvendinti tarpinius egzaminus laikant valstybinio brandos egzamino dalimi (dėmeniu) arba formuojant naują baigiamąjį įvertinimą, kuris apibendrintų tarpinių patikrinimų, valstybinių brandos egzaminų ir kitais vertinimo būdais gautus rezultatus. Kiti, ne vidurinio ugdymo pabaigoje laikomų egzaminų, dėmenys į baigiamuosius įvertinimus integruojami, pavyzdžiui, Prancūzijoje[25], Danijoje[26], Slovėnijoje.[27]

b) **Slenksčių** scenarijus. Tarpinių patikrinimų rezultatai naudojami kaip slenkstis. VBE laikyti gali tik surinkę minimalų balų skaičių tarpiniuose testuose. Slenksčiai taikomi Italijoje (centralizuoti testai laikomi prieš mokyklinius), Suomijoje (būtinas minimalus kreditų skaičius).

Papildomos prielaidos:

- Tarpiniuose patikrinimuose dalyvauja 30.000 mokinių.
- Mokinių darbų vertinimas laikomas ugdymo proceso dalimi (nereikalauja papildomų kaštų).
- Vertinimų moderavimo poreikis, kai vertinimas ne automatinis: 30 proc.
- Slenksčio aukštis / testų svoris baigiamajame įvertinime siekia ~ *40 proc.*
- Tarpinių testų skaičius: ~ *3 vienam dalykui*.
- Dalykų, kurių tarpiniai pasiekimai vertinami, skaičius: ~ *3 (lietuvių kalbos ir literatūros, matematikos ir užsienio kalbos).*
- Tarpinių testų įvairovė: *matematikos – 1 (automatinis vertinimas elektroninėje erdvėje), kalbų - 2 (mišrus vertinimas: elektroninis, vertintojui priskiriant balą, ir žodžiu).*

21 lentelė. Nacionalinių tarpinių patikrinimų rezultatų panaudojimo modelio, svorių scenarijaus (1a), vertinimas

Kaštai (kiek papildomai kainuos tiesioginis modelio įgyvendinimas?)	1,397 mln. Eur.
Naudos (kokie tikėtini teigiami pokyčiai pagal bendruosius įverčių kokybės kriterijus ir vertinamus šalutinius padarinius?)	Remiantis užsienyje atliktais tyrimais, papildomi tarpiniai patikrinimai gali turėti teigiamos įtakos mokinių pasiekimams. Remiantis dalimi tyrimų, šis poveikis tikėtinas nepriklausomai nuo to, ar patikrinimų rezultatai susieti su kitais mokiniui aktualiais ,,apdovanojimais“, kaip galimybė studijuoti aukštojoje mokykloje. Pavyzdžiui, pastebima, kad poreikis prisiminti ir panaudoti žinias testo laikymo metu gerina ilgalaikį informacijos įsiminimą (poveikio mechanizmai plačiau pristatomi 2.1 dalyje). Vis dėlto mokytis labiau motyvuoja vertinimai, kurie yra susieti su kitais mokiniui aktualiais ,,apdovanojimais“[28]. Bandomojo tarpinių patikrinimų išbandymo metu 2020 m. rugsėjo–2021 m. kovo mėnesiais ši intervencija neturėjo stabilios ir reikšmingos įtakos mokinių pasiekimams (vertinimo rezultatai pristatomi 2.3 dalyje). Vis dėlto išbandymo metu tarpinių patikrinimų rezultatai nebuvo susieti su baigiamaisiais balais ar kitais mokiniui svarbiais ,,apdovanojimais“. Todėl empirinių duomenų, kurie leistų vertinti nacionalinių tarpinių patikrinimų panaudojimo baigiamųjų balų formavimui įtaką mokinių pasiekimams, šiuo metu neturime.

[21] European Commission (EURIDYCE), France, *Assessment in General Upper Secondary Education* https://eacea.ec.europa.eu/national-policies/eurydice/content/assessment-general-upper-secondary-education-17_en

[22] European Commission (EURIDYCE), Denmark, *Assessment in General Upper Secondary Education* https://eacea.ec.europa.eu/national-policies/eurydice/content/assessment-general-upper-secondary-education-12_en

[23] European Commission (EURIDYCE), Slovenia, *Assessment in General Upper Secondary Education, https://eacea.ec.europa.eu/national-policies/eurydice/content/assessment-general-upper-secondary-education-56_en*

[28] OECD 2013

	Baigiamųjų įvertinimų **patikimumo augimas dėl didesnio stebėjimų skaičiaus**. Kitoms sąlygoms nekintant, iš kelių stebėjimų sudaryti įverčiai yra patikimesni nei pavienių stebėjimų, nes mažiau priklauso nuo atsitiktinių svyravimų, tokių kaip konkretaus testo turinys ar mokinių savijauta testo laikymo metu. Todėl galima tikėtis, kad baigiamieji balai, kurie susidėtų iš tarpinių patikrinimų ir baigiamųjų egzaminų rezultatų, mokinių žinias atspindėtų patikimiau. Remiantis NŠA atliktos apklausos rezultatais, moksleiviai ir mokytojai dažniausiai nepatikimiausiais įvardija lietuvių kalbos ir literatūros (52 proc.) ir matematikos egzaminus (19 proc.). Kitų dalykų egzaminus kaip nepatikimiausius įvardijo 4 proc. arba mažiau respondentų.[29]
	Galima tikėtis, kad tarpinių patikrinimų integravimas į baigiamąjį balą **sulauktų santykinai aukšto suinteresuotų šalių palaikymo**. Teiginiui, kad dalį brandos egzaminų baigiamojo įvertinimo mokiniai turėtų gauti iki baigiamojo egzamino, pritarė 82 proc. respondentų. Rinkdamiesi iš uždaro alternatyvų sąrašo respondentai dažniausiai nurodė, kad iki brandos egzamino sukaupta baigiamojo įvertinimo dalis turėtų susidėti iš elektroninių testų, vykdomų III–IV gimnazijos klasėse ir vertinamų automatiškai. Šiam variantui pritarė 49 proc. respondentų. Scenarijuje numatytas tarpinių patikrinimų svoris baigiamajame bale – 40 proc. – taip pat dera su suinteresuotų šalių nuostatomis. NŠA apklausos metu respondentai dažniausiai nurodė, kad iki baigiamojo egzamino mokiniai turėtų gauti 50 proc. brandos egzaminų baigiamojo įvertinimo. Taip manė 40 proc. respondentų. Dar 24 proc. respondentų nurodė, kad iki baigiamojo egzamino mokiniai turėtų gauti 30–40 proc. brandos egzaminų baigiamojo įvertinimo.
Prielaidos naudai gauti (sąlygos, kurios reikalingos, kad gautume laukiamą naudą)	**Baigiamieji balai susieti su mokiniui svarbiais „apdovanojimais“**: naudojamas priėmimui į aukštąsias mokyklas, valstybės finansuojamų vietų skirstymui. Ši prielaida svarbi, kad tarpiniai pasiekimų patikrinimai turėtų poveikį motyvacijai mokytis.
	Tarpinių patikrinimų turinys palankus mokinių pasiekimų ugdymui. Dalyje atliktų tyrimų pastebima, kad didesnį teigiamą poveikį turi tarpiniai patikrinimai, kurie pagal reikalaujamų gebėjimų lygį ne per daug lenkia mokinių turimą gebėjimų lygį. Tikėtina, kad didesnį teigiamą poveikį baigiamųjų egzaminų rezultatams taip pat turėtų patikrinimai, kurie savo turiniu ir struktūra būtų panašūs į baigiamuosius egzaminus. Panašumas įgalina lengvesnį tarpinių patikrinimų (ir jų aptarimų) metu įgytų žinių pritaikomumą vėlesniems patikrinimams. Vis dėlto vienareikšmiško atsakymo į klausimą, kokio turinio testai turi didžiausią poveikį mokinių pasiekimams Lietuvoje, šioje ataskaitoje pateikti nesiekiama.
	Testų rezultatai naudojami tolesniame ugdymo procese: pavyzdžiui, jų pagrindu teikiant grįžtamąjį ryšį moksleiviams ir tėvams, padėti nukreipti mokymo(si) pastangas į silpnąsias mokinių pasiekimų sritis, parinkti mokinio žinių lygį geriau atitinkantį ugdymo turinį.

[28] *Išorinio apibendrinamojo vertinimo sistemos kaita: elektroninė tėvų ir mokinių apklausa, 2019 m. gruodis.*

Naudos vertinimo būdai (kaip žinosime, kad gavome naudą?)	Administracinių ŠVIS duomenų analizė.
	Suinteresuotų šalių apklausos.
Priemonės naudai gauti (ką turime padaryti, kad gautume naudą?)	**Konkursinių balų pakeitimas baigiamaisiais balais**, atliekant reikiamus įstatymų pakeitimus.
	Tarpinių patikrinimų turinio tobulinimui reikalingų mokslinių bei taikomųjų **tyrimų finansavimas**.
	Mokytojų **kompetencijų** panaudoti testų rezultatus ugdymo proceso tobulinimui ugdymas.
Rizikos (kokie tikėtini neigiami pokyčiai pagal bendruosius įverčių kokybės kriterijus, baigiamojo balo funkcijas ir vertinamus šalutinius padarinius?)	**Neigiama įtaka ugdymo procesui** dėl papildomo krūvio mokiniams ir ypač – mokytojams. Pareiga dalyvauti tarpiniuose patikrinimuose ir jų vertinimas padidintų mokytojų darbo krūvį, dėl to nukentėtų kitos pedagoginės funkcijos. Dėl to savo ruožtu gali nukentėti ir mokinių pasiekimai: kai kurie tyrimai rodo, kad nors tarpiniai patikrinimai turi teigiamą poveikį mokinių pasiekimams, augant patikrinimų skaičiui šis poveikis mažėja.[30]
	Įverčių patikimumo mažėjimas dėl nepakankamų mokyklų išteklių (ir galimai paskatų) užtikrinti skaidrų, kokybišką patikrinimų organizavimą ir vertinimą.
Rizikos vertinimo būdai (kaip žinosime, kad rizika realizuojasi?)	Suinteresuotų šalių nuomonės tyrimai: apklausos, *fokus* grupės.
	Administracinių moksleivių pasiekimų duomenų analizė. Rodikliai: - Moderavimui pasirinktų darbų, kurių įvertinimai nesutampa, dalis (vertinimo kompetencijų rodiklis). - Atotrūkis tarp mokyklos mokytojų ir moderatorių vertinimų (vertinimo šališkumo rodiklis). - Atotrūkis tarp kontaktiniu ir nuotoliniu būdu patikrinimuose dalyvavusių mokinių pasiekimų (aktualu, jei nuotolinis laikymas išliks galimas). - Mokyklos lygmens atotrūkis tarp tarpinių patikrinimų ir baigiamųjų egzaminų rezultatų (alternatyvus vertinimo kompetencijų ir šališkumo rodiklis, naudotinas mažinant moderuojamų vertinimų dalį arba atsisakius moderavimo).
Rizikos valdymo priemonės (ką turime padaryti problemų prevencijai ir valdymui?)	Siekiant išvengti neigiamos įtakos ugdymo procesui, prasminga: - Pradėti nuo mažesnio tarpinių patikrinimų skaičiaus ir (arba) mažiau mokytojų darbo sąnaudų reikalaujančių vertinimo formų (automatizuoto vertinimo) ir modelį plėsti palaipsniui, atsižvelgiant į suinteresuotų šalių nuomonę. - Identifikavus pedagogų darbo krūvio augimą ir (arba) neigiamą poveikį ugdymo procesui, atsisakyti nuostatos, kad tarpinių patikrinimų vertinimas yra ugdymo proceso dalis, ir jam numatyti papildomas pedagogų darbo valandas (atitinkamai numatant papildomą finansavimą).

[28] Bangert-Drowns, Robert L., James A. Kulik, and Chen-Lin C. Kulik. „Effects of frequent classroom testing." The journal of educational research 85, no. 2 (1991): 89-99.

	Siekiant išvengti su vertinimo patikimumu susijusių problemų, prasminga: - Numatyti papildomas priemones ir finansavimą mokytojų dalykinių bei vertinimo kompetencijų kėlimui. - Pirmaisiais modelio diegimo metais numatyti galimybę (finansavimą) moderuojamų darbų dalį didinti patikrinimų sesijos eigoje, atsižvelgiant į ankstesniu laikotarpiu pastebėtą nesutampančių vertinimų dalį bei nesutapimų mastą. Moderuojamų darbų dalis galėtų būti didinama selektyviai, atsižvelgiant į mokyklos lygmens moderavimo rodiklius. - Kurti ir diegti nacionalinę tarpinių patikrinimų organizavimo tvarką, kuri apimtų ir kitas, nei moderavimas, priežiūros ir kontrolės priemones (pavyzdžiui, šiuo metu taikomas VBE organizavimui).

Slenksčių scenarijus (1b) daugeliu atvejų pasižymėtų panašiais kaštais, naudomis ir rizikomis, kaip ir svorių scenarijus (1a). Vis dėlto paminėtini keli svarbūs skirtumai:

- Slenksčio scenarijaus įgyvendinimas būtų paprastesnis, nes reikalautų mažesnių sistemos pokyčių. Tarpiniai patikrinimai atliktų panašią funkciją, kurią šiuo metu atlieka pažymių vidurkiai, o valstybiniai brandos egzaminai atestavimo ir priėmimo į aukštąsias mokyklas procesuose išlaikytų esamą funkciją. Vis dėlto šis modelis išspręstų ir mažiau problemų.
- Galima tikėtis, kad slenksčių taikymas motyvacinį poveikį turėtų tik žemesnių akademinių pasiekimų moksleiviams, kuriems tikimybė slenksčio neįveikti santykinai aukšta. Tarpinių patikrinimų rezultatų integravimas į konkursinį balą papildomas paskatas mokytis nuosekliai suteiktų platesniam ratui moksleivių.
- Slenksčių scenarijus nepadidintų moksleivių baigiamųjų balų patikimumo. Slenksčiai galėtų sumažinti atvejų, kai VBE laikančio moksleivio žinios neteisingai priskiriamos aukštesniam, nei tikrasis, lygiui, tačiau ne priešingai. Egzaminų sistema iš esmės liktų „didelių statymų“ (angl. *high-stakes examinations*) sistema. Todėl galima tikėtis, kad psichologinė įtampa baigiamojo egzamino laikymo metu reikšmingai nesumažėtų.
- Slenksčių scenarijus mažiau suderinamas su daugumai moksleivių ir jų tėvų būdinga nuostata, kad dalį brandos egzaminų baigiamojo įvertinimo mokiniai turėtų gauti iki baigiamojo egzamino.

2 modelis: VBE perlaikymas

Paliekama dabartinė sistema, numatant galimybes VBE perlaikyti tais pačiais metais. Pavyzdžiui, pirmasis egzaminas – gegužę, kiti du – birželį ir liepą (du perlaikymai numatyti Olandijoje[24]). Moksleiviams suteikiama galimybė pasirinkti geriausią rezultatą.[31]

Papildomos prielaidos:

- Perlaikymo kaštus visiškai finansuoja valstybė.

23 lentelė. Egzaminų perlaikymo modelio vertinimas

Kaštai (kiek papildomai kainuos tiesioginis modelio įgyvendinimas?)	Kiekvienas perlaikymas kainuotų *iki* vienos VBE sesijos organizavimo kainos: 1,7–2 mln.[32] Dėl egzamino organizavimo pastovių jų kaštų dalies (pavyzdžiui, užduočių parengimo) kaštai vienam perlaikančiajam priklausytų nuo aktyvumo perlaikant. Tikėtina, kad suteikiant galimybę abiturientui pasirinkti, kurio egzamino balą įtraukti į atestatą (pirmą kartą laikyto ar perlaikyto), aktyvumas būtų aukštas, nes perlaikantieji niekuo nerizikuotų, bet turėtų galimybę gauti geresnį rezultatą. Jei perlaikyme dalyvautų visi abiturientai, kaina vienam perlaikančiajam – 20–24 Eur. Kuo mažiau moksleivių rinktųsi perlaikyti egzaminus, tuo mažesnė būtų visuminė perlaikymų kaina ir didesnė kaina vienam perlaikančiajam.
Naudos (kokie tikėtini teigiami pokyčiai pagal bendruosius įverčių kokybės kriterijus ir vertinamus šalutinius padarinius?)	Remiantis literatūra, modelis gali turėti **teigiamos įtakos mokinių pasiekimams**. Pirma, modelis gali turėti teigiamos įtakos mokinių dalykinėms žinioms ir gebėjimams. Dalyje tyrimų pastebima, kad teigiamas poveikis žinioms ir gebėjimams pasireiškia, kai intervalas tarp egzaminų yra pakankamas mokinių gebėjimų tobulinimui.[33] Vis dėlto Lietuvoje šis poveikio mechanizmas empiriškai netikrintas.
	Galima tikėtis, kad dėl galimybės perlaikyti egzaminus **aukštojo mokslo prieinamumas** ypač didėtų moksleiviams, kilusiems iš žemų pajamų šeimų, nes jų sprendimas studijuoti (ir programos pasirinkimas) labiau priklauso nuo VBE rezultatų (aukštesnių pajamų šeimų moksleiviai yra labiau pajėgūs žemus balus kompensuoti mokėdami už studijas). Kadangi šiuo metu moksleiviams, kilusiems iš žemų pajamų šeimų, galimybės įgyti aukštąjį išsilavinimą yra mažesnės, tai galėtų prisidėti prie lygių galimybių kūrimo.
	Galima tikėtis, kad modelis sulauktų **suinteresuotų šalių palaikymo** dėl mažesnės psichologinės įtampos egzaminų metu.
Prielaidos naudai gauti (sąlygos, kurios reikalingos, kad gautume laukiamą naudą)	Intervalas tarp egzaminų perlaikymų turi būti pakankamas gebėjimų ir žinių tobulinimui.
	Galimybė perlaikyti egzaminą moksleiviams prieinama vienodai: nepriklausomai nuo šeimos finansinės padėties ir kitų charakteristikų.
Naudos vertinimo būdai (kaip žinosime, kad gavome naudą?)	Suinteresuotų šalių apklausos
	Poveikio vertinimas remiantis administraciniais duomenimis. Administracinių duomenų analizė leistų įvertinti: - modelio įtaką mokinių VBE rezultatams, - modelio įtaką aukštojo mokslo prieinamumui, - naudos pasiskirstymą tarp įvairių visuomenės grupių.
Priemonės naudai gauti (ką turime padaryti, kad gautume naudą?)	*Kitos, nei tiesioginės modelio įgyvendinimo, priemonės nereikalingos.*

[31] European Commission (EURIDYCE), Netherlands, *Assessment in General Upper Secondary Education* https://eacea.ec.europa.eu/national-policies/eurydice/content/assessment-general-lower-secondary-education-25_en

[32] Remiantis 2020 m. VBE kaina.

[33] Pavyzdžiui, Paff, Bradley. „Effect of test retakes on long-term retention." PhD diss., 2012.

Rizikos (kokie tikėtini neigiami pokyčiai pagal bendruosius įverčių kokybės kriterijus, baigiamojo balo funkcijas ir vertinamus šalutinius padarinius?)	Dėl ilgesnės egzaminų sesijos **gali sutrikti priėmimo į aukštąsias mokyklas procesas**.
	Tikėtina, kad dėl galimybės pasirinkti aukščiausią iš gautų įvertinimų ne tik augtų VBE rezultatų vidurkis, tačiau ir mažėtų sklaida. Priėmimo į aukštąsias mokyklas sistemai išliekant tai pačiai, dėl to galėtų nežymiai mažėti ir konkursinių balų variacija. Iš valstybės perspektyvos, tai galėtų **sukelti problemų sudarant geriausiųjų eilę priėmimui į studijų vietas**, dėl kurių paprastai konkuruoja ypač aukštų pasiekimų moksleiviai (kaip medicina). Iš asmens perspektyvos, tai **galėtų pabloginti geriau egzaminams pasiruošusių moksleivių, kurie aukštų rezultatų būtų pasiekę ir be perlaikymo galimybės, santykinę padėtį**.
Rizikos vertinimo būdai (kaip žinosime, kad rizika realizuojasi?)	Suinteresuotų šalių nuomonės tyrimai: apklausos, *fokus* grupės.
	Administracinių duomenų analizė. Galimi rodikliai: - VBE rezultatų variacija - Priėmimo į aukštąsias mokyklas konkursinių balų variacija
Rizikos valdymo priemonės (ką turime padaryti problemų prevencijai ir valdymui?)	Priėmimo į aukštąsias mokyklas tvarkaraščiai turėtų būti keičiami prisitaikant prie egzaminų perlaikymų sesijų tvarkaraščių.
	Siekiant išlaikyti konkursinių balų variaciją, reikėtų numatyti papildomus balo dėmenis – papildomus vertinimo kriterijus.

3 modelis: metinių pažymių panaudojimas

Į baigiamąjį balą integruojami metiniai pažymiai. Tai nėra unikali praktika, pavyzdžiui, Vokietijoje paskutinių 2 metų pažymiai sudaro 2/3 galutinio balo. Pažymiai į galutinį vertinimą įskaičiuojami ir kitose šalyse, pavyzdžiui, Danijoje.

Papildomos prielaidos:

- Mokinių vertinimas pažymiais toliau laikomas ugdymo proceso dalimi (nereikalauja papildomų kaštų).
- Vienkartiniai metinių pažymių ir VBE balų integravimo kaštai neįskaičiuojami į modelio įgyvendinimo kaštus.

22 lentelė. Metinių pažymių integravimo į baigiamąjį įvertinimą modelio vertinimas

Kaštai (kiek papildomai kainuos tiesioginis modelio įgyvendinimas?)	0 Eur. Kadangi vertinimas pažymiu yra ugdymo proceso dalis, modelio įgyvendinimas papildomai nekainuotų.
Naudos (kokie tikėtini teigiami pokyčiai pagal bendruosius įverčių kokybės kriterijus ir vertinamus šalutinius padarinius?)	Didesnis baigiamojo įvertinimo **validumas.**
	Remiantis literatūra, tinkamai įgyvendinamas modelis gali padidinti mokytojo vertinimų svorį ugdymo procese, o kartu paskatinti nuoseklesnį mokymąsi **ir geresnius rezultatus**. Vis dėlto šio modelio poveikis mokinių pasiekimams Lietuvoje empiriškai netikrintas.
	Pakeitimas potencialiai galėtų susilaukti **suinteresuotų šalių palaikymo**. Teiginiui, kad dalykų brandos egzaminai turėtų būti ne vienintelis instrumentas, skirtas vidurinės mokyklos abituriento pasiekimams įvertinti ir baigiamajam įvertinimui suformuoti, pritarė 84 proc. moksleivių ir jų tėvų. Variantas, kad dalį baigiamojo įvertinimo galėtų sudaryti mokytojų rašomi pažymiai, uždaruose klausimuose respondentams nepateiktas. Tarp savo pasiūlymą pateikusių respondentų šį variantą pateikė beveik pusė (44 proc.). Vis dėlto ši informacija neleidžia daryti išvadų apie varianto populiarumą visoje populiacijoje. Tai, kad alternatyvą respondentai nurodė dažnai, galėjo iš dalies lemti jos žinomumas.
Prielaidos naudai gauti (sąlygos, kurios reikalingos, kad gautume laukiamą naudą)	Mokytojai **geba** taikyti ir taiko vertinimo strategijas, kurios leidžia įvertinti aukštesnio lygmens, kompleksines kompetencijas, lyginant su VBE.
	Mokytojai pažymius rašo **nešališkai.**
	Baigiamasis įvertinimas susietas su mokiniui svarbiais „apdovanojimais“: pavyzdžiui, **naudojamas** priėmimui į aukštąsias mokyklas, valstybės finansuojamų vietų skirstymui. Ši prielaida svarbi, kad tarpiniai pasiekimų patikrinimai turėtų poveikį motyvacijai mokytis.
Naudos vertinimo būdai (kaip žinosime, kad gavome naudą?)	Suinteresuotų šalių **apklausos**.
	Poveikio vertinimas remiantis **administraciniais duomenimis**.
Priemonės naudai gauti (ką turime padaryti, kad gautume naudą?)	Mokytojų dalykinių ir vertinimo **kompetencijų kėlimas**.
	Konkursinių balų pakeitimas baigiamaisiais įvertinimais, atliekant reikiamus įstatymų pakeitimus.

Rizikos (kokie tikėtini neigiami pokyčiai pagal bendruosius įverčių kokybės kriterijus, baigiamojo įvertinimo funkcijas ir vertinamus šalutinius padarinius?)	Lyginant su standartizuotais nacionaliniais vertinimo įrankiais, **mažesnis patikimumas**. Dėl mokyklų vertinimo tvarkų skirtumų, vertintojo subjektyvumo, tikėtina, kad kito mokytojo pažymiai kitoje mokykloje skirtųsi.
	Prastesni suinteresuotų šalių santykiai: didesnis moksleivių ir jų šeimų spaudimas mokytojams ir mokykloms pasiekimus vertinti aukštesniais pažymiais.
	Tikėtina, kad **baigiamieji įvertinimai prasčiau atitiktų administracines funkcijas:** atestato teikimo ir priėmimo į aukštąsias mokyklas. Taip teigti galima dėl dviejų priežasčių. Pirma, dėl santykinai siauros skalės ir mažos variacijos pagal pažymius moksleiviai diferencijuojasi mažiau nei pagal VBE rezultatus. Antra, tikėtina, kad pažymių svarbos didinimas lemtų jų infliaciją: mokytojų rašomų pažymių augimą, mokinių žinioms ir gebėjimams nekintant. Ši rizika – numanomo moksleivių ir jų tėvų spaudimo rašyti aukštesnius pažymius bei laisvo mokyklos pasirinkimo pasekmė. Pažymių infliacija ne tik mažintų baigiamųjų įvertinimų atitiktį administracinėms funkcijoms švietimo sistemos lygiu, tačiau ir galėtų turėti neigiamų individualių pasekmių. Atliktuose tyrimuose pastebima, kad pažymių infliacija turi ypač neigiamą poveikį aukštesnių pasiekimų moksleiviams, kurie aukštus pažymius būtų pelnę ir infliacijai neegzistuojant[34].
Rizikos vertinimo būdai (kaip žinosime, kad rizika realizuojasi?)	**Administracinių duomenų** rodiklis: metinių pažymių ir VBE rezultatų atotrūkis, lyginant su pasirinktais metais (pavyzdžiui, metais iki sistemos pakeitimo).
Rizikos valdymo priemonės (ką turime padaryti problemų prevencijai ir valdymui?)	Siekiant pažymių palyginamumo nacionaliniu mastu: a) Baigiamosiose klasėse vykdomi **standartizuoti nacionaliniai testai**, kurie leidžia mokytojui geriau įvertinti mokinio pasiekimus nacionalinių tikslų atžvilgiu ir atitinkamai pakoreguoti savo mokymo bei vertinimo strategiją. b) Pakankamai **detalūs vertinimo kriterijai nustatomi nacionaliniame lygmenyje**, kiekvienam dalykui ir pažymiai rašomi tik remiantis kriteriniu vertinimu. c) Metinių pažymių ir VBE atotrūkio rodiklis įtraukiamas tarp mokytojų ir mokyklų vertinimo rodiklių.

[34] Wikström, C. and Wikström, M., 2005. Grade inflation and school competition: an empirical analysis based on the Swedish upper secondary schools. E*conomics of education Review,* 24(3), pp.309-322, OECD 2013

4. Išvados ir rekomendacijos

Šios apžvalgos **tikslas –** išanalizuoti alternatyvius tarpinių mokinių pasiekimų patikrinimų organizavimo ir jų rezultatų (ne)panaudojimo baigiamųjų įvertinimų formavimui (sertifikavimui) modelius bei pateikti rekomendacijas dėl tinkamiausio modelio taikymo. Siekiant šio tikslo, lig šiol įgyvendinti **uždaviniai**:

1) Įvertinti alternatyvių tarpinių pasiekimų patikrinimų *organizavimo* scenarijų kaštus bei efektyvumą.
2) Įvertinti tarpinių pasiekimų patikrinimų *poveikį mokinių pasiekimams*.
3) Įvertinti trijų alternatyvių tarpinių pasiekimų patikrinimų (ne)panaudojimo modelių naudas bei rizikas. Šie modeliai:
    1. nacionalinių tarpinių patikrinimų rezultatų panaudojimo (slenksčių ir svorių scenarijai),
    2. metinių pažymių panaudojimo,
    3. VBE perlaikymo.

Vertinant tarpinių patikrinimų organizavimo kaštus bei jų efektyvumą, pastebėta, kad daugeliu atvejų didėjant patikrinimų ir patikrinime dalyvaujančių mokinių skaičiui vieno mokinio patikrinimo kaštai mažėja (auga kaštų panaudojimo efektyvumas). Vis dėlto, daugėjant patikrinimų, suminiai jų organizavimo kaštai sparčiai auga. Drauge didėja darbo krūvis mokykloms ir mokytojams, o tai gali turėti neigiamos įtakos mokinių pasiekimams.

Remiantis tarpinių patikrinimų išbandymo 2020 m. rugsėjo–2021 m. kovo mėnesiais rezultatais, ši intervencija neturėjo stabilios ir reikšmingos įtakos mokinių pasiekimams. Remiantis šiais rezultatais, tikslo gerinti mokinių pasiekimus atžvilgiu tolesnis šios (nemodifikuotos) priemonės taikymas ir plėtimas nebūtų prasmingas. Vis dėlto, siekiant didesnio išvadų patikimumo, prasminga poveikio vertinimą kartoti po VBE sesijos. Tai leistų padidinti analizės imtį, matuoti didesnio tarpinių patikrinimų skaičiaus poveikį ir įvertinti tarpinių patikrinimų poveikį platesniam dalykinių žinių ir gebėjimų spektrui.

Siekiant didesnio tarpinių patikrinimų poveikio mokinių pasiekimams, svarbu nustatyti, kokios jų savybės ir taikymo aplinkybės ribojo jų poveikį 2020 m. rugsėjo–2021 m. kovo mėnesiais. Pavyzdžiui, gali būti, kad nacionaliniu mastu organizuojami tarpiniai pasiekimų patikrinimai ne papildė, o pakeitė kitas vyresniųjų klasių ugdymo procese įprastas priemones ir praktikas (kaip mokinių pasiekimų tarpinis tikrinimas klasėse, pasitelkiant ankstesnių metų VBE užduotis, klase.eduka.lt, egzaminatorius.lt ir panašias priemones). Tarpinių patikrinimų motyvacinį poveikį taip pat galėjo riboti maža šių patikrinimų svarba tolesniame švietimo procese: tarpinių patikrinimų rezultatai neturėjo tiesioginės įtakos mokinių galimybėms laikyti VBE, studijuoti aukštojoje mokykloje ir siekti kitų tikslų.

Tinkamiausio tarpinių patikrinimų (ne)panaudojimo modelio pasirinkimas priklauso nuo to, kokie modelio vertinimo kriterijai (arba modeliui keliami tikslai) laikomi svarbiausiais. Kriterijai, pagal kuriuos modeliai vertinti šioje ataskaitoje, tai validumas, patikimumas, kaštai ir poveikis pagal pasiekimų įverčiams balui keliamus tikslus arba funkcijas, kaip motyvavimas mokytis arba pasitarnavimas priėmimui į aukštąsias mokyklas. Vertinant panaudojimo scenarijus bei siūlant jų rizikų valdymo priemones, remtasi mokslinių ir viešosios politikos poveikio vertinimų rezultatais, užsienio šalių vertinimo-sertifikavimo praktikomis, kaštų modeliavimu, suinteresuotų šalių apklausų ir tarpinių pasiekimų organizavimo testavimo rezultatų duomenimis, taip pat kitomis autorių turimomis žiniomis (ekspertiniu vertinimu).

Nors tarpinių patikrinimų organizavimas siejamas su reikšmingais papildomais kaštais ir savaime nelaiduoja mokinių pasiekimų augimo, remiantis analizės metu turimais duomenimis tinkamas **nacionalinių tarpinių patikrinimų rezultatų panaudojimo modelio, svorių scenarijaus įgyvendinimas sietinas su plačiausiu naudų spektru ir santykinai mažomis bei lengvai valdomomis rizikomis.** Šio modelio bruožai:

- Tarpiniai patikrinimai integruojami į baigiamuosius įvertinimus, priskiriant jiems 40 proc. svorį. Tarpinių patikrinimų rezultatus į baigiamuosius įvertinimus galima integruoti dviem keliais. Pirma, tarpinius egzaminus laikant valstybinio brandos egzamino dalimi (dėmeniu). Antra, formuojant naują baigiamąjį įvertinimą, kuris apibendrintų tarpinių patikrinimų, valstybinių brandos egzaminų ir kitais vertinimo būdais gautus rezultatus.
- Vertinimų moderavimo poreikis, kai vertinimas ne automatinis: 30 proc.
- Tarpinių testų skaičius: 3 vienam dalykui.
- Dalykų, kurių tarpiniai pasiekimai vertinami, skaičius: 3 (lietuvių kalbos ir literatūros, matematikos ir užsienio kalbos).
- Tarpinių testų įvairovė: matematikos – 1 (automatinis vertinimas elektroninėje erdvėje), kalbų – 2 (mišrus vertinimas: elektroninis, vertintojui priskiriant balą, ir žodžiu).

Jei tarpiniuose patikrinimuose dalyvautų 30.000 mokinių, o darbų vertinimas būtų laikomas ugdymo proceso dalimi ir papildomai nefinansuojamas, **tokio modelio įgyvendinimas per metus galėtų kainuoti apytiksliai 1,397 mln. Eur**. Jeigu būtų mažinama moderuojamų darbų dalis ir (arba) pereinama prie automatizuoto elektroninės testo dalies, vertinimo kaštai galėtų mažėti link 1,029 mln. Eur.

Nors šis modelis reikalauja ne tik finansinių išteklių, tačiau ir kelia prisitaikymo iššūkių ugdymo procesui, jis gali prisidėti prie šiuo metu taikomos abiturientų vertinimo sistemos problemų sprendimo ir prisidėti siekiant kitų ugdymo tikslų.

Modelio įgyvendinimo naudos

Lig šiol Lietuvoje atliktas tarpinių patikrinimų poveikio vertinimas nesuteikia pagrindo manyti, kad tarpinių patikrinimų organizavimas savaime lemtų mokinių pasiekimų augimą. Vis dėlto šie rezultatai neleidžia daryti išvadų apie tai, kokį **poveikį pasiekimams** turėtų tarpinių pasiekimų rezultatų panaudojimas baigiamųjų balų formavimui. Išbandymo metu tarpinių patikrinimų rezultatai nebuvo susieti su baigiamaisiais balais ar kitais mokiniui svarbiais ,,apdovanojimais“. Remiantis užsienyje atliktais tyrimais, mokytis labiau motyvuoja būtent tie vertinimai, kurie yra susieti su kitais mokiniui aktualiais ,,apdovanojimais“[35].

Tarpinių patikrinimų rezultatų integravimas į baigiamuosius įvertinimus **gali padidinti jų patikimumą**. Kitoms sąlygoms nekintant, iš kelių stebėjimų sudaryti įverčiai yra patikimesni nei pavienių stebėjimų, nes mažiau priklauso nuo atsitiktinių svyravimų, tokių kaip konkretaus testo turinys ar mokinių savijauta testo laikymo metu.

Gali padidinti vertinimo sistemos atitiktį suinteresuotų šalių lūkesčiams. Teiginiui, kad dalį brandos egzaminų baigiamojo įvertinimo mokiniai turėtų gauti iki baigiamojo egzamino, pritarė 82 proc. respondentų. ¾ respondentų manė, kad iki baigiamojo egzamino mokiniai turėtų gauti 30–50 proc. brandos egzaminų baigiamojo įvertinimo. Rinkdamiesi iš uždaro alternatyvų sąrašo, respondentai dažniausiai nurodė, kad iki brandos egzamino sukaupta baigiamojo įvertinimo dalis turėtų susidėti iš elektroninių testų, vykdomų III–IV gimnazijos klasėse ir vertinamų automatiškai. Šiam variantui pritarė 49 proc. respondentų.

Siekiant įvertinti šių naudų pasiekimo mastą, po modelio įdiegimo svarbu tęsti suinteresuotų šalių nuomonės tyrimus ir poveikio mokinių pasiekimams vertinimą, remiantis administraciniais duomenimis. Visų pirma – įvertinti dalyvavimo bandomajame tarpinių patikrinimų sistemos išbandyme poveikį mokinių valstybinių brandos egzaminų rezultatams.

Prielaidos ir priemonės naudai gauti

Tam, kad tarpiniai patikrinimai turėtų maksimalų poveikį mokinių pasiekimams, svarbu, kad iš jų formuojami baigiamieji įvertinimai būtų susieti su mokiniui svarbiomis pasekmėmis. Jei tarpinių patikrinimų rezultatai būtų įskaičiuojami į galutinį VBE įvertinimą, veikiant dabartinei sistemai jie būtų automatiškai susieti su galimybėmis įgyti brandos atestatą ir siekti aukštojo mokslo. Jei tarpinių patikrinimų rezultatai būtų naudojami kitų baigiamųjų įverčių sudarymui, svarbu imtis papildomų veiksmų, kurie įverčius susietų su mokiniui svarbiomis pasekmėmis.

Kadangi patikrinimų poveikis priklauso nuo jų užduočių turinio, svarbu toliau vykdyti patikrinimų turinio tobulinimui reikalingų mokslinių bei taikomųjų **tyrimų finansavimas**.

Atsižvelgiant į tai, kad patikrinimų poveikis iš dalies priklauso nuo jų rezultatų panaudojimo tolesniame ugdymo procese, svarbu ugdyti tam reikalingas **mokytojų kompetencijas**.

Rizikos ir jų valdymo būdai

Didelis tarpinių patikrinimų skaičius gali turėti neigiamos įtakos studijų procesui dėl papildomo krūvio mokiniams ir ypač – mokytojams. Dėl to savo ruožtu gali nukentėti ir mokinių pasiekimai: kai kurie tyrimai rodo, kad nors tarpiniai patikrinimai turi teigiamą poveikį mokinių pasiekimams, augant patikrinimų skaičiui šis poveikis mažėja. Siekiant išvengti neigiamos įtakos ugdymo procesui prasminga:

- **Pradėti nuo mažesnio tarpinių patikrinimų skaičiaus ir (arba) mažiau mokytojų darbo sąnaudų reikalaujančių vertinimo formų (automatizuoto vertinimo)** ir modelį plėsti palaipsniui, atsižvelgiant į suinteresuotų šalių nuomonę. Pavyzdžiui, pirmaisiais modelio įgyvendinimo metais tarpinių patikrinimų skaičių būtų galima apriboti iki dviejų, o vertinimo formas iki vienos: automatizuotos.

[35] OECD 2013

- Identifikavus pedagogų darbo krūvio augimą ir (arba) neigiamą poveikį studijų procesui, **atsisakyti nuostatos, kad tarpinių patikrinimų vertinimas yra ugdymo proceso dalis ir jam numatyti papildomas pedagogų darbo valandas** (atitinkamai numatant papildomą finansavimą).

Dėl nepakankamų mokyklų išteklių ir galimai paskatų užtikrinti skaidrų, kokybišką patikrinimų organizavimą ir vertinimą tarpinių patikrinimų įvertinimai gali būti nepatikimi. Tokiu atveju tarpinių patikrinimų rezultatų integravimas į baigiamuosius įvertinimus keltų grėsmę ir jų patikimumui. Galimi tarpinių patikrinimų patikimumo rodikliai:

- Moderavimui pasirinktų darbų, kurių įvertinimai nesutampa, dalis (vertinimo kompetencijų rodiklis).
- Atotrūkis tarp mokyklos mokytojų ir moderatorių vertinimų (vertinimo šališkumo rodiklis).
- Mokyklos lygmens atotrūkis tarp tarpinių patikrinimų ir baigiamųjų egzaminų rezultatų (alternatyvus vertinimo kompetencijų ir šališkumo rodiklis, naudotinas mažinant moderuojamų vertinimų dalį arba ateityje atsisakius moderavimo).
- Atotrūkis tarp kontaktiniu ir nuotoliniu būdu patikrinimuose dalyvavusių mokinių pasiekimų (aktualu, jei nuotolinis laikymas išliks galimas).

Siekiant išvengti su vertinimo patikimumu susijusių problemų, prasminga:

- Numatyti papildomas priemones ir finansavimą mokytojų dalykinių bei vertinimo kompetencijų kėlimui.
- Pirmaisiais modelio diegimo metais numatyti galimybę (ir finansavimą) moderuojamų darbų dalį didinti patikrinimų sesijos eigoje, atsižvelgiant į ankstesniu laikotarpiu pastebėtą nesutampančių vertinimų dalį bei nesutapimų mastą. Moderuojamų darbų dalis galėtų būti didinama selektyviai, atsižvelgiant į mokyklos lygmens moderavimo rodiklius.
- Kurti ir diegti nacionalinę tarpinių patikrinimų organizavimo tvarką**,** kuri apimtų ir kitas, nei moderavimas, priežiūros ir kontrolės priemones (pavyzdžiui, šiuo metu taikomas VBE organizavimui).

Literatūra

Bangert-Drowns, Robert L., James A. Kulik, and Chen-Lin C. Kulik. „Effects of frequent classroom testing." The journal of educational research 85, no. 2 (1991): 89-99.

Baumert, J. and Demmrich, A., 2001. Test motivation in the assessment of student skills: The effects of incentives on motivation and performance. European Journal of Psychology of Education, 16(3), p.441.

Bishop, John. „Drinking from the fountain of knowledge: Student incentive to study and learn–externalities, information problems and peer pressure." Handbook of the Economics of Education 2 (2006): 909-944.

European Commission (EURIDYCE), Denmark, Assessment in General Upper Secondary Education https://eacea.ec.europa.eu/national-policies/eurydice/content/assessment-general-upper-secondary-education-12_en

European Commission (EURIDYCE), France, Assessment in General Upper Secondary Education https://eacea.ec.europa.eu/national-policies/eurydice/content/assessment-general-upper-secondary-education-17_en

European Commission (EURIDYCE), Slovenia, Assessment in General Upper Secondary Education, https://eacea.ec.europa.eu/national-policies/eurydice/content/assessment-general-upper-secondary-education-56_en European Commission (EURIDYCE), Netherlands, Assessment in General Upper Secondary Education https://eacea.ec.europa.eu/national-policies/eurydice/content/assessment-general-lower-secondary-education-25_en

Fitch, Mildred L., Arthur J. Drucker, and J. A. Norton Jr. „Frequent testing as a motivating factor in large lecture classes." Journal of Educational Psychology 42, no. 1 (1951): 1.

Kulik, J.A., Kulik, C.L.C. and Bangert, R.L., 1984. Effects of practice on aptitude and achievement test scores. American Educational Research Journal, 21(2), pp.435-447.

Levant, Beth, Wolfram Zückert, and Anthony Paolo. „Post-exam feedback with question rationales improves re-test performance of medical students on a multiple-choice exam." *Advances in Health Sciences Education* 23, no. 5 (2018): 995-1003.

OECD, Synergies for better learning: an international perspective on evaluation and assessment, 2013

Ou, Dongshu. „To leave or not to leave? A regression discontinuity analysis of the impact of failing the high school exit exam." Economics of Education Review 29, no. 2 (2010): 171-186

Paff, Bradley. „Effect of test retakes on long-term retention." PhD diss., 2012.

Roediger III, H.L. and Karpicke, J.D., 2006. The power of testing memory: Basic research and implications for educational practice. Perspectives on psychological science, 1(3), pp.181-210.

STRATA, Kaip skirtingo SES moksleiviams sukurti lygias galimybes įgyti aukštąjį išsilavinimą?, 2020

Szpunar, K.K., Moulton, S.T. and Schacter, D.L., 2013. Mind wandering and education: from the classroom to online learning. Frontiers in psychology, 4, p.495.

Wikström, C. and Wikström, M., 2005. Grade inflation and school competition: an empirical analysis based on the Swedish upper secondary schools. Economics of education Review, 24(3), pp.309-322

1 priedas. Mokinių ir mokyklų charakteristikų įtaka pirmojo tarpinio matematikos pasiekimų patikrinimo rezultatams

24 lentelė. Daugialypės tiesinės regresijos rezultatai

	Poveikio įvertis, surinktų taškų proc. p.
Laisvasis narys	-21.58***
Mokyklos absolventų matematikos VBE vidurkis (balais)	0.08.
Moksleivio matematikos PUPP rezultatas (balais)	**4.9*** **
Moksleivio lietuvių kalbos PUPP rezultatas (balais)	0.84*
Moksleivio matematikos pažymių vidurkis 11 kl.	4.6***
Lytis: vyras	-4***
Modelio R kvadratas	0.47
Modelio imtis	1078

P reikšmės: . - <0.1, * - <0,05,** - <0,01, ***-<0,001. P reikšmė yra tikimybė, kad tikroji statistika (tuo atveju, kai H0 teisinga) yra ne mažesnis už stebimą įvertį.

Duomenys: ŠVIS ir NŠA; skaičiavo autoriai.

Interpretacijos pavyzdys: moksleiviai, kurių matematikos PUPP pasiekimai buvo 1 balu aukštesni, pirmosios matematikos pasiekimų tarpinių patikrinimų bangos metu surinko 4.9 procentiniu punktu didesnę dalį taškų.

Į pristatomą modelį įtraukti tik kintamieji, kurių įtaka pirmosios bangos tarpinio matematikos pasiekimų patikrinimo rezultatams statistiškai reikšminga. Toliau įvardijami kintamieji, kurių įtaka tikrinta.

Mokyklos lygmens:
1) mokyklos tipas;
2) finansinę paramą gaunančių moksleivių dalis mokykloje;
3) mokyklos vietovė;
4) kalbos, kuriomis mokykloje vykdomas mokymas;
5) vidutiniai mokyklos absolventų lietuvių kalbos ir matematikos VBE rezultatai 2017–2019 metais;
6) vidutiniai mokyklos moksleivių lietuvių kalbos ir matematikos PUPP rezultatai 2017–2019 metais.

Individualaus lygmens:
7) lietuvių kalbos ir matematikos pažymių metinis vidurkis 11 klasėje;
8) lytis;
9) mokymosi kalba.

2 priedas. Mokinių ir mokyklų charakteristikų įtaka tikimybei dalyvauti trijuose tarpiniuose matematikos rezultatų patikrinimuose (patekti į tikslinę grupę)

25 lentelė. Logistinės regresijos rezultatai

	Poveikio įvertis, galimybių santykis (angl. *odds ratio*)
Laisvasis narys	0.02***
Vidutinė finansinę paramą gaunančių moksleivių dalis mokykloje (procentais)	1.26***
Mokinio kalbinė grupė: ne lietuvių	11.23***
Moksleivio matematikos PUPP rezultatas (balais)	**1.08***
Moksleivio lietuvių kalbos PUPP rezultatas (balais)	1.13*
Mokyklos absolventų matematikos VBE vidurkis (balais)	0.96***
Modelio R kvadratas	0.24
Modelio imtis	4046

P reikšmės: . - <0.1, * - <0,05,** - <0,01, ***-<0,001. P reikšmė yra tikimybė, kad tikroji statistika (tuo atveju, kai H0 teisinga) yra ne mažesnis už stebimą įvertį.

Duomenys: ŠVIS ir NŠA; skaičiavo autoriai.

Interpretacijos pavyzdys: moksleivių, kurių matematikos PUPP pasiekimai buvo 1 balu aukštesni, tikimybė dalyvauti trijuose tarpiniuose patikrinimuose (patekti į tikslinę grupę) buvo 8 proc. didesnė.

Į pristatomą modelį įtraukti tik kintamieji, kurių įtaka pirmosios bangos tarpinio matematikos pasiekimų patikrinimo rezultatams statistiškai reikšminga. Toliau įvardijami kintamieji, kurių įtaka tikrinta.

Mokyklos lygmens:
1) mokyklos tipas;
2) finansinę paramą gaunančių moksleivių dalis mokykloje;
3) mokyklos vietovė;
4) kalbos, kuriomis mokykloje vykdomas mokymas;
5) vidutiniai mokyklos absolventų lietuvių kalbos ir matematikos VBE rezultatai 2017–2019 metais;
6) vidutiniai mokyklos moksleivių lietuvių kalbos ir matematikos PUPP rezultatai 2017–2019 metais.

Individualaus lygmens:
7) lietuvių kalbos ir matematikos pažymių metinis vidurkis 11 klasėje;
8) lytis;
9) mokymosi kalba.

3 priedas. Pirmosios matematikos patikrinimų bangos poveikio antrosios bangos pasiekimams vertinimo rezultatai

26 lentelė. Tarpinio matematikos pasiekimų vertinimo įtaka mokinių rezultatams: apibendrinimas

Tikrinama hipotezė	Rodiklis		**Intervencijos poveikis**
H1. **Vidutiniškai** vertinimas turi teigiamą poveikį matuojamiems mokinių rezultatams 12 klasėje	Vidutinis poveikis		3.4
H2. Vertinimo poveikis priklauso nuo **testų turinio**	Poveikis moksleiviams, pirmojoje bangoje laikiusiems testo variantą (lyginant su kontroline grupe):	Nr. 1	+3.7*
		Nr. 2	-0.8
		Nr. 3	+4.3*
	1 testo varianto poveikis, lyginant su 2 (beta koeficientų skirtumas)		+4.2*
	1 testo varianto poveikis, lyginant su 3 (beta koeficientų skirtumas)		-0.6
	2 testo varianto poveikis, lyginant su 3 (beta koeficientų skirtumas)		+5**
H3. Tarpiniai patikrinimai turi didesnį teigiamą poveikį **moksleiviams, kilusiems iš žemų pajamų šeimų**	Poveikio skirtumas žemų pajamų šeimų moksleiviams (lyginant su likusiais)		*Duomenys vertinimo metu neprieinami
H4. Tarpiniai patikrinimai turi didesnį teigiamą poveikį **moksleiviams, iki intervencijos sukaupusiems žemesnes kompetencijas**	Poveikio skirtumas iki intervencijos aukštesnes kompetencijas (1 PUPP balu) sukaupusiems moksleiviams		+1.4*
H5. Tarpiniai patikrinimai turi didesnį teigiamą poveikį **moksleiviams, priklausantiems kalbinėms mažumoms**	Poveikio skirtumas ne lietuvių kalba besimokantiems (tautinių mažumų) moksleiviams (lyginant su likusiais)		+3.5

P reikšmės: . - <0.1, * - <0,05,** - <0,01, ***-<0,001. P reikšmė yra tikimybė, kad tikroji statistika (tuo atveju, kai H0 teisinga) yra ne mažesnis už stebimą įvertį.