

Зборник Института за педагошка истраживања

Година 57 • Број 1 • Јун 2025 • 5–28

УДК: 37.091.3::51

37.091.26-057.874(497.11)"2013/2019"

Примљено 16.12.2024; прихваћено 26.05.2025.

ISSN 0579-6431

ISSN 1820-9270 (Online)

DOI: 10.2298/ZIPI2501005V

Оригинални научни рад

СИМПСОНОВ ПАРАДОКС У ОЦЕНАМА ИЗ МАТЕМАТИКЕ: СТУДИЈА СЛУЧАЈА ИЗ ОСНОВНИХ ШКОЛА У СРБИЈИ*

Срђан Вербић** • ORCID: 0000-0003-1908-1421

ФЕФА, Универзитет Метрополићан, Београд, Србија

Марија Каплар • ORCID: 0000-0002-0920-8276

Факултет техничких наука, Универзитет у Новом Саду, Србија

Зорана Луžанин • ORCID: 0000-0002-7215-2252

Природно-математички факултет, Универзитет у Новом Саду, Нови Сад, Србија

Драгана Трнавац • ORCID: 0000-0002-0919-5049

Природно-математички факултет, Универзитет у Новом Саду, Нови Сад, Србија

Бранислав Ранђеловић • ORCID: 0000-0002-0643-0955

Електронски факултет, Универзитет у Нишу, Ниш, Србија

АПСТРАКТ

У раду је представљено емпиријско истраживање у коме се анализира успех из математике петнаестогодишњака на крају основног образовања у Србији. У анализи су коришћени резултати теста из математике на завршном испиту и оцене из математике. Истраживање обухвата период од седам година (2013–2019) и више од 440.000 ученика. Емпиријски налази потврђују да девојчице имају боље оцене из математике и постижу више резултате на националном тестирању. Међутим, појава Симпсоновог парадокса указује на то да постоји пристрасност у оцењивању у корист девојчица. Резултати додатно показују да девојчице у већој мери исказују спремност да покушају да реше задатке отвореног типа, чак и када нису успешне, док дечаки те задатке чешће остављају без одговора током тестирања. Квалитативни део истраживања обухватио је фокус групу са наставницима како би се боље разумели њихови ставови и искуства у вези са родним разликама у оцењивању.

Кључне речи:

оцењивање, национални испит, математика, родна пристрасност, Симпсон парадокс.

* Цитирати овај чланак на следећи начин: Verbić, S., Kaplar, M., Lužanin, Z., Trnavac, D. & Randelović, B. (2025). Simpsonov paradoks u ocenama iz matematike: studija slučaja iz osnovnih škola u Srbiji. *Zbornik Instituta za pedagoška istraživanja*, 57(1), 5–28. DOI: 10.2298/ZIPI2501005V.

** Мејл: sverbic@fefa.edu.rs

■ УВОД

Упис у средње школе, додељивање ученичких стипендија и доношење одлука о даљем образовању најчешће зависе од оцена остварених у основној школи и/или резултата спољашњих тестирања. Стога не изненађује чињеница да постоји изражено интересовање за школске оцене и постигнуће ученика на спољашњим проверама знања. Као последица тога, значајан број истраживања посвећен је анализи оцена и образовних постигнућа, при чему су бројне научне студије детаљно испитивале разлике у постигнућима међу половима. Посебан акценат стављен је на анализу разлика у оценама и постигнућима, са фокусом на „родни јаз у постигнућима” и „родни јаз у оцењивању”.

Термин „родни јаз”, такође познат као „родни јаз у постигнућима” најчешће означава разлику у учинку између девојчица и дечака. Овај феномен се обично квантификује мерењем разлика у њиховим резултатима на тестовима (Protivínský & Münich, 2018). С друге стране, „родни јаз у оцењивању” упућује на разлике у оценама које добијају ученици и ученице сличних способности, при чему су оцене најчешће резултат неанонимног оцењивања наставника, за разлику од анонимних тестова.

Ово истраживање је усмерено на испитивање присуства и интензитета родног јаза и јаз у оцењивању из математике у оквиру образовног система у Републици Србији. Предност овог рада препознаје се у обимним подацима, који се односе на седмогодишњи период, од 2013. до 2019. године, и више од шездесет хиљада ученика за сваку школску годину. Поред тога, анализирани су и задаци на које ученици нису ни покушали да одговоре, односу на пол и тип задатка, како би се дубље сагледао родни јаз. Ова анализа омогућава разумевање образаца у избегавању одговора на задатке, што може указивати на смањену мотивацију или невољност да се уложи додатни напор у решавање математичких проблема.

Бројна претходна истраживања у различитим образовним системима доследно указују на постојање родног јаз у оцењивању из математике, најчешће у корист девојчица (Protivínský & Münich, 2018). Према постојећим студијама, на овај јаз може утицати низ фактора, међу којима се као један од важнијих истиче интеракција између наставника и ученика (Protivínský & Münich, 2018). Имајући то у виду, у оквиру нашег истраживања спроведена је и квалитативна анализа путем фокус групе у којој су учествовали наставници

основних школа, што представља иновацију у односу на раније методолошке приступе. Овај квалитативни део истраживања осмишљен је с циљем дубљег разумевања ставова и искустава наставника у вези са родним разликама у оцењивању.

Структура рада организована је на следећи начин: у другом делу приказан је преглед релевантне литературе на којој се истраживање темељи. Трећи део садржи преглед институционалног оквира, коришћених података и примењене методологије. Четврти део доноси резултате квантитативног истраживања, у чијем су фокусу су родне разлике у различитим облицима оцењивања. У петом делу приказани су налази добијени из квалитативне студије. На крају, шести део обухвата дискусију добијених резултата и разматрање потенцијалних импликација за унапређивање образовне праксе.

■ ПРЕГЛЕД ЛИТЕРАТУРЕ

Родни јаз и родни јаз у оцењивању

Истраживања која се баве родним разликама у академским способностима доследно указују на сличне обрасце, без обзира на примењени метод оцењивања. Посматрано уопштено, примећује се да девојчице имају боље школске оцене, док дечаки остварују више резултате на стандардизованим тестовима. Студија коју је конципирао Теријер (Terrier, 2020) показала је да наставници у нижим средњим школама у Француској чешће додељују више оцене девојчицама. Такође, истраживање које су реализовали Гuez и сарадници (Guez et al., 2020) испитивало је постигнућа дечака и девојчица из француског језика и математике, користећи три различита типа оцењивања: процене наставника, националне испите и стандардизоване тестове знања. Резултати на репрезентативном узорка ученика нижих средњих школа у Француској показују да девојчице у просеку постижу бољи успех у француском језику, док су у математици постизале слабије постигнуће од дечака. Ђофре и сарадници (Giofrè et al., 2020) наводе да су и у италијанским школама дечаки остваривали боље резултате из математике у поређењу са девојчицама, док је обрнут тренд уочен у читању. Ова разлика у постигнућима расла је прогресивно од 2. до 8. разреда. Ди Либерто и сарадници (Di Liberto et al., 2022) упоредили су школске

оцене које додељују наставници са резултатима стандардизованих тестова код ученика у Италији и открили да су дечаци били оцењени неповољније од девојчица, и у математици и у језику. Сличне обрасце примећују Фалк и Напер (Falch & Naper, 2013) у Норвешкој. Њихово истраживање показује да девојчице добијају значајно више оцене од дечака када наставници процењују исте способности. Ова пристрасност у оцењивању у корист девојчица присутна је и у језицима и у математици. Противински и Минх (Protivínský & München, 2018) су истраживали родну пристрасност у оцењивању у Чешкој анализирајући школске оцене у поређењу са резултатима пријемних испита из математике и матерњег језика за петнаестогодишњаке. Њихови налази указују на изражену пристрасност у корист девојчица. Такође, Ферман и Фонтес (Ferman & Fontes, 2022) су доказали да постоји родна пристрасност у оцењивању из математике на узорку ученика нижих и виших разреда средњих школа у Бразилу, при чему су дечаци били у неповољнијем положају. Када је реч о Србији, резултати међународног истраживања TIMSS из 2003. године показали су да су петнаестогодишње девојчице у просеку биле успешније од дечака у математици. Слични резултати забележени су и 2019. године, иако је тадашњи циклус обухватио само ученике четвртог разреда (Đerić et al., 2021; Mullis et al., 2020). С друге стране, резултати PISA тестирања показују да су дечаци у Србији остваривали боље резултате у математици од девојчица, иако та разлика није увек била статистички значајна (OECD, 2023; Videnović & Ćaprić, 2020).

Џанга и Де Ђоанис (Zanga & De Gioannis, 2023) спровели су темељно истраживање које обухвата 37 објављених студија с циљем да идентификују особине ученика које могу довести до пристрасности или дискриминације у оцењивању. Већина тих студија фокусирана је на родну пристрасност, а више од половине (52%) потврђује постојање родно дискриминаторних образаца у оцењивању. Налази указују на то да наставници, у просеку, дају више оцене девојчицама него што би то било у складу са резултатима тестова који се оцењују „наслепо”. Супротно томе, дечаци често добијају ниже оцене у школском окружењу, упркос сличним резултатима на стандардизованим тестовима. Ипак, треба напоменути да мањи број студија бележи и случајеве пристрасности у корист дечака (Guttmann & Boudo, 1988; Lavy & Sand, 2015). Супротно томе, истраживање које су у Индији спровели Ракшит и Саху (Rakshit & Sahoo, 2023) није утврдило присуство пристрасности у наставничком

оцењивању из математике. Слично томе, истраживања у Шведској (Hinnerich et al., 2011) као и још једно истраживање у Индији (Hanna & Linden, 2012) указала су на одсуство родног јазу у оцењивању. Наведена испитивања указују на честу појаву родне пристрасности у оцењивању у корист девојчица, посебно у области математике. Такође, јасно је да је овај феномен широко распрострањен и присутан у различитим образовним системима широм света.

Ставови наставника о полу и математици

Праксе оцењивања које примењују наставници математике суштински су значајни јер често утичу на мотивацију ученика, њихове ставове и уверења. Бројна истраживања која се баве наставничким праксама оцењивања показала су да школске оцене не одражавају искључиво академска постигнућа ученика, већ такође откривају постојање пристрасности наставника према одређеним друштвеним категоријама. Лави (Lavy, 2008) је анализирао основне узроке родног јазу у оцењивању, испитујући да ли он проистиче из понашања наставника или ученика. Његови налази указују на то да је јаз блиско повезан са карактеристикама наставника, укључујући њихов пол, старост, радно искуство и величину породице, што имплицира да је понашање наставника, пре него понашање ученика, вероватнији узрок родног јазу у оцењивању.

Узимајући у обзир емпиријске доказе о пристрасности у оцењивању ученика из математике, поставља се питање да ли међу наставницима постоје родно заснована уверења (Dersch et al., 2022; Li, 1999; Lindner et al., 2022; OECD, 2015; Sumpter, 2016). Ли (Li, 1999) тврди да досадашња литература, иако не пружа коначне доказе, сугерише да наставници често имају различита очекивања и уверења у односу на ученике и ученице. Они су склони да стереотипно повезују математику са мушким полом, што се препознаје у њиховој склоности да прецењују математичке способности дечака, имају виша очекивања од њих и показују позитивније ставове према њима.

Дерш и сарадници (Dersch et al., 2022) идентификују три честа погрешна уверења код наставника: 1) *теорија емпатије и систематизације* тврди да су дечаци природно систематичнији, а девојчице емпатичније, што дечацима наводно даје предност у математици; 2) *хипотеза компензације девојчица* се заснива на претпоставци да дечаци морају да уложе мање труда да би постигли исте резултате као девојчице, јер поседују „природни” таленат за

математику и 3) *хитиоїеза некоїїензабилносїи девоїчица* која сугерише да, иако се девоїчице више труде, и даље заостају у математичким вештинама у поређењу са дечацима.

Тидеман (Tiedemann, 2002) примећује да наставници неуспех девоїчица у математици чешће приписују недостатку талента, док неуспехе дечака обично приписују недовољном труду, што указује на родну пристрасност у поступку оцењивања. У експерименталним условима, када наставници претпоставе да су постигнуће оствариле девоїчице, имају тенденцију да потцене те резултате у поређењу са резултатима дечака, као што је показано у студијама Авицора и сарадника (Avitzour et al., 2020) и Холдера и Кеселса (Holder & Kessels, 2017).

Ови налази стоје у контрасту са бројним истраживањима која указују на јаз у оцењивању у корист девоїчица. С једне стране, постоје наставничка уверења која фаворизују дечаке, док с друге стране, оцене које додељују наставници често су у корист девоїчица. Имајући то у виду, у нашем истраживању настојали смо да додатно расветлимо овај феномен кроз квалитативну анализу фокус групе, која је усмерена на испитивање ставова наставника о родно заснованом оцењивању.

Формат задатака и изостављање одговора

Током последњих деценија формати задатака коришћени у оцењивању били су предмет интензивног истраживања, што потврђује велики број емпиријских студија и накнадних метаанализа. Вишеструка емпиријска истраживања показала су да успешност ученика варира у зависности од формата задатка. Ова истраживања доследно указују на то да дечаци у просеку остварују боље резултате на задацима са понуђеним одговорима (*multiple-choice*), док девоїчице постижу боље резултате на задацима отвореног типа који захтевају објашњења или елаборацију одговора (Beller & Gafni, 2000; Birenbaum & Feldman, 1998; Garner & Engelhard, 1999; Riener & Wagner, 2018; Wester & Henriksson, 2000). Међутим, Белер и Гафни (Beller & Gafni, 2000) су кроз две одвојене студије дошли до опречних налаза. У једној студији родне разлике биле су израженије у задацима са понуђеним одговорима него у задацима отвореног типа, док су у другој студији разлике биле веће управо код задатака отвореног типа, у корист дечака. Ови контрадикторни резултати доводе у питање широко прихваћено мишљење да девоїчице увек постижу боље резултате на задацима отвореног типа, а тиме и претпоставку да се родне разлике у математичким постигнућима

могу објаснити искључиво форматом задатка. Важно је напоменути да су Белер и Гафни (Beller & Gafni, 2000) посебно нагласили улогу тежине задатка у објашњавању родних разлика у постигнућима. Њихови налази упућују на то да комплексност задатка има значајнији утицај на варијације у успеху међу половима него сам формат задатка. Конкретно, примећено је да дечаци показују боље резултате у поређењу са девојчицама како се тежина задатка повећава (Beller & Gafni, 2000). Склоност ка изостављању одговора, односно неодговарању на задатке, такође показује разлике у зависности од пола (Riener & Wagner, 2017). Фон Шрадер и Ензли (Von Schrader & Ansley, 2006) спровели су истраживање како би анализирали родне разлике у учесталости прескакања задатака на тестовима са понуђеним одговорима. Открили су да девојчице чешће прескачу математичке задатке, док су дечаци склонији да изоставе одговоре из задатака који се односе на вокабулар и читање. Овај образац се подудара са општим образовним трендовима, према којима девојчице генерално постижу боље резултате у читању, док дечаци остварују нешто боље резултате у математици – што потврђују и налази Бекера и Форсајта (Becker & Forsyth, 1990) и Хана и Хувера (Han & Hoover, 1994). Сходно томе, у многим случајевима, родна група која је чешће прескакала одговоре била је она која је имала слабије резултате у датом предмету (Von Schrader & Ansley, 2006). Ринера и Вагнер (Riener & Wagner, 2017) су показали да су девојчице склоније да прескоче задатке, посебно оне које сматрају тешким. Међутим, ова родна разлика нестаје када се уведу спољашње мотивационе награде (Riener & Wagner, 2017).

■ ПОДАЦИ И МЕТОДОЛОГИЈА

Истраживачки дизајн састоји се из два узастопна дела. Први део обухвата квантитативну анализу фокусирану на резултате завршног испита и школске оцене ученика, док други део подразумева квалитативно истраживање у коме су учествовали наставници математике.

Квантитативно истраживање

У оквиру ове студије спроведена је секундарна анализа резултата тестова из математике на завршном испиту и завршних оцена ученика на крају обавезног

основног образовања током седам узастопних година, у периоду пре пандемије COVID-19 (2013–2019). Овај временски оквир одабран је како би се избегле могуће неусклађености настале услед промена у начину реализације наставе и оцењивања током пандемије. У анализираном периоду (2013–2019) завршни испит у Србији састојао се од три теста: српски језик (или други матерњи језик), математика и комбиновани тест (биологија, историја, географија, физика, хемија). Сви ученици који завршавају осми разред основне школе, обично узраста између 14,5 и 15,5 година, морају да полажу завршни испит. Основна сврха ових тестова јесте селективна за упис у средње школе. Укупан број ученика који су у анализираном периоду полагали завршни испит износио је 446.814, при чему се број ученика по години кретао између 62.618 и 65.223. Удео девојчица међу испитаницима кретао се између 48,2% и 48,9%.

Тест из математике на завршном испиту у Србији се доследно састоји од 20 задатака које је потребно решити у року од два сата. Задаци покривају пет области: бројеви и рачунање, алгебра и функције, геометрија, мерење и анализа података. Ови задаци се класификују на три нивоа тежине: основни, средњи и напредни, и представљени су у различитим форматима. Током посматраних седам година спровођења испита коришћено је пет типова задатака: задаци са понуђеним одговорима (23 задатка), задаци за допуну (3 задатка), кратки одговори (36 задатака), задаци отвореног типа (74 задатка) и задаци упаривања (4 задатка). Сваки задатак се бодује једним поеном и укупан скор на тесту представља збир поена са свих задатака и креће се од 0 до максимално 20 поена.

Школске оцене у Србији се крећу у распону од 1 (недовољан) до 5 (одличан). На крају школске године ученици добијају закључну, односно завршну оцену из сваког предмета, која се уноси у сведочанство о завршеном образовању. Ученици са закљученом оценом 1 из било ког предмета полажу поправни испит и нису укључени у ово истраживање. Укупан број бодова који се користи за рангирање ученика при упису у средње школе представља линеарну комбинацију резултата теста и општег школског успеха у шестом, седмом и осмом разреду.

Истраживања јаза у оцењивању најчешће се спроводе упоређивањем школских оцена са резултатима спољашњег тестирања. Иако су обе врсте оцењивања усмерене на процену академског напретка ученика, између њих постоје значајне разлике (Marcenaro-Gutierrez et al., 2023). У спољашњем

тестирању се користе стандардизовани тестови које оцењују екстерни евалуатори анонимно, обезбеђујући упоредивост резултата свих ученика и конзистентност међу школама (Marcenaro-Gutierrez et al., 2023; Willingham et al., 2002). Ови тестови углавном мере компетенције ученика, односно њихове вештине и способности (Hanushek & Woessmann, 2012), док се оцењивање од стране наставника најчешће ослања на садржајну процену знања (Borghans et al., 2016; Lauermann et al., 2020; Marcenaro-Gutierrez et al., 2023).

Врсте испитивања, односно вредновања, се такође значајно разликују – спољашњи тестови се сматрају испитима високог ризика (*high-stakes*), док се школско оцењивање сврстава у оцењивање ниског ризика (*low-stakes*). Још једна кључна разлика односи се на време тестирања: у школама се знање тестира непосредно након учења, док код спољашњих тестова постоји временска дистанца. Такође, наставничко оцењивање се базира на вишемесечном праћењу, док се спољашње тестирање састоји од једног теста, који се полаже само једном. Упркос тим разликама, у литератури се сугерише да је поређење ова два типа оцењивања у одређеној мери могуће и да може дати корисне увиде у постојање јаза у оцењивању (Falch & Naper, 2013; Lavy, 2008; Matějů & Smith, 2014). У многим образовним системима како школске оцене, тако и резултати спољашњих тестова користе се у процесу селекције и уписа на наредне нивое образовања. Будући да таква пракса постоји и у Србији, ова студија упоређује два облика оцењивања како би идентификовала могуће разлике.

Подаци о резултатима завршног испита из математике преузети су из базе Завода за вредновање квалитета образовања и васпитања, који је задужен за припрему и спровођење завршног испита. За сваког ученика који је учествовао на испиту забележени су резултати за сваки задатак, као и закључне оцене из математике, физике, хемије, биологије, историје и географије, уз податке о полу и школи коју је ученик похађао. Подаци су достављени у потпуности анонимно чиме је обезбеђена заштита приватности и поверљивост информација.

За тестирање хипотеза, поред дескриптивне статистике, коришћени су одговарајући статистички тестови, укључујући t-тест и тест једнакости пропорција. Због великог обима узорка, примењени су и показатељи величине ефекта, попут Коеновог d и Коеновог h . Статистичка анализа и визуализација података спроведене су коришћењем софтверског алата R.

Квалитативно истраживање

Фокус група је формирана како би се боље разумеле родне разлике у школском постигнућу ученика, нарочито у вези са оценама и успехом (Adler et al., 2019; Lindqvist et al., 2020). Постављена су два конкретна питања: 1) Које су опсервације наставника у вези са родним разликама у оценама и постигнућима ученика из математике? 2) Како наставници тумаче или објашњавају своја запажања?

У истраживању је учествовало седам наставника са различитим нивоима професионалног искуства, који су запослени у основним школама током другог полугодишта школске 2022/2023. године, из четири различита школска округа. Сви учесници имали су најмање пет година наставничког искуства у периоду који обухвата квантитативно истраживање. Интервју фокус групе спроведен је у присуству једног главног модератора и два помоћна модератора. На почетку разговора наставници су замољени да изнесу своја запажања о родним разликама у постигнућима и оценама ученика, ослањајући се на лична искуства из праксе. Прво питање које је истраживач поставио односило се на то да ли примећују постојање родног јаза и јазу у оцењивању у сопственој наставној пракси, као и да изнесу своје мишљење с тим у вези и да објасне те појаве. Након тога, учесницима су приказани квантитативни подаци који указују на то да девојнице у просеку добијају више оценое од дечака. Конкретно, наставницима је презентован графички приказ резултата добијених анализом квантитативних података, који јасно показује постојање јазу у оцењивању, а затим су замољени да изнесу своја тумачења таквих налаза.

Модератор је потом усмеравао разговор кроз различите подтеме, које су произашле спонтано из дискусије, а не на основу унапред дефинисаних питања. Ове подтеме су обухватале различита питања, укључујући разлике у расподели оцена, узроке јазу у оцењивању и утицај оцена на будуће образовне и професионалне изборе. Дискусија је реализована у синхронизованом онлајн окружењу преко дискусионог форума. Ова метода је изабрана јер подстиче рефлексивно промишљање и међусобну размену коментара (Newell et al., 2002), као и учешће особа које су у непосредним (уживо) интеракцијама мање склоне изношењу ставова (Groth, 2006). Разговор је трајао два сата. Три истраживача су независно анализирали и транскрибовали све одговоре учесника током фокус групе. Након дискусије, истраживачи су заједнички ускладили белешке како би креирали целовиту и обједињену транскрипцију.

Тематско кодирање података спроведено је индуктивним приступом, при чему су се кодови развијали директно из садржаја, а не на основу унапред дефинисаних категорија. Ради обезбеђивања поузданости процеса кодирања, три истраживача су независно формулисала почетне кодове, а затим су се састали како би упоредили, дискутовали и усагласили своја тумачења. Све несугласице су решене дискусијом док није постигнут потпуни консензус. Иако формални показатељи међукодерске сагласности, попут Коенове капе, нису израчунавани, сам процес сарадничког кодирања обезбедио је довољну интерну конзистентност. За структурисање података коришћена је матрица развијена према оквиру Мајлса и Хубермана (Miles & Huberman, 1994), што је омогућило идентификацију понављајућих тема и односа између наставничких опсервација и њихових тумачења. Матрица је укључивала како запажања наставника, тако и објашњења која су они давали у вези са тим запажањима. Подаци су у оквиру матрице систематски категорисани по тематским целинама које су се развиле током дискусије.

РЕЗУЛТАТИ

Оцене и постигнућа на завршном испиту

Табела 1 и 2 приказују просечне вредности завршних оцена из математике и постигнућа на тесту из математике на завршном испиту за дечаке и девојчице. Релативна разлика између средњих вредности приказана је помоћу Коеновог d . Током свих анализираних година, девојчице су биле успешније од дечака и у завршним оценама и у резултатима тестирања. Уочљиво је да је просечна вредност завршне оцене из математике прилично стабилна и да се не мења значајно током година (Слика 1). Такође, опажа се да је релативна разлика у завршним оценама између дечака и девојчица знатно већа него разлика у просечним вредностима постигнућа на тесту. С обзиром на то да је број степени слободе у свим случајевима већи од 60.000, све ове разлике су статистички високо значајне.

ТАБЕЛА 1. Број испитаника, проценат завршних оцена према полу из математике, просечне вредности завршних оцена за дечаке и девојчице, стандардна девијација, t-вредност разлике и Коенов d

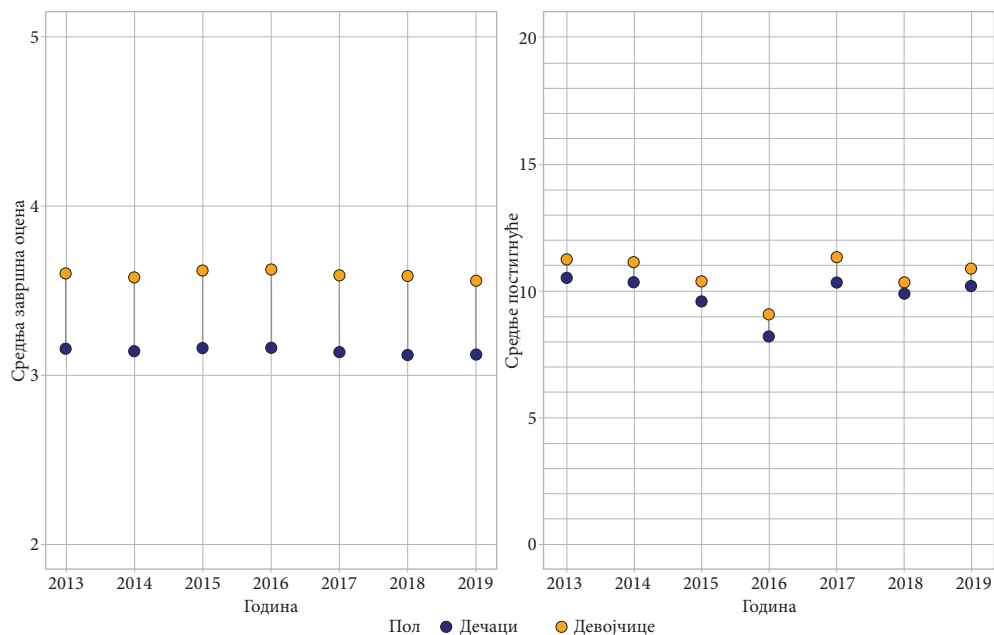
Година	Пол	N	Завршна оцена				Средња оцена	SD	t-вредност	Коенов d
			2	3	4	5				
2013	Дечаци	32981	42,3%	21,1%	16,1%	20,5%	3,15	1,21	48,3*	0,37
	Девојчице	31433	26,5%	21,1%	18,0%	34,4%	3,60			
2014	Дечаци	32452	43,1%	21,0%	15,3%	20,7%	3,14	1,22	46,5*	0,37
	Девојчице	30166	27,4%	21,0%	17,6%	33,9%	3,58			
2015	Дечаци	33445	42,1%	21,0%	15,6%	21,3%	3,16	1,22	49,1*	0,38
	Девојчице	31778	26,1%	20,4%	18,6%	34,9%	3,62			
2016	Дечаци	33134	42,2%	21,2%	15,6%	21,0%	3,16	1,22	50,3*	0,39
	Девојчице	31609	25,9%	20,5%	18,5%	35,1%	3,63			
2017	Дечаци	32019	43,2%	20,9%	15,2%	20,7%	3,13	1,22	48,1*	0,38
	Девојчице	30599	27,2%	20,3%	18,3%	34,2%	3,59			
2018	Дечаци	32948	43,8%	20,7%	15,1%	20,4%	3,12	1,22	49,3*	0,38
	Девојчице	31472	27,4%	20,3%	18,5%	33,8%	3,59			
2019	Дечаци	32157	43,8%	20,8%	15,2%	20,2%	3,12	1,21	46,3*	0,36
	Девојчице	30621	27,9%	21,0%	18,7%	32,5%	3,56			

*p<0,01

ТАБЕЛА 2. Број испитаника и просечне вредности постигнућа на испиту из математике за дечаке и девојчице, стандардна девијација постигнућа, t-вредност разлике и Коенов d

Година	Пол	n	Средња вредност	SD	t-вредност	Коенов d
2013	Дечаци	32981	10,49	5,03	19,1*	0,15
	Девојчице	31433	11,25			
2014	Дечаци	32452	10,33	5,26	19,0*	0,15
	Девојчице	30166	11,13			
2015	Дечаци	33445	9,58	5,03	19,3*	0,15
	Девојчице	31778	10,34			
2016	Дечаци	33134	8,18	5,27	21,9*	0,17
	Девојчице	31609	9,08			
2017	Дечаци	32019	10,32	4,98	25,0*	0,20
	Девојчице	30599	11,31			
2018	Дечаци	32948	9,9	4,28	11,7*	0,09
	Девојчице	31472	10,3			
2019	Дечаци	32157	10,19	4,28	20,0*	0,16
	Девојчице	30621	10,87			

СЛИКА 1. Поређење завршних оцена из математике и постигнућа на тесту из математике за дечаке и девојчице



* $p < 0,01$

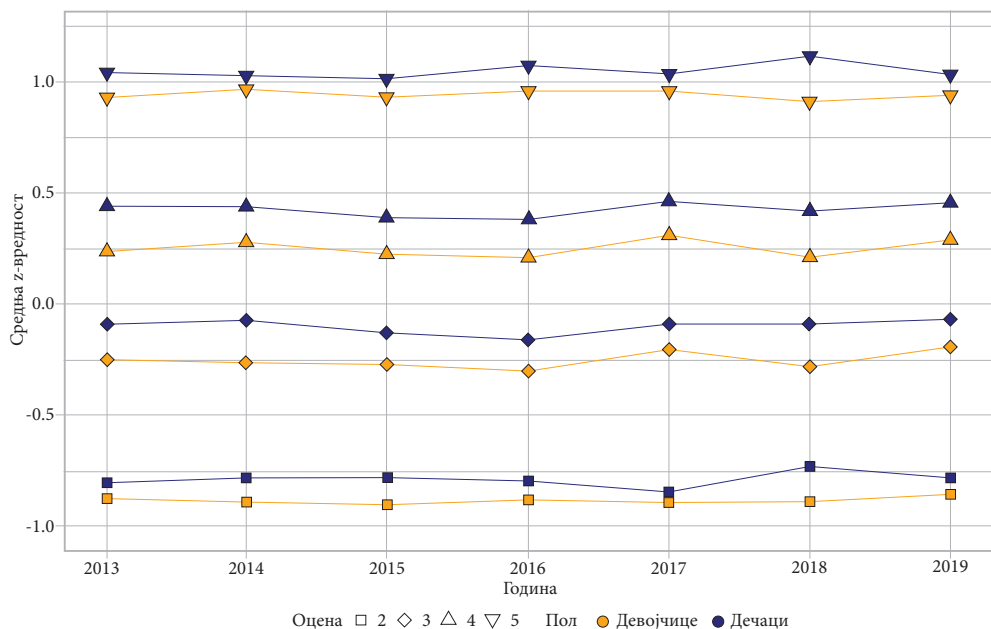
На основу графичког приказа резултата завршног теста на Слици 1 може се претпоставити да тежина теста није била иста током свих разматраних година. На пример, 2016. године тест је био захтевнији, што се препознаје у нижим просечним резултатима како дечака, тако и девојчица. Због поређења тестова из различитих година, као и због усклађивања са сличним истраживањима из литературе, постигнућа су трансформисана у z -вредности. У Табели 3 приказане су просечне z -вредности резултата на тесту из математике, унакрсно распоређене према завршним оценама из математике и полу ученика. Стандардне грешке свих ових вредности износе 0,01 или мање. Занимљиво је да су z -вредности постигнућа дечака биле веће од z -вредности девојчица за све оцене и у свим годинама. Иако су девојчице укупно гледано биле успешније, дечаци су постизали веће просечне резултате у свакој групи према оцени. Ова појава Симпсоновог парадокса последица је велике неравнотеже у расподели завршних оцена између дечака и девојчица. Симпсонов парадокс настаје када се у више подгрупа података јавља одређени тренд, али он нестаје или се преокреће када се подаци обједине. До тога долази јер скривена или споредна варијабла утиче на резултате, што може довести до погрешних закључака ако се тај утицај не узме у обзир.

ТАБЕЛА 3. Просечне z-вредности на испиту из математике, унакрсно распоређене према завршним оценама из математике и полу. Стандардне грешке за све приказане z-вредности износе 0,01 или мање, што чини све приказане родне разлике статистички високо значајним

Година	Пол	Просечне z-вредности према завршној оцени			
		2	3	4	5
2013	Дечаци	-0,80	-0,09	0,45	1,04
	Девојчице	-0,87	-0,25	0,24	0,93
2014	Дечаци	-0,78	-0,07	0,44	1,02
	Девојчице	-0,89	-0,26	0,28	0,97
2015	Дечаци	-0,77	-0,12	0,39	1,01
	Девојчице	-0,90	-0,27	0,23	0,93
2016	Дечаци	-0,80	-0,16	0,38	1,07
	Девојчице	-0,88	-0,30	0,22	0,96
2017	Дечаци	-0,84	-0,08	0,47	1,03
	Девојчице	-0,89	-0,20	0,31	0,96
2018	Дечаци	-0,73	-0,08	0,42	1,11
	Девојчице	-0,89	-0,28	0,21	0,91
2019	Дечаци	-0,78	-0,07	0,46	1,03
	Девојчице	-0,85	-0,19	0,29	0,94

Разлике у постигнућима између дечака и девојчица (изражене као z-вредности) и завршна оцена прате исти образац у свим посматраним годинама. Те разлике су приказане за све године на Слици 2.

СЛИКА 2. Поређење z-вредности ученика за све завршне оцене и све посматране године.



Задаци на које ученици нису одговорили

Поставили смо хипотезу да ниже оцене, које дечаци добијају, могу бити последица тога што наставници несвесно оцењују и ненамерно вреднују некогнитивне особине ученика, као што су труд, став према школи, понашање и одговорност, односно особине које у идеалним условима не би требало да утичу на оцењивање.

Међу овим карактеристикама једина коју можемо непосредно повезати са променљивама из наше базе података јесте труд. Прескакање задатака, односно одсуство било каквог одговора на задатак, може се тумачити као недостатак мотивације или неспремност да се уложи већи напор у решавање математичких проблема. Утврдили смо да дечаци чешће не одговарају на задатке него девојчице. Укупно гледано, дечаци прескачу 24,1% задатака, док девојчице прескачу 18,4%.

Будући да различити типови задатака захтевају различите нивое труда и начине одговарања, очекивали смо да на отворене и сложеније задатке ученици чешће неће одговорити. Статистички смо упоредили учесталост изостављених одговора за три типа задатака: задаци са понуђеним одговорима, кратки одговори и задаци отвореног типа.

У Табели 4 приказани су проценти неодговорених и нетачно решених задатака за ова три типа задатака, и за дечаке и за девојчице. Као што се и очекивало, учесталост прескакања одговора код задатака са понуђеним одговорима знатно је нижа у поређењу са задацима који захтевају конструисани одговор, посебно код задатака отвореног типа. Такође, примећује се да девојчице ређе остављају задатке не одговоривши у односу на дечаке, али истовремено имају и већи број нетачних одговора код задатака отвореног типа.

ТАБЕЛА 4. Проценти задатака на које ученици нису одговорили и нетачни одговори према типу задатка (MC – задаци са понуђеним одговорима, SA – кратки одговори, OA – одговори отвореног типа) и према полу ученика за све тестове из математике у периоду од 2013. до 2019. године

	% неодговорених задатака			% нетачних задатака		
	MC	SA	OA	MC	SA	OA
Дечаци	1,20%	6,39%	41,3%	20,7%	20,3%	26,7%
Девојчице	0,78%	4,62%	31,7%	20,1%	19,1%	30,8%
Коенов h	0,04	0,08	0,20	0,01	0,03	-0,09

Резултати показују да дечаци изражавају знатно већу склоност ка изостављању одговора у поређењу са девојчицама, при чему је та разлика статистички високо значајна. Ефекат величине разлике је најизраженији код задатака отвореног типа, што указује на то да је највећа разлика у спремности да се одговори, између дечака и девојчица, управо код задатака који захтевају отворене одговоре. Ово запажање имплицира да су девојчице склоније покушају решавања задатка, чак и када нису сигурне у решење. Ова особина у понашању дечака могла би бити један од разлога због којих они генерално бележе нижа постигнућа из математике.

Ставови наставника

У различитим образовним контекстима како на нивоу сопственог искуства тако и у ширем школском окружењу, наставници су из својих перспектива углавном закључили да не постоје значајне разлике у школским оценама између девојчица и дечака. Овакав образац примећен је и у урбаним и у руралним срединама, што је приказано у Табели 5. Као одговор на приказане податке, који указују на то да девојчице у просеку имају више оцене од дечака, наставници су понудили објашњења заснована на опаженим родним карактеристикама. Навели су да девојчице углавном показују већу марљивост, одговорност и педантност у учењу. Примећено је да се девојчице више припремају за тестове и пажљиво проверавају своја решења, али су мање склоне ризику и показују израженији страх од грешака.

С друге стране, наставници су приметили да су дечаци склонији такмичењу, боље се сналазе у раду са технологијом и спремнији су да ризикују. Међутим, такође су истакли да се дечаци генерално мање припремају за тестове и лакше одустају. Наставници су такође дискутовали о утицају дисциплинских мера на оцене дечака, наводећи примере када је успостављање дисциплине кроз оцењивање доводило до нижих оцена за дечаке (видети Табелу 5).

ТАБЕЛА 5. Запажања наставника – понашање дечака и девојчица

Запажања наставника	Образложења наставника
Нема разлике	Немам утисак да девојчице имају боље оцене. То се мења из године у годину. Моја школа је мала, сеоска. Можда је узорак мали. У нашој школи (градској школи) нема такве разлике. Имају исте оцене. Код дечака је то вероватно присилно, то је за њих спољашња мотивација.

Понашање девојчица	Девојчице су ујорније, марљивије, раде до краја и проверавају решења. Девојчице су послушније, једанјније и одговорније. Однос према раду је друјачији. Девојчице вежбају за шесћ. Девојчице се плаше да не направе грешке.
Понашање дечака	Дечаци су такмичарски настрјенији. Више воле да се надмећу. Дечаци се не припремају, па праве тривијалне грешке. Дечаци су доминантни у вези са технолојом. Када раде неку примену математике, дољи су. Често добијамо врхунске резултате од дечака, али они често и одустају. Више дечака излази на таблу, преузимају више ризика и ојуштенији су. У нашој школи сваке године ученик генерације је дечак који је нешто освојио на математичком такмичењу.
Дисциплиновање	Одржавање дисциплине може резултирајти лошим оценама. Када је ученик немиран, наставник га често изведе пред таблу, што често доводи до негативне оцено. У већини случајева то су дечаци.

ТАБЕЛА 6. Запажања наставника – ниже оцено дечака и њихове последице

Запажања наставника	Образложења наставника
Слабија оцено не утиче на ученике	Мислим да средњошколци немају појма коју су оцено добили у осмом разреду и да их то заправо не занима. Верујем да ниже оцено не производе последице. Њих то не занима.
Слабија оцено не утиче на упис у средњу школу	Сви се уписују у школу и сви иду даље без обзира на оцено. То их много не поља. У Врдасу сви уписују природно-математички смер. И они са двојкама. У средњим школама увек има места, осим у Београду.
Слабија оцено не утиче на будућа постигнућа	Није одлучујуће шта су знали и коју оцено су имали са 15 година. Имам колеу који је докторирао математику у Гену, а имао је тројку из математике у средњој школи.

ДИСКУСИЈА

Јаз у оцењивању

Налази ове студије указују на то да девојчице остварују боље резултате од дечака у математици, што потврђују њихове више просечне оцено из математике и виши просечни резултати на завршном испиту из математике (Табела 1, Табела 2 и Слика 1). Међутим, разлика у просечним оценама између дечака и девојчица већа је од разлике у постигнућима на самом испиту, што указује на могућност постојања пристрасности у оцењивању у корист девојчица.

Запажање да девојчице у просеку постижу више резултате на завршном испиту разликује се од налаза других стандардизованих тестова, попут PISA истраживања, према којем дечаци из Србије остварују боље резултате из математике, иако та разлика није увек статистички значајна (OECD, 2023; Videnović & Čarpić, 2020). У овом истраживању, ефекат величине разлике у постигнућима на тесту из математике креће се од 0,15 до 0,20 стандардних девијација. Ови налази се разликују од резултата које су приказали Противински и Минх (Protivínský & München, 2018): утврдили су родну разлику од приближно 0,29 стандардних девијација, али у корист дечака. Такође, наши налази се не подударају са резултатима Машина и Пекаринена (Machin & Pekkarinen, 2008), који су анализом глобалних PISA података утврдили да петнаестогодишњи дечаци постижу боље резултате од девојчица у математици за 0,10 стандардних девијација.

Додатна анализа према оценама потврдила је постојање пристрасности у оцењивању, показујући да дечаци остварују боље резултате на завршном испиту из математике међу ученицима са истим закљученим оценама (Табела 3). Дати налази, који представљају пример Симпсоновог парадокса, додатно потврђују постојање пристрасности у оцењивању у корист девојчица. Добијени резултати се поклапају са налазима других студија (Di Liberto et al., 2022; Graetz & Karimi, 2022; Protivínský & München, 2018; Terrier, 2020), али се разликују од студије у којој је идентификован јаз у оцењивању у корист дечака, али је коришћена нешто другачија методологија (Lavy & Sand, 2015).

Резултати ове студије показују да јаз у оцењивању превазилази било коју појединачну генерацију ученика и да се доследно јавља из године у годину. Анализа података за период од седам година открива образац који се понавља: у просеку, девојчице добијају више оцене и, уопштено гледано, бележе боља постигнућа на завршном испиту из математике. Ипак, међу ученицима који имају исте оцене дечаци постижу више резултате од девојчица на самом испиту.

Утврђено је да је јаз у оцењивању постојан и да траје током низа узастопних година. Претходна истраживања су испитивала факторе који доприносе овом јазу, идентификујући варијабле, попут интензитета конкуренције у току оцењивања, динамике интеракције наставник–ученик, као и утицаја некогнитивних вештина на оцењивање (Falch & Naper, 2013; Lavy, 2008). Други фактори се односе на степен анонимности у оцењивању и наставни план, који може истицати различите компетенције, као и разлике у начину интеракције и временском размаку између наставе и оцењивања (Lavy, 2008).

За разлику од претходних студија, у овом испитивању су биле у фокусу и перспективе наставника како би се боље разумеле њихове перцепције о јазу у оцењивању и могућим узроцима који леже у основи овог феномена.

Како наставници доживљавају јаз у оцењивању?

Иако резултати спроведеног истраживања указују на очигледан јаз у оцењивању, наставници га често не препознају у оквиру сопствене наставне праксе или унутар својих школа. Склони су да узроке јаз у оцењивању припишу особинама ученика, а не пристрасности у сопственом оцењивању. Наставници препознају различито понашање ученика за које верују да доприноси расподели оцена у којој девојчице чешће добијају више оцене (Li, 1999).

Запажања наставника донекле се поклапају са налазима истраживања која показују да дечаци постижу боље резултате у конкурентским окружењима и под повећаним стресом (Gneezy et al., 2003; Ors et al., 2013). Ова запажања такође су у складу са налазима Јурајдеа и Минха (Jurajda & Münich, 2011), који показују да девојчице често остварују слабије резултате на тестовима високог ризика. Међутим, истраживање које су спровели Фалк и Напер (Falch & Naper, 2013) у Норвешкој испитивало је постојање јаз у оцењивању кроз различите врсте тестирања, не ограничавајући се само на поређење школских оцена и тестова високог ризика. То истраживање, заједно са налазима Лавија (Lavy, 2008) и Матијеја и Смита (Matějů & Smith, 2014) потврђује да је јаз у оцењивању једнако присутан како у тестовима високог, тако и ниског ризика, као и у анонимним тестирањима. Стога, на основу поменутих студија, може се закључити да утицај конкурентности окружења у ком се врши процена постаје упитан (Protivínský & Münich, 2018).

Наставници су указали на то да су особине попут марљивости, упорности, посвећености, педантности и темељне провере решења израженије код девојчица, што може позитивно утицати на њихова постигнућа. У квантитативном делу анализе посматран је број задатака на које ученици нису одговорили, разврстано према типовима задатака. Налази показују да девојчице доследно покушавају да одговоре на задатке свих типова, при чему су разлике у сваком типу статистички значајне. Најизраженија разлика у изостављању одговора забележена је код задатака отвореног типа, где девојчице показују знатно већу спремност да покушају да реше задатак.

Ови налази делимично се поклапају са резултатима истраживања (Beller & Gafni, 2000; Graetz & Karimi, 2022; Riener & Wagner, 2018; Wester & Henriksson, 2000) која указују на то да су девојчице успешније у задацима отвореног типа. Таква склоност одговара запажањима наставника да девојчице показују већу упорност и марљивост, као и спремност да покушају да реше задатке чак и када нису сигурне у тачност свог одговора. Ова запажања су у контрасту са налазима фон Шрадера и Ензлија (Von Schrader & Ansley, 2006) и Ринера и Вагнера (Riener & Wagner, 2017), који су известили да дечаци ређе остављају задатке без одговора у математици. Кључно запажање односи се на упорност девојчица да покушају да одговоре на задатке, без обзира на степен сигурности – особина која је идентификована као један од могућих фактора који доприносе јазу у оцењивању.

Наставници који су учествовали у овом истраживању изразили су уверење да ниже оцене дечака неће имати негативан утицај на њих, њихов упис у средње школе нити на њихова будућа постигнућа. Овакав став заснивају на чињеници да у неким средњим школама често остају слободна места. Ово мишљење наставника међутим није у складу са налазима истраживања Лавија и Санда (Lavy & Sand, 2015) и Федеричове (Federičová, 2016), јер показују да оцене имају важну улогу у обликовању ученикове перцепције сопствених постигнућа.

Дакле, пристрасност у оцењивању може утицати на будући живот ученика на различите начине. Ученици и родитељи обично посматрају оцене као поуздану повратну информацију о учениковим постигнућима и напретку и ослањају се на њих приликом доношења образовних и професионалних одлука. Ако је та повратна информација пристрасна, те одлуке могу бити субоптималне, што потенцијално доводи до неефикасности на тржишту рада (Protivínský & Münich, 2018).

У Србији постоји релативно ниско интересовање за упис у гимназије, што често резултира слободним местима у тим школама. Ипак, у многим школама потребан је велики број бодова за упис, при чему оцене из математике имају значајну улогу. Поред тога, општи успех, укључујући оцене из математике, значајно утиче и на право на смештај у домовима ученика и на доделу стипендија.

ЗАКЉУЧЦИ

У презентованој студији анализиране су школске оцене ученика и њихова постигнућа на завршном испиту из математике на крају обавезног основног

образовања у Републици Србији. На први поглед, запажање да девојчице у просеку добијају више оцене и постижу боље резултате на завршном испиту није деловало неуобичајено. Међутим, детаљнијом анализом откривено је да је разлика у оценама између дечака и девојчица знатно већа од разлике у њиховим просечним резултатима на испиту, што указује на постојање јаз у оцењивању у корист девојчица. Накнадна анализа према оценама потврдила је постојање пристрасности у оцењивању, показујући да међу ученицима који имају исту завршну оцену дечаци постижу боље резултате од девојчица на завршном испиту из математике. Континуирано присуство ове појаве у више година анализе указује на системски, трајан јаз у оцењивању у образовном систему у Србији. Бројне студије бавиле су се могућим узроцима овог јаз. Посебан допринос овог истраживања представља примена квалитативне методологије, односно фокус група у којој су учествовали наставници како би се испитала њихова перцепција родног јаз. Иако је утврђено постојање јаз у оцењивању, наставници углавном нису препознали његово постојање, већ су разлике у оценама приписивали унутрашњим карактеристикама ученика, што свакако може допринети уоченом јазу.

Једно од значајних запажања ове студије јесте да су девојчице склоније да покушају да одговоре на задатак, чак и када нису сигурне у решење. Наставници су изнели помало неконвенционално мишљење да нешто ниже оцене код дечака вероватно неће имати значајније последице по њих. Овакав став занемарује добро документован утицај који оцене имају на одређене животне прилике попут уписа у средње школе и на факултете, стицање права на стипендије, добијање места у ученичким и студентским домовима и општу перцепцију ученикових способности.

Ово запажање указује на важност детаљног испитивања ставова наставника и подизање вести о улози и последицама наставничких пракси у оцењивању. Иако је квалитативни део овог истраживања донео нове увиде, он има одређена ограничења – на пример ограничен број наставника у фокус групи и временски размак између квантитативне и квалитативне фазе истраживања. Ова ограничења указују на потребу за будућим истраживањима у којима би додатно били испитани, потврђени или оспорени налази представљени у овом раду. Дата студија покреће кључно питање поузданости наставничких оцена у поређењу са резултатима завршног испита. У актуелном систему, оба елемента имају важну улогу у процени подобности ученика за

наставак школовања. Студија, као и шири образовни оквир, суочавају се са ограничењима у домену процене да ли се садржај који оцењују наставници подудара са садржајем завршног испита. Такође се поставља питање упоредивости континуираног оцењивања, које обухвата вишемесечни рад и труд ученика, са једним испитом који превасходно мери знање из математике. Постаје јасно да оцена из математике одражава више од самог знања јер она обухвата и шири спектар ученикових особина и понашања. Упркос уоченим разликама између школских оцена и резултата на завршном испиту, остаје отворена дебата о томе да ли би ове метрике требало ускладити, имајући у виду да оцењују различите димензије учениковог знања и постигнућа.

Сукоби интереса. Аутори изјављују да немају финансијске интересе нити личне односе који би могли утицати на рад приказан у овом раду.

Финансирање: Нема финансирања.

Напомена. Ово истраживање је подржано од стране Министарства науке, технолошког развоја и иновација Републике Србије (уговори бр. 451-03-65/2024-03/200125, 451-03-65/2024-03/200156, 451-03-65/2024-03/200102, 451-03-137/2025-03/200251, 451-03-137/2025-03/200102, 451-03-65/2025-03/200156) и Факултета техничких наука Универзитета у Новом Саду кроз пројекат *Научноистраживачки и уметнички рад истраживача у наставним и сарадничким звањима на Факултету техничких наука Универзитета у Новом Саду* (бр. 01-3394/1).

КОРИШЋЕНА ЛИТЕРАТУРА

- Adler, K., Salanterä, S., & Zumstein-Shaha, M. (2019). Focus group interviews in child, youth, and parent research: An integrative literature review. *International Journal of Qualitative Methods*, 18, 1–15. DOI: 10.1177/1609406919887274
- Avitzour, E., Choen, A., Joel, D., & Lavy, V. (2020). On the origins of gender-biased behavior: The role of explicit and implicit stereotypes (NBER Working Paper No. 27818). *National Bureau of Economic Research*. DOI:10.3386/w27818
- Becker, D. F., & Forsyth, R. A. (1990, April). Gender differences in academic achievement in grades 3 through 12: A longitudinal analysis. Paper presented at the Annual Meeting of the American Educational Research Association, Boston, MA.
- Beller, M., & Gafni, N. (2000). Can item format (Multiple Choice vs. Open-Ended) Account for gender differences in mathematics achievement? *Sex Roles*, 42, 1–21. DOI:10.1023/A:1007051109754
- Birenbaum, M., & Feldman, R. A. (1998). Relationships between learning patterns and attitudes towards two assessment formats. *Educational Research*, 40(1), 90–98. DOI:10.1080/0013188980400109
- Borghans L. Golsteyn B. H. Heckman J. J., & Humphries J. E. (2016). What grades and achievement tests measure. *Proceedings of the National Academy of Sciences*, 113(47) 13354–13359. DOI:10.1073/pnas.1601135113
- Đerić, I. Gutvajn, N., Jošić, S. & Ševa, N. (Eds.) (2021). *TIMSS 2019 u Srbiji*. Beograd: Institut za pedagoška istraživanja.
- Dersch, A. S., Heyder, A., & Eitel, A. (2022). Exploring the nature of teachers' math-gender stereotypes: The math-gender misconception questionnaire. *Frontiers in Psychology*, 13, Article 820254. DOI:10.3389/fpsyg.2022.820254

- Di Liberto, A., Casula, L., & Pau, S. (2022). Grading practices, gender bias and educational outcomes: evidence from Italy. *Education Economics*, 30(5), 481–508. DOI:10.1080/09645292.2021.2004999
- Falch, T., & Naper, L. R. (2013). Educational evaluation schemes and gender gaps in student achievement. *Economics of Education Review*, 36, 12–25.
- Federičová, M. (2016). Gender gap in application to selective schools: Are grades a good signal? (CERGE-EI Working Paper No. 550).
- Ferman, B., & Fontes, L. F. (2022). Assessing knowledge or classroom behavior? Evidence of teachers' grading bias. *Journal of Public Economics*, 216, Article 104773. DOI:10.1016/j.jpubeco.2022.104773
- Garner, M., & Engelhard, G., Jr. (1999). Gender differences in performance on multiple-choice and constructed response mathematics. *Applied Measurement in Education*, 12(1), 29–51.
- Giofrè, D., Cornoldi, C., Martini, A., & Toffalini, E. (2020). A population level analysis of the gender gap in mathematics: Results on over 13 million children using the INVALSI dataset. *Intelligence*, 81, 101467. DOI:10.1016/j.intell.2020.101467
- Gneezy, U., Niederle, M., & Rustichini, A. (2003). Performance in competitive environments: Gender differences. *The Quarterly Journal of Economics*, 118(3), 1049–1074.
- Graetz, G., & Karimi, A. (2022). Gender gap variation across assessment types: Explanations and implications. *Economics of Education Review*, 91, 102313. DOI:10.1016/j.econedurev.2022.102313
- Groth, R. E. (2006). Analysis of an online case discussion about teaching stochastics. *Mathematics Teacher Education and Development*, 7, 53–71.
- Guez, A., Peyre, H., & Ramus, F. (2020). Sex differences in academic achievement are modulated by evaluation type. *Learning and Individual Differences*, 83–84, 101935. DOI:10.1016/j.lindif.2020.101935
- Guttmann, J., & Boudo, M. (1988). Teachers' evaluations of pupils' performance as a function of pupils' sex, family type and past school performance. *Educational Review*, 40(1), 105–113. DOI:10.1080/0013191880400108
- Han, L., & Hoover, H. D. (1994, April). Gender differences in achievement test scores. Paper presented at the Annual Meeting of the National Council on Measurement in Education, New Orleans, LA.
- Hanna, R. N., & Linden, L. L. (2012). Discrimination in grading. *American Economic Journal: Economic Policy*, 4(4), 146–168.
- Hanushek, E. A. & Woessmann L. (2012). Do better schools lead to more growth? Cognitive skills economic outcomes and causation. *Journal of Economic Growth*, 17(4) 267–321. DOI:10.1007/s10887-012-9081-x
- Hinnerich, B. T., Högl, E., & Johannesson, M. (2011). Are boys discriminated in Swedish high schools? *Economics of Education Review*, 30(4), 682–690.
- Holder, K., & Kessels, U. (2017). Gender and ethnic stereotypes in student teachers' judgments: A new look from a shifting standards perspective. *Social Psychology of Education*, 20, 471–490.
- Jurajda, Š., & Münich, D. (2011). Gender gap in performance under competitive pressure: Admissions to Czech universities. *The American Economic Review*, 101(3), 514–518.
- Lauermann, F., Meißner A., & Steinmayr R. (2020). Relative importance of intelligence and ability self-concept in predicting test performance and school grades in the math and language arts domains. *Journal of Educational Psychology*, 112(2), 364. DOI:10.1037/edu0000377
- Lavy, V. (2008). Do gender stereotypes reduce girls' or boys' human capital outcomes? Evidence from a natural experiment. *Journal of Public Economics*, 92(10), 2083–2105.
- Lavy, V., & Sand, E. (2015). On the origins of gender human capital gaps: Short and long term consequences of teachers' stereotypical biases (NBER Working Paper No. 20909). National Bureau of Economic Research. <https://www.nber.org/papers/w20909>
- Li, Q. (1999). Teachers' beliefs and gender differences in mathematics: A review. *Educational Research*, 41(1), 63–76. DOI:10.1080/0013188990410106
- Lindner, J., Makarova, E., Bernhard, D., & Brovelli, D. (2022). Toward gender equality in education – Teachers' beliefs about gender and math. *Education Sciences*, 12(6), Article 373. DOI:10.3390/educsci12060373

- ☞ Lindqvist, A.-K., Weurlander, M., Wernerson, A., & Thornberg, R. (2020). A focus group interview study of the experience of stress amongst Swedish schoolchildren. *International Journal of Environmental Research and Public Health*, 17(11), 4021. DOI:10.3390/ijerph17114021
- ☞ Machin, S., & Pekkarinen, T. (2008). Global sex differences in test score variability. *Science*, 332(5906), 1331–1332.
- ☞ Marcenaro-Gutierrez, O. D., Prieto-Latorre, C., & Sánchez Rodríguez, M. I. (2023). Gender differences between teachers' assessments and test-based assessments: Evidence from Spain. *Assessment in Education: Principles, Policy & Practice*. DOI:10.1080/0969594X.2023.2251715
- ☞ Matějů, P., & Smith, M. (2014). *Are boys that bad? Gender gaps in measured skills, grades and aspirations in Czech elementary schools*. British Journal of Sociology of Education, 36(6), 871–895.
- ☞ Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis* (2nd ed.). Thousand Oaks, CA: Sage.
- ☞ Mullis, I. V. S., Martin, M. O., Foy, P., Kelly, D. L., & Fishbein, B. (2020). *Average achievement by gender – TIMSS 2019 International results in mathematics and science*. TIMSS & PIRLS International Study Center, Boston College.
- ☞ Newell, G., Wilsman, M., Langenfeld, M., & McIntosh, A. (2002). Online professional development: Sustained learning with friends. *Teaching Children Mathematics*, 8, 505–508.
- ☞ OECD. (2015). *The ABC of Gender Equality in Education*. OECD. DOI: 10.1787/9789264229945-en
- ☞ OECD. (2023). Serbia student performance (PISA 2022). Education. <https://gpseducation.oecd.org>
- ☞ Ors, E., Palomino, F., & Peyrache, E. (2013). Performance gender gap: Does competition matter? *Journal of Labor Economics*, 31(3), 443–499.
- ☞ Protivínský, T., & Münich, D. (2018). Gender bias in teachers' grading: What is in the grade. *Studies in Educational Evaluation*, 59, 141–149. DOI:10.1016/j.stueduc.2018.07.006
- ☞ Rakshit, S., & Sahoo, S. (2023). Biased teachers and gender gap in learning outcomes: Evidence from India. *Journal of Development Economics*, 161, Article 103041. DOI:10.1016/j.jdevco.2022.103041
- ☞ Riener, G., & Wagner, V. (2017). Shying away from demanding tasks? Experimental evidence on gender differences in answering multiple-choice questions. *Economics of Education Review*, 59, 43–62. DOI:10.1016/j.econedurev.2017.06.005
- ☞ Riener, G., & Wagner, V. (2018). Gender differences in willingness to compete and answering multiple-choice questions – The role of age. *Economics Letters*, 164, 86–89. DOI:10.1016/j.econlet.2018.01.012
- ☞ Sumpter, L. (2016). Investigating upper secondary school teachers' conceptions: is mathematical reasoning considered gendered? *International Journal of Science and Mathematics Education*, 14, 347–362. DOI: 10.1007/s10763-015-9634-5
- ☞ Terrier, C. (2020). Boys lag behind: How teachers' gender biases affect student achievement. *Economics of Education Review*, 77, 101981. DOI:10.1016/j.econedurev.2020.101981
- ☞ Tiedemann, J. (2002). Teachers' gender stereotypes as determinants of teacher perceptions in elementary school mathematics. *Educational Studies in Mathematics*, 50, 49–62. DOI:10.1023/A:1020518104346
- ☞ Videnović, M., & Čaprić, G. (2020). *PISA report for 2018, for the Republic of Serbia*. Ministry of Education, Science and Technological Development. <https://ceo.edu.rs/wp-content/uploads/2022/11/PISA-izvestaj-2018.pdf>
- ☞ Von Schrader, S., & Ansley, T. (2006). Sex differences in the tendency to omit items on multiple-choice tests: 1980–2000. *Applied Measurement in Education*, 19(1), 41–65.
- ☞ Wester, A., & Henriksson, W. (2000). The interaction between item format and gender differences in mathematics performance based on TIMSS data. *Studies in Educational Evaluation*, 26(1), 79–90. DOI:10.1016/s0191-491x(00)00007-9
- ☞ Willingham, W. W., Pollack, J. M. & Lewis, C. (2002). Grades and test scores: Accounting for observed differences. *Journal of Educational Measurement*, 39(1) 1–37. DOI:10.1111/j.1745-3984.2002.tb01133.x
- ☞ Zanga, G., & De Gioannis, E. (2023). Discrimination in grading: A scoping review of studies on teachers' discrimination in school. *Studies in Educational Evaluation*, 78, 101284.