ISSN 1392-0340 (Print)
ISSN 2029-0551 (Online)
https://doi.org/10.15823/p.2026.162.7

Pedagogika / Pedagogy
2026, t. 162, Nr. 2, p. 124–150 / Vol. 162, No. 2, pp. 124–150, 2026

VYTAUTO DIDŽIOJO UNIVERSITETO ŠVIETIMO AKADEMIJA

Kokios vertinimo užduotys yra autentiškos? Instrumento užduočių autentiškumui vertinti kūrimas ir pirminis validavimas

Gintarė Lübeck

Vytauto Didžiojo universitetas, Socialinių mokslų fakultetas, Sociologijos katedra, Jonavos g. 66, LT-44138 Kaunas, Lietuva, gintare.luebeck@vdu.lt

Anotacija. Šio tyrimo tikslas buvo sukurti ir validuoti instrumentą, skirtą aukštojo mokslo vertinimo užduočių autentiškumui vertinti. Instrumentas, sukurtas remiantis autentiško vertinimo teorijomis, buvo išbandytas 40-yje vertinimo situacijų. Rezultatai atskleidė aukštą vidinį suderinamumą (α = .97) bei gerą vertintojų tarpusavio suderinamumą. Siekiant pademonstruoti šio instrumento pritaikomumą, straipsnyje analizuojamas ryšys tarp vertinimo autentiškumo ir studentų suvokiamos vidinės užduoties vertės.

Esminiai žodžiai: *autentiškas vertinimas, aukštasis mokslas, užduoties autentiškumo matavimas, suvokiama užduoties vertė.*

Įvadas

Pastarieji dešimtmečiai pasižymėjo gausa tyrimų, siekiančių atskleisti studentų motyvacijos ir įsitraukimo į studijų procesą stiprinimo strategijas (Bols & Freeman, 2011; Bovill, 2020; Healey et al., 2006). Į studentą orientuoti dėstymo metodai plačiai pripažįstami veiksminga alternatyva tradiciniams, į dėstytoją orientuotiems didaktiniams metodams – tyrimai rodo, kad jie gali spartinti studentų mokymąsi (Barral

Copyright © 2026. Gintarė Lübeck. Published by Vytautas Magnus University, Lithuania. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

et al., 2018), gerinti studijų rezultatus (Freeman et al., 2014) bei stiprinti studentų saviveiksmingumą (Ballen et al., 2017).

Vis dėlto, nepaisant vis dažnesnio į studentą orientuotų metodų taikymo universitetuose, daugelyje studijų programų egzaminai išlieka dominuojančiu pasiekimų vertinimo formatu (Kent-Water et al., 2018). Studentai įprastus egzaminus dažnai suvokia kaip nesąžiningus ir netiksliai atspindinčius tikrąsias jų žinias, keliančius nerimą ir stresą (Kent-Water et al., 2018; Lübeck et al., 2025). Mokslinė literatūra taip pat atskleidžia, kad egzaminai gali sudaryti nepalankias sąlygas mažumų grupėms priklausantiems studentams (Lübeck et al., 2022; Putwain, 2008). Tradiciniai vertinimo metodai vis dažniau kritikuojami ir dėl to, kad nepakankamai atspindi, kaip studentų žinios bus vertinamos profesinėje veikloje ar platesniuose gyvenimo kontekstuose (National Forum for the Enhancement of Teaching and Learning in Higher Education, 2017). Tad alternatyvių, lankstesnių ir labiau į studentą orientuotų vertinimo metodų poreikis yra visapusiškai pagrįstas.

Viena iš perspektyviausių mokslinėje literatūroje aprašomų alternatyvų tradiciniams vertinimo būdams yra autentiško vertinimo metodika. Autentiškos užduotys itin teigiamai vertinamos dėl gebėjimo įvertinti aukštesniojo lygmens kognityvinius gebėjimus bei įtraukti studentus į prasmingas, realaus pasaulio kontekstus atspindinčias užduotis (Gulikers et al., 2004; Swaffield, 2011; Villarroel et al., 2020). Studentai autentiškas vertinimo užduotis dažnai suvokia kaip asmeniškai reikšmingesnes ir prasmingesnes nei įprastus egzaminus (Dall'Alba & Barnacle, 2007; Miller & Konstantinou, 2022; Sokhanvar et al., 2021). Be to, autentiškas vertinimas gali prisidėti prie teisingesnės ir įtraukesnės mokymosi aplinkos kūrimo, nes leidžia taikyti lankstesnius vertinimo formatus, galinčius geriau atliepti skirtingus studentų poreikius (Abramenka-Lachheb & de Siqueira, 2021; Hanafin et al., 2007; Kearney & Perkins, 2014; Tai et al., 2023; Tanner, 2001). Be to, tyrimai nuosekliai atskleidžia teigiamą autentiško vertinimo poveikį įvairiems ugdymo rezultatams. Nustatyta, kad autentiško vertinimo elementų integravimas gali didinti studentų motyvaciją (Darling-Hammond, 1995; Näsström et al., 2021), stiprinti pasitikėjimą savo gebėjimais (Ismail et al., 2023; Raymond et al., 2013; Serrano et al., 2018) bei skatinti gilesnį mokymąsi (Ashford-Rowe et al., 2014; Dochy & McDowell, 1997).

Nepaisant augančio susidomėjimo autentišku vertinimu, pats autentiškumo konceptas išlieka daugialypis ir nevienareikšmis. Skirtingi teoriniai požiūriai bei praktiniai kontekstai akcentuoja skirtingus autentiškumo aspektus, todėl autentiškas vertinimas literatūroje operacionalizuojamas labai įvairiai. Nors egzistuoja įvairūs modeliai ir rekomendacijos, skirtos padėti dėstytojams kurti autentiškesnes vertinimo užduotis, trūksta priemonių, leidžiančių sistemiškai ir kiekybiškai įvertinti vertinimo autentiškumo lygį. Be to, tyrimų autentiško vertinimo tema itin trūksta kituose nei Vakarų Europos ir JAV švietimo kontekstuose, o Lietuvoje tokių tyrimų iki šiol nebuvo nė vieno.

Tyrimo tikslas. Šio tyrimo tikslas buvo sukurti ir validuoti vieną pirmųjų kiekybinių instrumentų, skirtų studijų vertinimo užduočių autentiškumui išmatuoti[1]. Tyrime siekta operacionalizuoti autentiško vertinimo konceptą ir sukurti instrumentą, leidžiantį sistemiškai vertinti skirtingų vertinimo užduočių autentiškumą įvairių studijų dalykų kontekstuose. Buvo itin svarbu, kad instrumentas būtų aiškus, praktiškas ir lengvai pritaikomas ne tik tyrėjams, bet ir dėstytojams ar kitiems švietimo specialistams, neturintiems išsamių žinių apie autentiško vertinimo teorijas ar strategijas. Instrumentas buvo grindžiamas autentiško vertinimo literatūroje išskiriamais principais.

Tyrime buvo keliami šie uždaviniai:

1. Sukurti autentiškumo instrumentą, remiantis autentiško vertinimo literatūra ir teoriniais modeliais.
2. Įvertinti instrumento vidinį nuoseklumą ir skirtingų vertintojų tarpusavio suderinamumą.
3. Ištirti, kaip skiriami autentiškumo balai priklauso nuo skirtingų vertinimo formatų.
4. Įvertinti praktinį instrumento pritaikomumą analizuojant ryšius tarp vertinimo autentiškumo ir vidinės vertės, kurią studentai priskiria užduočiai.

Teorinė problemos lauko analizė

Autentiško vertinimo samprata

Anot mokslinės literatūros, nors autentiškumas yra vienas kertinių šiuolaikinio vertinimo elementų, pati autentiškumo samprata išlieka sudėtinga ir nevienareikšmė (Gulikers et al., 2004). Istoriškai autentiškas vertinimas dažnai buvo siejamas su aktyvia veikla grįstu vertinimu (angl. *performance-based assessment*) bei profesinės praktikos imitavimu. Dėl šios priežasties ankstyvosiose autentiško vertinimo konceptualizacijose didelis dėmesys buvo skiriamas tam, kiek vertinimo užduotys atspindi realias profesines situacijas ir reikalauja taikyti praktikoje reikalingas žinias bei gebėjimus. Viena dažniausiai cituojamų autentiško vertinimo operacionalizacijų priklauso Gulikers ir kt. (2004), kurie autentišką vertinimą apibrėžia kaip vertinimą, kuris užduočiai atlikti reikalauja taikyti tas kompetencijas, žinias ir gebėjimus, kurie būtini profesinėje veikloje. Šiame modelyje autentiškumas siejamas su tikroviškomis užduotimis, autentiška vertinimo aplinka, prasmingais užduoties rezultatais ir vertinimo kriterijų aiškumu. Vėlesnės autentiško vertinimo konceptualizacijos akcentuoja ne tik užduoties tikroviškumą, bet ir gilesnius mokymosi procesus. Pavyzdžiui, Ashford-Rowe ir kt. (2014) autentiškumo operacionalizacija kreipia dėmesį į vertinimo užduoties sudėtingumą ir galimybę perkelti žinias bei pripažįsta metakognityvinių procesų ir grįžtamojo ryšio

[1] Kuriant šio tyrimo mokslinį projektą, toks instrumentas dar neegzistavo.

svarbą. Apskritai autentišką vertinimą galima apibrėžti kaip platų alternatyvių vertinimo metodų, kuriais siekiama sustiprinti ryšį tarp universitetinių studijų ir platesnių visuomenės bei darbo rinkos poreikių, spektrą (Ajjawi et al., 2020).

Autentiškas vertinimas Lietuvos aukštojo mokslo kontekste

Autentiško vertinimo aktualumas Lietuvos aukštojo mokslo kontekste yra glaudžiai susijęs su aukštojo mokslo pokyčiais platesniame, Europos kontekste. Lietuvos Respublikos švietimo, mokslo ir sporto ministerijos dokumentuose pastaraisiais metais vis dažniau akcentuojamas poreikis pritaikyti švietimo sistemą sparčiai besikeičiančiai darbo rinkai bei geriau atliepti tiek studentų, tiek platesnės bendruomenės poreikius (Lietuvos Respublikos švietimo ir mokslo ministerija, 2014; Lietuvos Respublikos švietimo, mokslo ir sporto ministerija, 2026). Valstybinėje švietimo 2013–2022 metų strategijoje buvo pabrėžta, kad tuometė gebėjimų vertinimo sistema nepakankamai atliepė šiuolaikines mokymosi ir darbo realijas, buvo neadekvati vertinant tam tikras kompetencijas, o 2021–2030 m. Švietimo plėtros programoje disonansas tarp universitete įgyjamų kompetencijų ir darbo rinkos poreikių įvardijamas kaip viena pagrindinių švietimo sistemos problemų. Tad akivaizdu, jog tai, kas ir kaip vertinama Lietuvos aukštajame moksle, išlieka prioritetiniu Lietuvos švietimo politikos klausimu.

Mokslinė literatūra patvirtina, kad švietimas Lietuvoje vis dar stipriai orientuotas į standartizuotą vertinimą arbitraliomis sąlygomis (Raudienė et al., 2022; Raudienė & Mažeikienė, 2023; Žibėnienė, 2014). Nors Žibėnienė (2014) pažymi, kad vertinimo kultūra Lietuvoje palaipsniui pereina nuo klasikinės normatyvinės prie labiau humanistinės ir į studentą orientuotos paradigmos, kur vertinimas suprantamas ne tik kaip žinių patikrinimas, bet ir kaip studento pažangos, kompetencijų bei gebėjimo taikyti žinias demonstravimas, Raudienė ir Mažeikienė (2023) atskleidžia, kad praktikoje gebėjimų vertinimas vis dar yra standartizuoto pasiekimų matavimo sinonimas, kuomet mokymasis orientuojamas į vertinimo situaciją ir pasiruošimą testui. Raudienė ir kt. (2022) pažymi, jog tai lemia stiprios centralizuotos kontrolės bei standartizacijos tradicijos Lietuvoje.

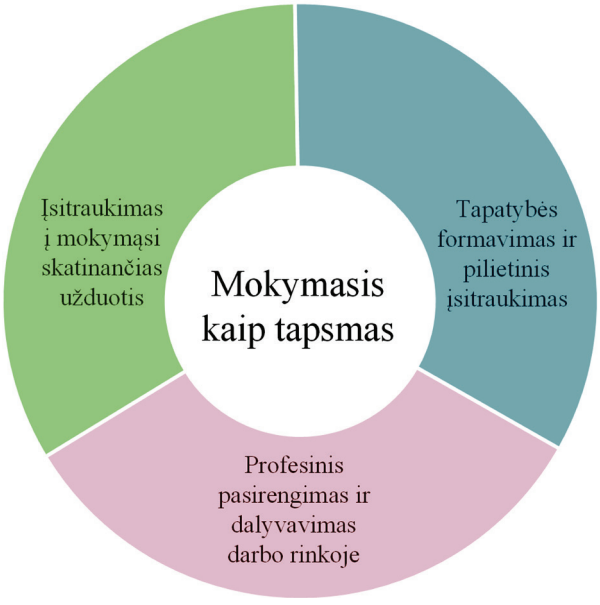
Tokia vertinimo kultūra plačiai kritikuojama kaip mažinanti mokymosi prasmingumą ir skatinanti paviršinį mokymąsi (Raudienė & Mažeikienė, 2023), nesudaranti sąlygų studentui prasmingai įsitraukti į vertinimo užduotį (Lenkauskaitė et al., 2021) ir netinka kompleksinėms kompetencijoms ir kontekstualioms žinioms vertinti (Kaminskienė & Stasiūnaitė, 2013). Tyrimai Lietuvos kontekste siūlo platų vertingesnių alternatyvų spektrą. Raudienė (2018) pabrėžia formuojamojo vertinimo metodų, grįžtamojo ryšio bei studentų aktyvaus įsitraukimo į vertinimo procesą svarbą mokymąsi palaikančiai vertinimo kultūrai puoselėti. Šliogerienė (2012) pažymi, kad alternatyvių vertinimo užduočių, tokių kaip užduočių aplanko (angl. *Portfolio*), atlikimas daugiau dėmesio skiria refleksijai, kuri itin svarbi skatinant gilesnį mokymąsi, savireguliaciją ir autonomiją. Kituose tyrimuose akcentuojama konstruktyvistinių mokymo praktikų, užtikrinančių aktyvų studento dalyvavimą tiek mokymosi, tiek paties vertinimo

procese, svarba ir pabrėžiama, kad vertinimą svarbu suprasti kaip mokymosi dalį (Lenkauskaitė et al., 2021). Tokie metodai kaip problemų sprendimu ar bendradarbiavimu grįstas mokymas(is) bei mokymas(is), užtikrinantis pakankamai galimybių teikti ir gauti grįžtamąjį ryšį bei studentams patiems įsivertinti savo progresą, padeda kurti prasmingesnes mokymosi ir vertinimo patirtis, kuriose studentai labiau įgalinami, prisiima didesnę atsakomybę ir geba geriau valdyti savo mokymąsi (Burkšaitienė, 2012, 2020; Lenkauskaitė et al., 2021).

Mokslinę literatūrą apie vertinimo praktikas Lietuvos švietimo kontekste vienija prasmingesnių, į studentą orientuotų vertinimo metodų paieška. Vis dėlto tokie metodai dažnai sistematizuojami po formuojamojo vertinimo sąvoka, o autentiško vertinimo sąvoka Lietuvos mokslinėje literatūroje apskritai neegzistuoja. Nacionaliniuose metodiniuose projektuose vertinimas vis dažniau konceptualizuojamas išskiriant tris prieigas – mokymosi vertinimą (angl. *assessment OF learning*), vertinimą mokymuisi (angl. *assessment FOR learning*) ir vertinimą kaip mokymąsi (angl. *assessment AS learning*). Vertinimas mokymuisi ir vertinimas kaip mokymasis sulaukia daugiausia dėmesio dėl šių praktikų formuojamojo pobūdžio. Tačiau šiame tyrime į vertinimą siekiama pažvelgti dar plačiau, todėl pasitelkiamas Sambell ir Brown (2022) mokymosi kaip tapsmo modelis (1 pav.). Ši platesnė perspektyva dera su naujesnėmis vertinimo interpretacijomis, kurios vertinimą sieja ne tik su gilesniu mokymusi ar aukštesniais pasiekimais, bet ir su studentų tapatybės, autonomijos bei pilietiškumo raida.

1 paveikslas

Mokymosi kaip tapsmo modelis (Sambell & Brown, 2022)



Tyrimo metodologija

Tyrimo dizainas ir renkami duomenys

Tyrime buvo taikoma mišri metodologinė prieiga, renkami trijų tipų duomenys. Pirma, buvo atliekamas struktūruotas vertinimo situacijų stebėjimas, skirtas analizuoti vertinimo užduočių charakteristikas ir jų įgyvendinimo kontekstą. Antra, autentiškumui kiekvienoje vertinimo situacijoje nustatyti buvo taikomas šiame tyrime sukurtas instrumentas. Trečia, studentų buvo prašoma užpildyti apklausą ir nurodyti, kokią vertę skiria atliktai užduočiai. Šiame straipsnyje analizuojama tik vidinė užduoties vertė, kuri glaudžiai siejama su tuo, kiek vertinimo užduotis studentams pasirodė įdomi ir kiek jiems patiko ją atlikti.

Duomenų rinkimo eiga ir tyrimo dalyviai

Duomenys buvo renkami pirmo kurso bakalauro studijų moduliuose dviejuose didžiausiuose (pagal 2024 m. priėmimo duomenis) Lietuvos universitetuose. Vertinimo situacijos buvo atrenkamos pasitelkus atsitiktinės atrankos principus. Pirmiausia buvo sudaryti abiejų universitetų pirmo kurso modulių ir jų dėstytojų sąrašai, tuomet kiekvienam dėstytojui priskirtas atsitiktinis skaičius nuo 0 iki 1, o sąrašai surikiuoti didėjimo tvarka pagal priskirtas reikšmes. Dėstytojai buvo kviečiami dalyvauti tyrime etapais, laikantis iš anksto numatytos atrankos strategijos. Atrankos tikslas buvo užtikrinti proporcingą vertinimo situacijų pasiskirstymą pagal universitetų pirmo kurso studentų skaičių. Kvotos buvo apskaičiuotos remiantis 2024 m. naujų studijų sutarčių skaičiais (LAMA BPO, 2024). Atsižvelgiant į bendrą planuojamą 36 vertinimo situacijų imtį ir nuspėjamą maždaug 30 % dėstytojų atsako potencialą (q = 0,30), pirmajame kvietimų etape buvo susisiekta su 120 dėstytojų. Antrame etape išsiųsti papildomi 40 kvietimų. Galiausiai dėstytojai iš 40 skirtingų studijų dalykų sutiko, kad jų dalyko atsiskaitymuose būtų atliktas vertinimo užduoties stebėjimas ir autentiškumo vertinimas, o 843 studentai iš šių dalykų sutiko dalyvauti apklausoje. Pagrindinės apklausoje dalyvavusių studentų charakteristikos pateikiamos 1 lentelėje.

1 lentelė

Tyrimo dalyvių charakteristikos

Charakteristika	*n*	% / *M* (*SD*)	Intervalas
Dalyvių skaičius	843		
Amžius	843	19,5 (2,17)	17–46
Lytis			
Moteris	462	54,8	
Vyras	338	40,1	
Kita	33	3,9	
Nenurodė	10	1,2	

Charakteristika	*n*	% / *M* (*SD*)	Intervalas
Universitetas			
Universitetas A	182	21,6	
Universitetas B	661	78,4	

Etiniai tyrimo aspektai

Tyrimui vykdyti buvo gautas Vytauto Didžiojo universiteto Socialinių mokslų fakulteto Sociologijos krypties mokslinių tyrimų atitikties pagrindiniams tyrimų profesionalumo ir etikos principams vertinimo komiteto pritarimas (2025 m. balandžio 14 d., protokolo Nr. 20250414). Vėliau atlikti teorinio modelio ir tyrimo instrumentų pakeitimai buvo papildomai pateikti komitetui vertinti ir patvirtinti 2025 m. rugsėjo 18 d. (protokolo Nr. 20250918-GL). Šie pakeitimai nepakeitė tyrimo dalyvių atrankos ir informavimo, informuoto sutikimo, duomenų apsaugos ar kitų etinių tyrimo procedūrų.

Visiems potencialiems tyrimo dalyviams buvo pateikiama informacija apie tyrimo tikslą ir pobūdį, dalyvavimo sąlygas, konfidencialumo užtikrinimą ir teisę atsisakyti dalyvauti. Informacija studentams buvo pateikiama ne vėliau kaip likus dviem savaitėms iki duomenų rinkimo, o duomenų rinkimo dieną tyrėja juos dar kartą supažindindavo su tyrimu. Tiek tyrime dalyvaujantys dėstytojai, tiek studentai suteikė informuotą sutikimą. Apie studentus, kurie nesuteikė informuoto sutikimo, nebuvo renkami jokie individualaus lygmens duomenys. Dalyvavimas buvo savanoriškas, dalyviai galėjo bet kuriuo metu pasitraukti iš tyrimo be jokių neigiamų pasekmių.

Instrumento kūrimas

Šio tyrimo instrumento kūrimas pirmiausia rėmėsi trimis plačiai autentiško vertinimo literatūroje naudojamomis operacionalizacijomis: Gulikers ir kt. (2004) penkių dimensijų modeliu, Ashford-Rowe ir kt. (2014) aštuonių kritinių autentiškumo elementų modeliu bei Villarroel ir kt. (2018) keturių žingsnių modeliu. Šie modeliai sudarė pagrindinį teorinį instrumento dimensijų ir kriterijų formavimo pagrindą. 2 lentelėje detalizuojama, kaip pasirinkti modeliai atspindėjo naujo instrumento struktūrą ir į kokius argumentus buvo atsižvelgiama formuojant instrumento dimensijas.

2 lentelė

Autentiškumo instrumento dimensijų formavimo teorinis pagrindas

Instrumento dimensija	Pagrindiniai teoriniai šaltiniai	Pagrindinės teorinės idėjos
Užduoties sudėtingumas ir kognityvinis iššūkis	Gulikers et al., 2004; Ashford-Rowe et al., 2014; Villarroel et al., 2018	Autentiškos užduotys turi reikalauti aktyvaus žinių kūrimo, teorijos taikymo, problemų sprendimo, kūrybiškumo bei skirtingų žinių, gebėjimų ir nuostatų integravimo. Vertinimo užduotis turėtų reikalauti naudoti aukštesnio lygmens mąstymą, skatinti analizuoti bei formuluoti išvadas, o ne tik atkartoti informaciją.

Užduoties tikroviškumas ir žinių pritaikomumas	Gulikers et al., 2004; Ashford-Rowe et al., 2014; Villarroel et al., 2018	Autentiškumas siejamas su prasmingomis, praktinėmis ir gyvenimiškomis užduotimis, kurios padeda ugdyti perkeliamuosius gebėjimus bei parodyti, kaip žinios gali būti taikomos praktikoje. Užduočių atlikimo kontekstas ir naudojami ištekliai turėtų kuo tiksliau atitikti profesinę realybę, tačiau tikroviškumas gali būti kuriamas ir suteikiant praktinį kontekstą tradicinėms užduotims.
Metakognicija ir refleksija	Ashford-Rowe et al., 2014; Gulikers et al., 2004	Refleksija ir metakognicija laikomos svarbiais autentiško mokymosi aspektais, padedančiais suprasti mokymąsi platesniame kontekste bei skatinančiais gilųjį mokymąsi. Savęs vertinimas ir refleksija prisideda prie savistabos ir asmeninio tobulėjimo, o studentams turėtų būti suteikiama autonomija patiems spręsti, kokią užduotį ir kaip ją atlikti.
Bendradarbiavimas ir socialinis kontekstas	Gulikers et al., 2004; Ashford-Rowe et al., 2014; Villarroel et al., 2018	Mokymasis ir kompetencijų demonstravimas suprantami kaip socialinė praktika, todėl autentiškos užduotys turėtų atspindėti bendradarbiavimo pobūdį profesinėje veikloje. Vertinimas turi skatinti komunikacinių, darbo komandoje, tarpasmeninių, loginio mąstymo ir retorinių gebėjimų ugdymą. Taip pat pabrėžiama, kad užduotys turėtų būti prasmingos ne tik studentui ar dėstytojui, bet ir platesnei visuomenei, klientams ar darbdaviams.
Užduoties rezultatas	Gulikers et al., 2004; Ashford-Rowe et al., 2014; Villarroel et al., 2018	Autentiškas vertinimas orientuojasi į prasmingo produkto sukūrimą arba veiklos atlikimą, aktyviai demonstruojant įgytas žinias ir gebėjimus. Užduoties rezultatas turėtų atspindėti tikras profesinės praktikos formas, leisti studentams įvairiais būdais parodyti kompetencijas bei pristatyti ir pagrįsti savo darbą auditorijai. Pabrėžiama, kad galutinis rezultatas dažnai yra reikšmingesnis už patį atlikimo procesą.
Vertinimo kriterijai ir sąlygos	Gulikers et al., 2004; Villarroel et al., 2018; Ashford-Rowe et al., 2014	Vertinimo sąlygos – laikas, aplinka ir ištekliai – turėtų kuo labiau atitikti profesinę realybę. Studentams turi būti aiškiai pateikiamos vertinimo rubrikos ir kriterijai prieš vertinimą, tačiau svarbu ne tik formaliai supažindinti su kriterijais, bet ir ugdyti gebėjimą vertinti savo bei kitų darbus, suprasti numanomus ir metakriterijus. Studentai turėtų dalyvauti apibrėžiant vertinimo kriterijus.

Grįžtamasis ryšys ir iteracija	Ashford-Rowe et al., 2014; Villarroel et al., 2018	Grįžtamasis ryšys laikomas neatsiejama sėkmingos profesinės praktikos dalimi ir turėtų būti integruotas į pačią vertinimo struktūrą. Formuojamasis vertinimas padeda studentams geriau suprasti savo gebėjimus ir kryptingiau organizuoti mokymąsi. Pabrėžiama, kad grįžtamasis ryšys turi būti teikiamas nuosekliai, tapti formalia vertinimo dalimi ir sudaryti sąlygas tobulinti darbą.

Formuluojant konkrečius kriterijus ir operacionalizuojant autentiškumo sampratą taip pat buvo atsižvelgta į platesnę teorinę diskusiją apie autentiško vertinimo reikšmę ir ribas. Ypatingos įtakos šiuo požiūriu turėjo Forsyth ir Evans (2019) bei McArthur (2023) darbai, kritikuojantys siaurą autentiškumo sampratą. McArthur (2023) skatina galvoti apie „realistiškas" ir „tikro pasaulio" užduotis ne tik kaip apie užduotis, artimas profesinei praktikai, bet ir kaip prasmingas užduotis, svarbias platesnei visuomenei. Autorė pabrėžia, jog studentus turėtume matyti ne kaip ateities darbo jėgą, bet kaip vertingus visuomenės narius, kurie siekia prasmingai prisidėti prie visuomenės ir kurti socialinį gėrį. Tuo tarpu Forsyth ir Evans (2019) kalba apie tai, kad autentiškas vertinimas turėtų sudaryti sąlygas naujoms žinioms ir perspektyvoms legimitizuoti ir užtikrinti, kad studentams būtų suteikta autonomija pasirinkti vertinimo užduoties kryptį bei patiems apibrėžti, kas jiems svarbu ir aktualu. Pasak autorių, tikras autentiškumas pasiekiamas tuomet, kai studentai atsižvelgia į savo socialinį kontekstą ir interesus, prasmingai įsitraukia į užduotį ir atlieka veiklą, kuri turi apčiuopiamos socialinės vertės.

Forsyth Evans ir McArthur argumentai paskatino atidžiai formuluoti autentiškumo kriterijus, stengiantis užtikrinti, kad autentiškumas juose reikštų ne tik atspindimą darbo rinkos kontekstą. Tad dalis instrumento kriterijų buvo formuluojami taip, kad būtų vertinamas ne tik užduoties ryšys su profesine praktika, bet ir studento autonomija, galimybė prasmingai įsitraukti į užduotį bei platesnis jos aktualumas visuomenei. Antrame paveiksle pristatomas visas instrumentas su apibrėžtomis dimensijomis, kriterijais ir orientaciniais klausimais. Kiekvienas kriterijus buvo vertinamas atsakant į klausimą „Kaip gerai vertinimo užduotis atitinka šį kriterijų?", naudojant Likerto skalę nuo 0 („visiškai neatitinka") iki 3 („labai gerai atitinka"). Susumavus visų kriterijų balus buvo apskaičiuojamas bendras vertinimo autentiškumo balas, o maksimali galima instrumento balų suma siekė 60 taškų.

2 paveikslas

Siūlomas instrumentas autentiškumui vertinti



Instrumento novatoriškumas

Svarbu pažymėti, kad nors šio tyrimo pradžioje literatūroje nepavyko identifikuoti kito kiekybinio instrumento, skirto sistemiškai vertinti užduočių autentiškumą, tyrimo vykdymo laikotarpiu Baines ir kt. (2025) publikavo pirmą kiekybinę autentiško vertinimo tipologiją, pagal kurią vertinamas atskirų vertinimo užduočių autentiškumas, o gauti balai naudojami autentiškumo pasiskirstymui studijų programose analizuoti ir vizualizuoti. Ši prieiga daugiausia grindžiama Gulikers ir kt. (2004) autentiško vertinimo modeliu, papildytu Ashford-Rowe ir kt. (2014) išskirtu metakognicijos aspektu, o vertinami požymiai apibendrinami į dvi platesnes, vertinimo produkto ir proceso, kategorijas.

Šiame tyrime pristatomas instrumentas buvo kuriamas remiantis kitokia metodologija ir skirtas kitokiam tyrimų tikslui. Šio instrumento teorinis pagrindas formuotas aprėpiant platesnę ir naujesnę autentiško vertinimo literatūrą bei atsižvelgiant į autentiško vertinimo sampratos kritiką. Instrumentas sukurtas vertinimo užduoties autentiškumui operacionalizuoti konkrečios vertinimo situacijos lygmeniu ir sudaryti galimybę autentiškumo balą sieti su studentų motyvaciniais įsitikinimais, suvokiama užduoties verte ir mokymosi patirtimi. Toks analizės lygmuo ypač svarbus tiriant mokymąsi kaip tapsmą bei kontekstualius ir konkrečiai užduočiai būdingus studentų įsitikinimus. Todėl, nors šis instrumentas negali būti vadinamas pirmąja kiekybine priemone autentiškumui vertinti, jo teorinis pagrindas ir paskirtis išlieka unikalūs.

Duomenų analizė

Duomenų analizė buvo atliekama naudojant „R“ statistinę programinę įrangą (versija 4.1.2). Pirminėje analizėje buvo apskaičiuota instrumento dimensijų aprašomoji statistika, analizuojamas bendras autentiškumo balų pasiskirstymas tarp skirtingų vertinimo situacijų.

Instrumento vidiniam nuoseklumui įvertinti buvo apskaičiuoti Cronbacho alfa koeficientai tiek visam instrumentui, tiek atskiroms dimensijoms. Instrumentui validuoti buvo pasitelkti išoriniai vertintojai. Penkios atrinktos stebėtos vertinimo situacijos buvo pateiktos dviem nešališkiems švietimo ekspertams. Jiems buvo pateikti stebėjimo protokolai, nuosekliai aprašantys vertinimo užduotis, ir instrumentas autentiškumui vertinti. Išoriniams vertintojams užpildžius instrumentą ir pateikus autentiškumo balus buvo analizuojamas nuoseklumas tarp instrumento autorės (tyrėjos) ir išorinių vertintojų pateiktų autentiškumo balų.

Potencialiai instrumento vertei ir pritaikomumui švietimo tyrimų kontekste įvertinti buvo analizuojami ryšiai tarp autentiškumo balų ir studentų suvokiamos vidinės vertinimo užduoties vertės. Šiam tikslui studentų apklausos duomenys buvo agreguojami vertinimo situacijos lygmeniu ir lyginami su atliktos vertinimo užduoties autentiškumo balu. Ryšiams tarp kintamųjų įvertinti buvo taikoma Pearsono koreliacinė analizė. Vertinimo situacijų lygmeniu agreguotų studentų atsakymų patikimumui ir

rezultatų stabilumui esant nevienodam respondentų skaičiui įvertinti, papildomai apskaičiuoti intraklasinės koreliacijos koeficientai ICC(1) ir ICC(2) bei atliktos jautrumo analizės.

Tyrimo rezultatai

Instrumento veikimas

1. Vidinė struktūra ir balų pasiskirstymas

Instrumento vidinei struktūrai ir diskriminaciniam pajėgumui įvertinti buvo analizuojama aprašomoji statistika ir balų pasiskirstymas visose stebėtose vertinimo situacijose ($N = 40$). Analizė buvo orientuota į bendrųjų autentiškumo balų pasiskirstymą, dimensijų lygmens tendencijas bei priemonės gebėjimą diferencijuoti tarp skirtingų vertinimo formatų.

Bendrieji autentiškumo balai tarp vertinimo situacijų pasižymėjo didele variacija. Balai svyravo nuo 2 iki 54 iš galimų 60, o vidurkis siekė 21,12 (SD = 16,57). Platus balų intervalas ir santykinai didelis standartinis nuokrypis rodo, jog instrumentas geba diferencijuoti tarp skirtingų autentiškumo lygių. Nebuvo nustatyta nei *floor*, nei *ceiling* efektų, kadangi balai pasiskirstė beveik visame teoriniame instrumento balų intervale.

Bendrųjų autentiškumo balų pasiskirstymas buvo vidutiniškai teigiamai asimetriškas (0.73). Tai reiškia, kad žemesnio autentiškumo užduotys stebėtoje imtyje buvo dažnesnės nei aukšto autentiškumo užduotys. Toks pasiskirstymas nestebina, atsižvelgiant į tai, kad aukštajame moksle vis dar dominuoja tradiciniai, į egzaminus orientuoti vertinimo formatai. Ekscesas buvo neigiamas (−0,95). Tai rodo, jog balų pasiskirstymas buvo santykinai platus, o ne koncentruotas ties vidurkiu. 3 lentelėje pateikiamas bendrųjų autentiškumo balų pasiskirstymas visuose vertinimo atvejuose.

3 lentelė

Bendrųjų autentiškumo balų aprašomoji statistika

n	M	SD	Md	Min	Max	Asimetrija	Ekscesas
40	21,12	16,57	14,50	2	54	0,73	−0,95

2. Dimensijų lygmens balų pasiskirstymas

Didžiausi vidurkiai buvo nustatyti dimensijose „Užduoties sudėtingumas ir kognityvinis iššūkis“ (M = 5,08, SD = 2,87) bei „Užduoties tikroviškumas ir žinių pritaikomumas“ (M = 4,40, SD = 3,33). Tai atskleidžia, kad į daugelį vertinimo užduočių buvo integruoti bent tam tikri elementai, padedantys sukurti kompleksiškas, kontekstualias ir disciplinai prasmingas vertinimo užduotis. Abi dimensijos taip pat pasižymėjo

plačiais balų intervalais (0–9) ir santykinai simetriškais pasiskirstymais, demonstruojančiais reikšmingą variaciją tarp vertinimo atvejų.

Tuo tarpu dimensijos „Metakognicija ir refleksija“ (M = 2,0, SD = 2,82), „Bendradarbiavimas ir socialinis kontekstas“ (M = 1,57, SD = 2,81) bei „Grįžtamasis ryšys ir iteracija“ (M = 1,45, SD = 1,60) pasižymėjo mažesnėmis medianomis ir stipria teigiama asimetrija. „Metakognicijos ir refleksijos“ ir „Socialinio konteksto ir bendradarbiavimo“ dimensijų medianos lygios 0. Tai rodo, kad daugelyje stebėtų vertinimų studentams nebuvo sudaroma galimybių reflektuoti, įsivertinti savo darbą, bendradarbiauti ar peržiūrėti ir patobulinti atliktą užduotį. Grįžtamojo ryšio ir iteracijos dimensija pasižymėjo stipriausia teigiama asimetrija (1,72) ir teigiamu ekscesu (2,28). Tai rodo, jog galimybės gauti formuojamąjį grįžtamąjį ryšį bei peržiūrėti ir tobulinti atliekamą užduotį vertinimo praktikoje buvo retos, tačiau keliose vertinimo užduotyse – itin ryškios. Visų autentiškumo dimensijų aprašomoji statistika pateikiama 4 lentelėje.

4 lentelė

Autentiškumo dimensijų aprašomoji statistika

Dimensija	n	M	SD	Md	Min	Max	Asimetrija	Ekscesas
Užduoties sudėtingumas ir kognityvinis iššūkis	40	5,08	2,87	5	0	9	0,08	−1,31
Užduoties tikroviškumas ir žinių pritaikomumas	40	4,40	3,33	3	0	9	0,20	−1,59
Metakognicija ir refleksija	40	2,00	2,82	0	0	9	1,02	−0,49
Bendradarbiavimas ir socialinis kontekstas	40	1,57	2,81	0	0	9	1,49	0,61
Užduoties rezultatas	40	3,72	2,77	3	0	9	0,52	−0,89
Vertinimo kriterijai ir sąlygos	40	2,90	2,24	2	0	8	0,76	−0,63
Grįžtamasis ryšys ir iteracija	40	1,45	1,60	1	0	6	1,72	2,28

3. Tiriamoji faktorinė analizė

Norint įvertinti instrumento vidinę struktūrą, papildomai buvo atlikta tiriamoji faktorinė analizė (EFA). Atsižvelgiant į nedidelį vertinimo situacijų skaičių (N = 40), rezultatai interpretuojami tik preliminariai. Duomenų tinkamumo rodikliai buvo geri: bendras KMO rodiklis siekė .90, atskirų kriterijų MSA reikšmės svyravo nuo .79 iki .96, o Bartletto sferiškumo testas buvo statistiškai reikšmingas, $\chi^2(190) = 950{,}81$, $p < .001$. Lygiagrečioji analizė rekomendavo vieno faktoriaus sprendinį. Vieno faktoriaus MINRES modelyje visų 20 kriterijų faktorinės apkrovos buvo teigiamos ir svyravo nuo .54 iki .89, o bendras faktorius paaiškino 65,5 % kriterijų dispersijos.

Papildomai išbandytas septynių faktorių MINRES sprendinys su įstrižąja rotacija, siekiant patikrinti, ar empiriškai galima atkurti septynias teoriškai išskirtas instrumento

sritis. Šis sprendinys nebuvo stabilus: rotacija nekonvergavo po 1 000 iteracijų, o vieno kriterijaus bendrumo balas viršijo 1, t. y. nustatytas ultra-Heywood atvejis. Todėl septynių faktorių struktūra šioje imtyje negali būti patvirtinta. Nors šie rezultatai yra preliminarūs ir turėtų būti patvirtinti su didesne imtimi, tikslingiausia instrumentą interpretuoti kaip veikiantį stipraus bendro autentiškumo faktoriaus pagrindu.

4. Žinomų grupių validumas

Žinomų grupių (angl. *known-groups*) validumas buvo vertinamas lyginant autentiškumo balus tarp teoriškai skirtingų vertinimo kategorijų. Stebėtose vertinimo situacijose identifikuoti šie vertinimo formatai – testai su keliais atsakymo variantais (angl. *MCQs*, $n = 5$), tradiciniai kolokviumai ir egzaminai ($n = 9$), mišrūs testai[2] ($n = 2$), multikomponentiniai užsienio kalbos testai ($n = 4$), analitinės ir atvirų išteklių užduotys ($n = 5$), praktinių ir techninių gebėjimų vertinimas ($n = 4$), kūrybinės ir projektinės užduotys ($n = 4$), pristatymai ir gynimai ($n = 5$) bei akademinės analizės užduotys ir tyrimai ($n = 3$). Kaip ir tikėtasi, skirtingi vertinimo formatai pasižymėjo reikšmingais autentiškumo skirtumais.

Vertinimai, priskirti testų su keliais atsakymo variantais kategorijai, pasižymėjo žemiausiais autentiškumo balais ($M = 4{,}20$). Ši kategorija pasižymėjo žemomis medianomis ir ribota variacija visose dimensijose. Toks rezultatas buvo nuspėtinas, atsižvelgiant į tai, jog testai su keliais atsakymo variantais dažniausiai yra pagrįsti žinių atkartojimu, o ne prasmingu jų taikymu. Tradiciniai kolokviumai ir egzaminai taip pat pasižymėjo žemais autentiškumo balais ($M = 7{,}56$), nors šioje kategorijoje egzistavo didesnė variacija. Ši variacija dažnai buvo susijusi su disciplininiu autentiškumu. Tai reiškia, kad teorinės užduotys buvo realistiškai kontekstualizuojamos disciplinos atžvilgiu.

Tarpinę poziciją tarp tradicinių egzaminų ir autentiškiausių vertinimo formų užėmė mišrūs testai ($M = 13{,}0$), multikomponentiniai užsienio kalbos testai ($M = 14{,}75$), praktinių ir techninių gebėjimų vertinimas ($M = 21{,}67$) bei analitinės ir atvirų išteklių užduotys ($M = 22{,}40$). Šioms užduotims atlikti dažnai reikėjo, kad studentai taikytų žinias, interpretuotų duomenis, spręstų problemas ar demonstruotų praktinius gebėjimus, todėl atliepė kai kuriuos autentiško vertinimo kriterijus. Vis dėlto jos dažniausiai buvo atliekamos griežtai apibrėžtomis sąlygomis, pasižymėjo individualiu užduoties atlikimu ir tik iš dalies atitiko platesnius autentiško vertinimo kriterijus, susijusius su studentų autonomija, bendradarbiavimu, refleksija ar autentiško produkto kūrimu.

Aukščiausiais autentiškumo balais pasižymėjo kūrybinės ir projektinės užduotys ($M = 33{,}0$), pristatymai ir gynimai ($M = 41{,}60$) bei akademinės analizės užduotys ir tyrimai ($M = 51{,}33$). Šie vertinimo tipai geriau atliepė studentų autonomijos, viešo užduoties pristatymo, prasmingos veiklos ar produkto kūrimo, bendradarbiavimo ir refleksijos kriterijus bei suteikė daugiau galimybių perkelti žinias už disciplinos

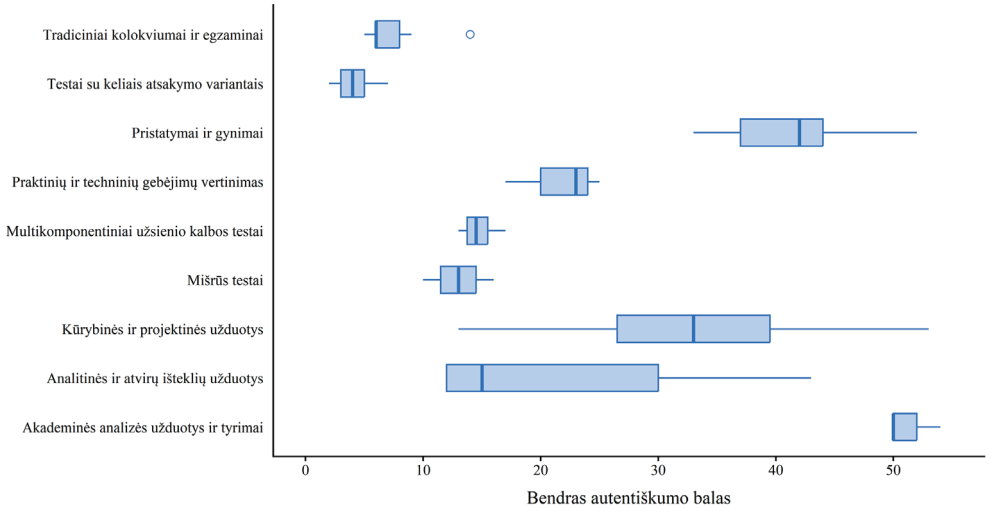
[2] Mišrūs testai šiame tyrime negalėjo būti priskirti nei testų su keliais atsakymo variantais kategorijai, nei tradiciniams kolokviumams ir egzaminams, nes pasižymėjo abiejų kategorijų elementais.

ribų. Ypač aukšti akademinės analizės užduočių ir tyrimų autentiškumo balai rodo, kad užduotys, orientuotos į savarankišką tyrimų atlikimą ir disciplininių problemų nagrinėjimą, stipriausiai atitiko instrumente operacionalizuotus autentiško vertinimo principus.

Akivaizdu, jog vertinimo kategorijos, teoriškai siejamos su autentiškesnėmis vertinimo praktikomis, nuosekliai demonstravo aukštesnius autentiškumo balus nei tradiciniai testai ir egzaminai. Vis dėlto autentiškumas neturėtų būti suprantamas kaip paprasta dichotomija tarp „tradicinių“ ir „autentiškų“ vertinimų, nes reikšminga variacija buvo pastebima ir vertinimo formatų kategorijų viduje (žr. 3 pav.).

3 paveikslas

Autentiškumo balų pasiskirstymas pagal vertinimo užduoties kategorijas



5. **Vidinis suderinamumas ir ryšiai tarp dimensijų**

Bendras instrumento ir atskirų jo dimensijų patikimumas buvo vertinamas atliekant vidinio suderinamumo analizę. Cronbacho alfa koeficientai buvo apskaičiuoti tiek visai 20 kriterijų priemonei, tiek kiekvienai konceptualiai dimensijai atskirai.

Bendra priemonė pasižymėjo itin aukštu vidiniu suderinamumu ($\alpha = .97$). Kriterijų ir bendro balo koreliacijos buvo nuosekliai aukštos, o atskirų kriterijų pašalinimas reikšmingai nepagerino patikimumo rodiklių, taigi nė vienas iš kriterijų nebuvo perteklinis, visi prasmingai prisidėjo prie bendro konstrukto matavimo.

Dimensijų lygmens analizė parodė stiprų visų konceptualių dimensijų vidinį suderinamumą: Cronbacho alfa koeficientai svyravo nuo .74 iki .94. Dimensijos „Užduoties sudėtingumas ir kognityvinis iššūkis“ ($\alpha = .90$), „Užduoties tikroviškumas ir žinių pritaikomumas“ ($\alpha = .92$), „Metakognicija ir refleksija“ ($\alpha = .94$) bei „Bendradarbiavimas ir socialinis kontekstas“ ($\alpha = .94$) pasižymėjo ypač stipriu vidiniu suderinamumu.

Kiek žemesni vidinio suderinamumo koeficientai buvo nustatyti dimensijų „Vertinimo kriterijai ir sąlygos“ (α = .74) bei „Grįžtamasis ryšys ir iteracija“ (α = .74). Santykinai mažesni šių dimensijų rodikliai greičiausiai atspindi didesnį konceptualų šių dimensijų heterogeniškumą. Pavyzdžiui, skaidrūs vertinimo kriterijai, realistiški atlikimo standartai ir studentų įtraukimas į vertinimo procesą nebūtinai visada pasireiškia kartu toje pačioje vertinimo užduotyje. Taip pat vertinimo užduotys dažnai suteikia grįžtamąjį ryšį be galimybės peržiūrėti ar patobulinti darbą.

5 lentelė

Instrumento dimensijų aprašomoji statistika ir vidinio suderinamumo rodikliai

Dimensija	Kriterijai	*M*	*SD*	Min	Max	α
Užduoties sudėtingumas ir kognityvinis iššūkis	3	5,08	2,87	0	9	.90
Užduoties tikroviškumas ir žinių pritaikomumas	3	4,40	3,33	0	9	.92
Metakognicija ir refleksija	3	2,00	2,82	0	9	.94
Bendradarbiavimas ir socialinis kontekstas	3	1,57	2,81	0	9	.94
Užduoties rezultatas	3	3,72	2,77	0	9	.86
Vertinimo kriterijai ir sąlygos	3	2,90	2,24	0	8	.74
Grįžtamasis ryšys ir iteracija	2	1,45	1,6	0	6	.74

Siekiant detaliau įvertinti ryšius tarp dimensijų, buvo apskaičiuoti Pearsono koreliacijos koeficientai tarp visų dimensijų balų. Tarp visų dimensijų nustatytos stiprios teigiamos koreliacijos, jų koeficientai svyravo nuo *r* = .64 iki *r* = .87. Ypač stiprūs ryšiai nustatyti tarp dimensijų „Užduoties sudėtingumas ir kognityvinis iššūkis“ bei „Užduoties rezultatas“ (*r* = .87), „Užduoties tikroviškumas ir žinių pritaikomumas“ ir „Užduoties rezultatas“ (*r* = .85), „Bendradarbiavimas ir socialinis kontekstas“ ir „Užduoties rezultatas“ (*r* = .86).

Atsižvelgiant į šiuos rezultatus panašu, kad skirtingų dimensijų autentiško vertinimo charakteristikos vertinimo užduotyse dažnai pasireiškia kartu. Užduotys, kuriose reikalaujama atlikti prasmingas veiklas ar sukurti prasmingą produktą, dažnai yra sudėtingos ir praktiškos, atliekamos tikroviškuose kontekstuose ir pagrįstos bendradarbiavimu.

Įdomu tai, kad dimensijos „Grįžtamasis ryšys ir iteracija“ bei „Vertinimo kriterijai ir sąlygos“ demonstravo santykinai silpnesnes koreliacijas su kai kuriomis kitomis dimensijomis. Taip gali būti todėl, jog net ir itin realistiškos ar sudėtingos užduotys nebūtinai aiškiai perteikia vertinimo kriterijus ar suteikia prasmingų galimybių grįžtamajam ryšiui ir darbo peržiūrai.

6. Tyrėjos ir išorės vertintojų balų suderinamumas

Siekiant atskleisti, ar instrumentas padeda objektyviai įvertinti užduoties autentiškumą, buvo palyginti trijų vertintojų balai – tyrėjos ir dviejų išorės vertintojų. Tyrėjai įvertinus visas 40 vertinimo situacijų, buvo tikslingai atrinktos penkios skirtingo formato ir autentiškumo lygio užduotys. Išorės vertintojai šias užduotis nepriklausomai įvertino naudodami tą patį autentiškumo vertinimo instrumentą. Tyrėjos ir išorės vertintojų rezultatai pateikiami 6 lentelėje.

6 lentelė

Bendrieji autentiškumo balai pagal vertintojus ir vertinimų variacija

Vertinimo užduotis	1 vertintojas	2 vertintojas	Tyrėja
25	51	53	52
34	5	10	7
35	7	19	17
36	18	25	23
39	6	7	5

Bendras vertintojų tarpusavio suderinamumas buvo aukštas, bet atskirų užduočių vertinimai atskleidė tam tikrą variaciją. Mažiausi skirtumai nustatyti vertinant 25 ir 39 užduotis, o didžiausias – 35 užduotį, kurios bendrieji autentiškumo balai svyravo nuo 7 iki 19. Tai rodo, kad, nepaisant aukšto bendro suderinamumo, atskirų autentiškumo požymių interpretacija tam tikrose vertinimo situacijose gali skirtis.

Dimensijų lygmens analizė parodė, kad mažiausi vidutiniai skirtumai tarp vertintojų nustatyti „Bendradarbiavimo ir socialinio konteksto“ bei „Metakognicijos ir refleksijos“ dimensijose (po 0,10 balo), taip pat „Grįžtamojo ryšio ir iteracijos dimensijoje“ (0,20 balo). Didžiausi vidutiniai skirtumai nustatyti „Užduoties rezultato“ (0,64 balo), „Užduoties sudėtingumo ir kognityvinio iššūkio“ (0,57 balo) bei „Užduoties tikroviškumo ir žinių pritaikomumo“ (0,43 balo) dimensijose (7 lentelė). Šie rezultatai leidžia preliminariai identifikuoti instrumento sritis, kuriose vertintojų interpretacijos gali labiau skirtis ir kurioms tolesniuose instrumento validavimo etapuose gali reikėti tikslesnių vertinimo gairių ar pavyzdžių.

Nepaisant to, bendras vertinimų nuoseklumas išliko aukštas, o vertintojų skirtumai dažniausiai neturėjo įtakos bendroms vertinimo tendencijoms (t. y. ar užduotis vertinama kaip mažai, vidutiniškai ar labai autentiška) ar santykiniam vertinimų išsidėstymui pagal autentiškumo lygį (t. y. kuri pateikta užduotis buvo vertinama kaip autentiškiausia ir kuri kaip mažiausiai autentiška).

7 lentelė

Vidutiniai skirtumai tarp vertintojų pagal instrumento dimensijas

Dimensija	Vid. skirtumas tarp tyrėjos ir 1 vertintojo	Vid. skirtumas tarp tyrėjos ir 2 vertintojo	Bendras vidurkis
Užduoties rezultatas	0,67	0,60	0,64
Bendradarbiavimas ir socialinis kontekstas	0,07	0,13	0,10
Vertinimo kriterijai ir sąlygos	0,47	0,33	0,40
Grįžtamasis ryšys ir iteracija	0,20	0,20	0,20
Metakognicija ir refleksija	0,13	0,07	0,10
Užduoties tikroviškumas ir žinių pritaikomumas	0,53	0,33	0,43
Užduoties sudėtingumas ir kognityvinis iššūkis	0,47	0,67	0,57

Ryšys tarp autentiškumo ir studentų suvokiamos vidinės vertės

Instrumento pritaikomumui aukštojo mokslo kontekste pademonstruoti buvo analizuojamas ryšys tarp bendro autentiškumo balo ir studentų suvokiamos vidinės vertės, agreguotos vertinimo užduoties lygmeniu. Vidinės vertės konstruktas šiame tyrime buvo grindžiamas Eccles ir Wigfield (2024) sukurta situatyvia lūkesčių ir vertės teorija (angl. *Situated Expectancy-Value Theory* – SEVT). SEVT pabrėžia, kad studentų motyvaciją formuoja ne tik individualūs įsitikinimai apie savo gebėjimus, bet ir konkretūs mokymosi kontekstai bei užduotys, su kuriomis studentai susiduria (Gladstone et al., 2022; Tang et al., 2022). Remiantis šia teorija, tokie studentų įsitikinimai kaip sėkmės lūkesčiai bei subjektyvi užduočiai skiriama vertė yra pagrindiniai studentų įsitraukimą, pasiekimus ir motyvaciją prognozuojantys veiksniai.

Vidinė vertė SEVT kontekste apibrėžiama kaip susidomėjimas užduotimi ir patiriamas malonumas ją atliekant ir dažnai priklauso nuo konkrečių užduočių ar disciplinų patrauklumo studentui (Gladstone et al., 2022). Studentų suvokiama vidinė vertė buvo vertinama naudojant trumpą Likerto tipo skalę. Skalės teiginiai buvo parengti remiantis ankstesniais Wigfield ir Eccles (2000), Gaspard ir kt. (2016, 2017, 2018) bei Muenks ir kt. (2023)klausimynais. Studentų buvo prašoma įvertinti savo sutikimą su šiais teiginiais: „Man patiko šiandienos vertinimo užduotis“; „Ši vertinimo užduotis buvo įdomi“; „Šią vertinimo užduotį atlikti buvo smagu“.

Apklausą užpildė 843 studentai, o vertinimo situacijos apėmė platų disciplininių kontekstų spektrą tiek modulių, tiek studijų programų lygmeniu. Į tyrimą buvo įtraukti moduliai iš matematikos ir fizinių mokslų, kompiuterijos ir inžinerijos, socialinių mokslų, humanitarinių mokslų, verslo ir ekonomikos, gyvybės mokslų, sveikatos mokslų, menų bei profesinių / akademinių įgūdžių sričių. Tuo tarpu studijų programų lygmeniu buvo atstovaujama tokioms kryptims kaip verslas ir ekonomika, kompiuterija ir inžinerija, socialiniai mokslai, humanitariniai mokslai, gyvybės mokslai, sveikatos mokslai, pedagoginiai mokslai, menai bei mišrios STEM kryptys.

8 lentelė

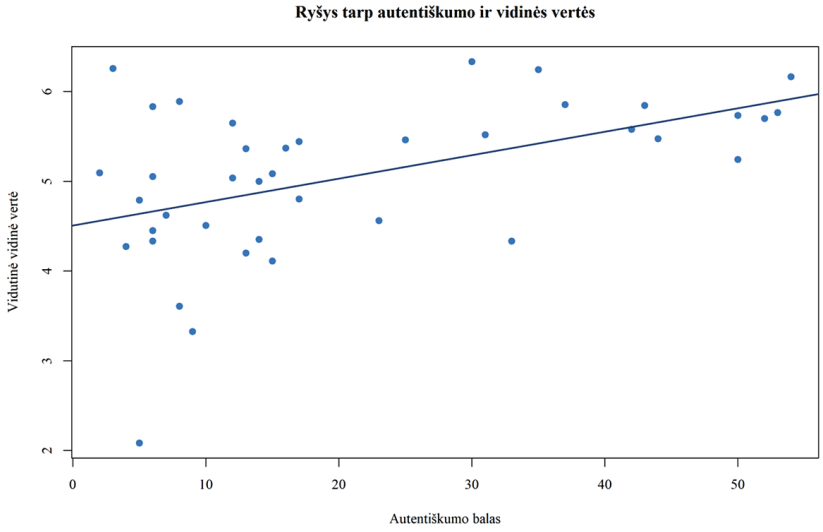
Imties aprašomoji statistika

Rodiklis	Reikšmė
Respondentų skaičius	843
Vertinimo situacijų skaičius	40
Vidutinis respondentų skaičius vienoje vertinimo situacijoje	21,08
Respondentų skaičiaus mediana	14
Standartinis nuokrypis	19,31
Mažiausias respondentų skaičius vertinimo situacijoje	3
Didžiausias respondentų skaičius vertinimo situacijoje	110

Kadangi instrumentu buvo matuojamas autentiškumas vertinimo užduoties lygmeniu, studentų apklausos duomenys taip pat buvo agreguoti vertinimo situacijos lygmeniu, apskaičiuojant vidutinius vidinės vertės balus kiekvienai vertinimo užduočiai. Koreliacinė analizė parodė vidutinio stiprumo teigiamą ryšį tarp bendro autentiškumo balo ir studentų suvokiamos vidinės vertės (r = .49, 95 % PI [.21, .70], p = .001). Papildoma analizė parodė, kad teigiamas ryšys tarp autentiškumo ir vidinės vertės buvo stebimas ir atskirų vidinės vertės teiginių lygmeniu. Stipriausias ryšys nustatytas tarp autentiškumo ir studentų teiginio „Ši vertinimo užduotis buvo įdomi“ (r = .54, p < .001). Šis rezultatas dera su autentiško vertinimo literatūra, kurioje pabrėžiama, kad studentai linkę aukščiausiai vertinti ir labiausiai mėgautis tomis užduotimis, kurios yra praktiškos ir suteikia galimybę prasmingai įsitraukti į užduoties atlikimą bei kurių atlikimo būdą bent iš dalies gali patys pasirinkti (Quinlan et al., 2025).

4 paveikslas

Ryšys tarp autentiškumo ir vidinės vertės



Agreguotų įverčių patikimumas ir jautrumo analizė

Kadangi studentų vidinės užduoties vertės įverčiai pagrindinėje analizėje buvo agreguojami vertinimo situacijos lygmeniu, papildomai įvertintas grupavimosi mastas ir grupių vidurkių patikimumas. ICC(1) siekė .173, taigi 17,3 % individualių studentų vidinės vertės balų variacijos buvo susiję su priklausymu tai pačiai vertinimo situacijai. ICC(2) siekė .689, o tai rodo pakankamą agreguotų vertinimo situacijų vidurkių patikimumą.

Siekiant įvertinti, ar vertinimo situacijos, kuriose apklausoje dalyvavo nedaug studentų, galėjo neproporcingai paveikti nustatytą ryšį, atliktos jautrumo analizės (9 lentelė). Visose 40 vertinimo situacijų nustatytas vidutinio stiprumo teigiamas ryšys tarp bendro autentiškumo balo ir vidutinės studentų vidinės užduoties vertės (r = .493, 95 % PI [.214; .697], p = .001). Pašalinus 12 vertinimo situacijų, kuriose buvo mažiau nei 10 respondentų, likusiose 28 situacijose nustatytas kiek stipresnis ryšys (r = .579, 95 % PI [.262; .783], p = .001).

Vis dėlto mažas respondentų skaičius ne visais atvejais reiškė mažą tiriamos studijų grupės aprėptį, nes dalis tyrime dalyvavusių studijų programos grupių pačios buvo labai mažos. Todėl papildomoje kontekstinėje jautrumo analizėje iš analizės pašalintos mažiau nei 10 respondentų turėjusios situacijos, išskyrus atvejus, kai visa studijų grupė buvo klasifikuota kaip labai maža (1–9 studentai). Šioje analizėje liko 34 vertinimo situacijos, o nustatytas ryšys (r = .506, 95 % PI [.202; .721], p = .002) buvo labai panašus į visos imties rezultatą. Taigi autentiškumo ir vidinės užduoties vertės ryšio dydis bei kryptis išliko stabilūs taikant skirtingus vertinimo situacijų įtraukimo kriterijus.

9 lentelė

Autentiškumo ir užduoties vertės ryšio jautrumo analizė

Analizė	*N*	*r*	95 % PI	*p*
Visos vertinimo situacijos	40	.493	[.214; .697]	.001
≥10 respondentų	28	.579	[.262; .783]	.001
≥10 respondentų arba labai maža studijų grupė	34	.506	[.202; .721]	.002

Diskusija ir išvados

Šio tyrimo tikslas buvo sukurti vieną pirmųjų kiekybinių instrumentų, skirtų vertinimo užduoties autentiškumui aukštojo mokslo kontekste įvertinti, ir atlikti pirminį jo validavimą. Gauti rezultatai suteikia pirminį empirinį pagrindą instrumentą taikyti bendram vertinimo užduoties autentiškumui vertinti: nustatytas aukštas vidinis suderinamumas ir geras vertintojų tarpusavio suderinamumas, instrumentas diferencijavo skirtingo autentiškumo vertinimo užduotis, o bendras autentiškumo balas buvo reikšmingai susijęs su studentų suvokiama vidine užduoties verte.

Vidinės instrumento struktūros analizės rezultatai rodo, kad teoriškai išskirtus autentiškumo požymius sieja stiprus bendras konstruktas. Visi 20 kriterijų turėjo teigiamas faktorines apkrovas bendrame faktoriuje, o atskirų kriterijų analizė neidentifikavo kriterijų, kurių pašalinimas reikšmingai pagerintų instrumento vidinį suderinamumą. Nors itin aukštas Cronbacho alfa koeficientas (α = .97) ir stiprios kriterijų tarpusavio koreliacijos rodo glaudų jų susietumą, šiame tyrime nenustatyta empirinio pagrindo atskirus kriterijus laikyti pertekliniais. Kartu EFA nepatvirtino septynių atskirų faktorių struktūros, todėl teoriškai išskirtas instrumento sritis šiame validavimo etape tikslinga interpretuoti kaip konceptualiai skirtingas bendro autentiškumo konstrukto sritis, o pagrindiniu instrumento rodikliu laikyti bendrą autentiškumo balą. Atsižvelgus į nedidelį vertinimo situacijų skaičių (N = 40), tikslesnis instrumento vidinės struktūros vertinimas turėtų būti tęsiamas didesnėse imtyse.

Tyrimo rezultatai taip pat atskleidė, kad autentiškumas dažnai yra susijęs su vertinimo formatu. Aukščiausiais autentiškumo balais buvo įvertintos akademinės analizės užduotys ir tyrimai bei pristatymai ir gynimai, o tradiciniai egzaminai ir kolokviumai bei testai su keliais atsakymo variantais pasižymėjo žemiausiais autentiškumo balais. Šios tendencijos dera su ankstesne literatūra, kurioje autentiškumas dažniausiai siejamas su sudėtingomis, praktiškai pritaikomomis, refleksiją ir prasmingą veiklą skatinančiomis užduotimis, o ne informacijos atkartojimu griežtai apribotame kontekste (Ashford-Rowe et al., 2014; Gulikers et al., 2004; Villarroel et al., 2018).

Reikšmingas vidutinio stiprumo ryšys tarp autentiškumo ir studentų suvokiamos vidinės vertės taip pat turi svarbių teorinių implikacijų. Šis rezultatas dera su SEVT

prieiga, pagal kurią studentų motyvaciją formuoja konkrečios mokymosi ir vertinimo užduočių bei kontekstų charakteristikos (Eccles & Wigfield, 2020; Gladstone et al., 2022). Ryšio kryptis ir dydis išliko stabilūs ir atlikus jautrumo analizes, todėl gauti rezultatai leidžia teigti, kad autentiškesnės vertinimo užduotys yra susijusios su didesne studentų suvokiama vidine užduoties verte ir pozityvesne užduoties atlikimo patirtimi.

Tyrimo ribotumai

Tyrimo rezultatus reikėtų vertinti atsižvelgus į keletą ribotumų. Nors potencialūs tyrimo dalyviai buvo kviečiami pagal iš anksto atsitiktine tvarka sudarytą eiliškumą, galutinė imtis priklausė nuo savanoriško dėstytojų ir studentų dalyvavimo. Todėl negalima atmesti saviatrankos poveikio: sutikę dalyvauti dėstytojai galėjo sistemingai skirtis nuo nesutikusiųjų, o apklausoje dalyvavę studentai – nuo joje nedalyvavusių. Tyrimas taip pat vykdytas tik dviejuose Lietuvos universitetuose ir apėmė pirmo kurso bakalauro studijų vertinimo situacijas. Pirmasis studijų kursas pasirinktas tikslingai, tačiau ši tyrimo aprėptis riboja rezultatų apibendrinimą kitoms studijų pakopoms ir instituciniams kontekstams.

Kitas ribotumas susijęs su instrumento pirminio validavimo apimtimi. Nors tyrime dalyvavo 843 studentai, instrumento analizės vienetas buvo vertinimo situacija, todėl instrumento struktūros analizė rėmėsi 40 atvejų. Toks imties dydis yra ribotas 20 kriterijų instrumento faktorinei struktūrai vertinti, todėl tiriamosios faktorinės analizės rezultatai laikytini preliminariais, o instrumento vidinė struktūra turėtų būti toliau tikrinama didesnėje ir įvairesnėje vertinimo situacijų imtyje.

Galiausiai pagrindinį vertinimo situacijų stebėjimą ir autentiškumo vertinimą atliko viena tyrėja, todėl negalima visiškai atmesti vertintojo subjektyvumo negalima, nors to išvengti buvo siekiama taikant teoriškai apibrėžtus kriterijus, struktūruotą stebėjimo ir vertinimo procedūrą bei pasitelkiant nepriklausomus išorės vertintojus. Be to, tyrimas yra koreliacinio pobūdžio, todėl nustatyti ryšiai negali būti interpretuojami kaip priežastiniai.

Literatūros sąrašas

Abramenka-Lachheb, D. V., & de Siqueira, D. A. (2021). Authentic assessments through the lenses of diversity, equity, inclusion and justice in a fully online course. *Journal of Teaching and Learning with Technology*, *11*(1). https://doi.org/10.14434/jotlt.v11i1.34591

Ashford-Rowe, K., Herrington, J., & Brown, C. (2014). Establishing the critical elements that determine authentic assessment. *Assessment & Evaluation in Higher Education*, *39*(2), 205–222. https://doi.org/10.1080/02602938.2013.819566

Baines, S., Otermans, P., Tree, D., & Worsfold, N. (2025). Measuring and mapping authentic assessment with a novel quantitative typology. *Teaching in Higher Education*, *30*(3), 663–682. https://doi.org/10.1080/13562517.2024.2424823

Ballen, C. J., Wieman, C., Salehi, S., Searle, J. B., & Zamudio, K. R. (2017). Enhancing diversity in undergraduate science: Self-efficacy drives performance gains with active learning. *CBE Life Sciences Education*, *16*(4), ar56. https://doi.org/10.1187/cbe.16-12-0344

Barral, A. M., Ardi-Pastores, V. C., & Simmons, R. E. (2018). Student learning in an accelerated introductory biology course is significantly enhanced by a flipped-learning environment. *CBE–Life Sciences Education*, *17*(3), ar38. https://doi.org/10.1187/cbe.17-07-0129

Bols, A., & Freeman, R. (2011). Engaging students in shaping their curriculum. *Educational Developments*, *12*(2), 5–9. https://www.seda.ac.uk/wp-content/uploads/2020/09/Educational-Developments-12.2.pdf

Bovill, C. (2020). Co-creation in learning and teaching: The case for a whole-class approach in higher education. *Higher Education*, *79*(6), 1023–1037. https://doi.org/10.1007/s10734-019-00453-w

Burkšaitienė, N. (2012). Promoting student learning through feedback in higher education. *Socialinių mokslų studijos*, *4*(1), 33–46. https://etalpykla.lituanistika.lt/fedora/objects/LT-LDB-0001:J.04~2012~1367182639125/datastreams/DS.002.1.01.ARTIC/content

Burkšaitienė, N. (2020). Self-assessment of ESP oral performance: A tool for learning and fostering learner awareness. *Journal of Education Culture and Society*, *11*(1), 146–162. https://doi.org/10.15503/jecs2020.1.146.162

Dall'Alba, G., & Barnacle, R. (2007). An ontological turn for higher education. *Studies in Higher Education*, *32*(6), 679–691. https://doi.org/10.1080/03075070701685130

Darling-Hammond, L. (1995). Setting standards for students: The case for authentic assessment. *The Educational Forum*, *59*(1), 14–21. https://doi.org/10.1080/00131729409336358

Dochy, F. J. R. C., & McDowell, L. (1997). Assessment as a tool for learning. *Studies in Educational Evaluation*, *23*(4), 279–298. https://doi.org/10.1016/S0191-491X(97)86211-6

Eccles, J. S., & Wigfield, A. (2020). From expectancy-value theory to situated expectancy-value theory: A developmental, social cognitive, and sociocultural perspective on motivation. *Contemporary Educational Psychology*, *61*, 101859. https://doi.org/10.1016/j.cedpsych.2020.101859

Eccles, J. S., & Wigfield, A. (2024). The development, testing, and refinement of Eccles, Wigfield, and colleagues' situated expectancy-value model of achievement performance and choice. *Educational Psychology Review*, *36*(2), 51. https://doi.org/10.1007/s10648-024-09888-9

Forsyth, H., & Evans, J. (2019). Authentic assessment for a more inclusive history. *Higher Education Research & Development*, *38*(4), 748–761. https://doi.org/10.1080/07294360.2019.1581140

Freeman, S., Eddy, S. L., McDonough, M., Smith, M. K., Okoroafor, N., Jordt, H., & Wenderoth, M. P. (2014). Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(23), 8410–8415. https://doi.org/10.1073/pnas.1319030111

Gaspard, H., Dicke, A.-L., Flunger, B., Häfner, I., Brisson, B. M., Trautwein, U., & Nagengast, B. (2016). Side effects of motivational interventions? Effects of an intervention in math classrooms on motivation in verbal domains. *AERA Open*, *2*(2), 2332858416649168. https://doi.org/10.1177/2332858416649168

Gaspard, H., Häfner, I., Parrisius, C., Trautwein, U., & Nagengast, B. (2017). Assessing task values in five subjects during secondary school: Measurement structure and mean level differences across grade level, gender, and academic subject. *Contemporary Educational Psychology*, *48*, 67–84. https://doi.org/10.1016/j.cedpsych.2016.09.003

Gaspard, H., Wigfield, A., Jiang, Y., Nagengast, B., Trautwein, U., & Marsh, H. W. (2018). Dimensional comparisons: How academic track students' achievements are related to their expectancy and value beliefs across multiple domains. *Contemporary Educational Psychology*, *52*, 1–14. https://doi.org/10.1016/j.cedpsych.2017.10.003

Gladstone, J. R., Wigfield, A., & Eccles, J. S. (2022). Situated expectancy-value theory, dimensions of engagement, and academic outcomes. In A. L. Reschly & S. L. Christenson (Eds.), *Handbook of research on student engagement* (pp. 57–76). Springer International Publishing. https://doi.org/10.1007/978-3-031-07853-8_3

Gulikers, J. T. M., Bastiaens, T. J., & Kirschner, P. A. (2004). A five-dimensional framework for authentic assessment. *Educational Technology Research and Development*, *52*(3), 67–86. https://doi.org/10.1007/BF02504676

Hanafin, J., Shevlin, M., Kenny, M., & Neela, E. M. (2007). Including young people with disabilities: Assessment challenges in higher education. *Higher Education*, *54*(3), 435–448. https://doi.org/10.1007/s10734-006-9005-9

Healey, M., Bradley, A., Fuller, M., & Hall, T. (2006). Listening to students: The experiences of disabled students of learning at university. In M. Adams & S. Brown (Eds.), *Towards inclusive learning in higher education: Developing curricula for disabled students* (pp. 32–43). Routledge. https://doi.org/10.4324/9780203088623

Ismail, S. M., Nikpoo, I., & Prasad, K. D. V. (2023). Promoting self-regulated learning, autonomy, and self-efficacy of EFL learners through authentic assessment in EFL classrooms. *Language Testing in Asia*, *13*(1), 27. https://doi.org/10.1186/s40468-023-00239-z

Kaminskienė, L., & Stasiūnaitė, E. (2013). Validity of assessment and recognition of non-formal and informal learning achievements in higher education. *Aukštojo mokslo kokybė*, *1,* 28–47. https://doi.org/10.7720/1822-1645.10.2

Kearney, S., & Perkins, T. (2014). Engaging students through assessment: The success and limitations of the ASPAL (Authentic Self and Peer Assessment for Learning) model. *Journal of University Teaching and Learning Practice*, *11*(3), Article 2. https://doi.org/10.53761/1.11.3.2

Kent-Water, J., Seago, O., & Smith, L. (2018). *A compendium of assessment techniques in higher education: From students' perspectives* (D. S. Pugh, Ed.). Leeds Institute for Teaching Excellence. https://teachingexcellence.leeds.ac.uk/wp-content/uploads/sites/89/2018/10/PUGHcompendiumcomplete.pdf

LAMA BPO. (2024). *2024 m. Bendrojo priėmimo į Lietuvos aukštąsias mokyklas apžvalga*. https://lamabpo.lt/wp-content/uploads/2024/09/2024-m.-bendrojo-priemimo-rezultatu-apzvalga-1.pdf

Lenkauskaitė, J., Bubnys, R., Masiliauskienė, E., & Malinauskienė, D. (2021). Participation in the assessment processes in problem-based learning: Experiences of the students of social sciences in Lithuania. *Education Sciences*, *11*(11), Article 678. https://doi.org/10.3390/educsci11110678

Lenkauskaitė, J., Colomer, J., & Bubnys, R. (2020). Students' social construction of knowledge through cooperative learning. *Sustainability*, *12*(22), Article 9606. https://doi.org/10.3390/su12229606

Lietuvos Respublikos švietimo ir mokslo ministerija. (2014). *Valstybinė švietimo 2013–2022 metų strategija*. Švietimo ir mokslo ministerijos Švietimo aprūpinimo centras.

Lietuvos Respublikos švietimo, mokslo ir sporto ministerija. (2026). *2021–2030 m. Švietimo plėtros programa*. https://smsm.lrv.lt/public/canonical/1778822126/7693/1.%20SPP_Nutarimas_2026.05.13.pdf

Lübeck, G., Seery, M. K., & Ryan, B. J. (2022). Eight steps to facilitating more equitable education in undergraduate sciences. *The Biochemist*, *44*(5), 21–26. https://doi.org/10.1042/bio_2022_123

Lübeck, G., Seery, M. K., & Ryan, B. J. (2025). Supporting female and gender diverse students in undergraduate STEM courses through enhancing self-efficacy. *Teaching in Higher Education, 30*(7), 1663–1686. https://doi.org/10.1080/13562517.2025.2449660

McArthur, J. (2023). Rethinking authentic assessment: Work, well-being, and society. *Higher Education*, *85*(1), 85–101. https://doi.org/10.1007/s10734-022-00822-y

Miller, E., & Konstantinou, I. (2022). Using reflective, authentic assessments to embed employability skills in higher education. *Journal of Work-Applied Management*, *14*(1), 4–17. https://doi.org/10.1108/JWAM-02-2021-0014

Muenks, K., Miller, J. E., Schuetze, B. A., & Whittaker, T. A. (2023). Is cost separate from or part of subjective task value? An empirical examination of expectancy-value versus expectancy-value-cost perspectives. *Contemporary Educational Psychology*, *72*, 102149. https://doi.org/10.1016/j.cedpsych.2023.102149

Näsström, G., Andersson, C., Granberg, C., Palm, T., & Palmberg, B. (2021). Changes in student motivation and teacher decision making when implementing a formative assessment practice. *Frontiers in Education*, *6*, Article 616216. https://doi.org/10.3389/feduc.2021.616216

National Forum for the Enhancement of Teaching and Learning in Higher Education. (2017). *Authentic assessment in Irish higher education*. https://doi.org/10.6084/m9.figshare.4786348

Putwain, D. W. (2008). Test anxiety and GCSE performance: The effect of gender and socio-economic background. *Educational Psychology in Practice*, *24*(4), 319–334. https://doi.org/10.1080/02667360802488765

Quinlan, K. M., Sellei, G., & Fiorucci, W. (2025). Educationally authentic assessment: Reframing authentic assessment in relation to students' meaningful engagement. *Teaching in Higher Education*, *30*(3), 717–734. https://doi.org/10.1080/13562517.2024.2394042

Raudienė, I. (2018). Formuojamojo vertinimo raiška Lietuvos mokyklose: Ką rodo nacionaliniai mokinių pasiekimų tyrimai? *Socialinis ugdymas*, *50*(3), 103–116. https://doi.org/10.15823/su.2018.22

Raudienė, I., Kaminskienė, L., & Cheng, L. (2022). The education and assessment system in Lithuania. *Assessment in Education: Principles, Policy & Practice*, *29*(3), 383–394. https://doi.org/10.1080/0969594X.2022.2103516

Raudienė, I., & Mažeikienė, N. (2023). Delving into pragmatic horizons: Critical ethnographic research on the assessment culture in school. *Pedagogika (Vilnius, Lithuania)*, *149*(1), 35–52. https://doi.org/10.15823/p.2023.149.2

Raymond, J., Homer, C., Smith, R., & Gray, J. (2013). Learning through authentic assessment: An evaluation of a new development in the undergraduate midwifery curriculum. *Nurse Education in Practice*, *13*(5), 471–476. https://doi.org/10.1016/j.nepr.2012.10.006

Sambell, K., & Brown, S. (2022). *Assess for success – broadening conception of authentic assessment*. Maynooth University. https://www.maynoothuniversity.ie/sites/default/files/assets/document//NUIM_T3%20_O4_Broadening%20conception%20of%20authentic%20assessment%20%281%29.pdf

Serrano, M. M., O'Brien, M., Roberts, K., & Whyte, D. (2018). Critical pedagogy and assessment in higher education: The ideal of 'authenticity' in learning. *Active Learning in Higher Education*, *19*(1), 9–21. https://doi.org/10.1177/1469787417723244

Sokhanvar, Z., Salehi, K., & Sokhanvar, F. (2021). Advantages of authentic assessment for improving the learning experience and employability skills of higher education students: A systematic literature review. *Studies in Educational Evaluation*, *7*. 101030. https://doi.org/10.1016/j.stueduc.2021.101030

Šliogerienė, J. (2012). Students' reflection analysis in portfolio-based learning. *Santalka: filologija, edukologija, 20*(1), 98–105. https://doi.org/10.3846/cpe.2012.10

Swaffield, S. (2011). Getting to the heart of authentic assessment for learning. *Assessment in Education: Principles, Policy & Practice*, *18*, 433–449. https://doi.org/10.1080/0969594X.2011.582838

Tai, J., Ajjawi, R., Bearman, M., Boud, D., Dawson, P., & Jorre de St Jorre, T. (2023). Assessment for inclusion: Rethinking contemporary strategies in assessment design. *Higher Education Research & Development*, *42*(2), 483–497. https://doi.org/10.1080/07294360.2022.2057451

Tang, X., Lee, H. R., Wan, S., Gaspard, H., & Salmela-Aro, K. (2022). Situating expectancies and subjective task values across grade levels, domains, and countries: A network approach. *AERA Open*, *8*, 23328584221117168. https://doi.org/10.1177/23328584221117168

Tanner, D. E. (2001). Authentic assessment: A solution, or part of the problem? *The High School Journal*, *85*(1), 24–29. https://doi.org/10.1353/hsj.2001.0020

Villarroel, V., Boud, D., Bloxham, S., Bruna, D., & Bruna, C. (2020). Using principles of authentic assessment to redesign written examinations and tests. *Innovations in Education and Teaching International*, *57*(1), 38–49. https://doi.org/10.1080/14703297.2018.1564882

Wigfield, A., & Eccles, J. S. (2000). Expectancy–value theory of achievement motivation. *Contemporary Educational Psychology*, *25*(1), 68–81. https://doi.org/10.1006/ceps.1999.1015

Žibėnienė, G. (2014). *Studijų rezultatų vertinimas: Teorinis ir empirinis pagrindimas.* Mokslo studija. Mykolo Romerio universitetas.

What Makes an Assessment Task Authentic? Development and Initial Validation of an Instrument for Measuring Assessment Authenticity

Gintarė Lübeck

Vytautas Magnus University, Faculty of Social Sciences, Department of Sociology, 66 Jonavos St., LT-44138 Kaunas, Lithuania, gintare.luebeck@vdu.lt

Summary

The aim of this study was to develop and initially validate an instrument for assessing the authenticity of higher education assessment tasks. Drawing on established theoretical models, the instrument operationalised authenticity through seven theoretically derived domains. Data from 40 first-year assessment situations across the two largest Lithuanian universities demonstrated high internal consistency for the overall instrument ($\alpha = .97$) and good inter-rater reliability. Exploratory factor analysis indicated a strong general authenticity factor, although these findings are interpreted as preliminary due to the limited number of assessment situations. The instrument differentiated between assessment formats, with multiple-choice tests and traditional examinations receiving the lowest authenticity scores, and research tasks, presentations, and project-based assessments receiving the highest. To examine the instrument's practical applicability, the relationship between assessment authenticity and students' perceived intrinsic value of assessment tasks was analysed, revealing a moderate positive correlation ($r = .49$, $p = .001$). The findings provide preliminary evidence supporting the reliability and theoretical meaningfulness of the instrument as a measure of overall assessment authenticity in higher education, while further validation with larger and more diverse samples of assessment situations is required.

Keywords: *authentic assessment, higher education, assessment authenticity measurement, perceived task value.*

Gauta 2026 06 04 / Received 04 06 2026
Priimta 2026 09 07 / Accepted 07 09 2026