

# POMEN IN VLOGA MEDNARODNIH PRIMERJAV ZNANJA V VZGOJI IN IZOBRAŽEVANJU

*Mojca Štraus*

*Pedagoški inštitut, Ljubljana*

---

## Uvod

Čeprav so prvi rezultati mednarodnih primerjav znanja običajno objavljene kot lestvice povprečnih dosežkov sodelujočih držav na merjenem področju, rezultati teh primerjav služijo tudi bolj poglobljenemu ugotavljanju kakovosti izobraževalnih sistemov. Oblikovanje kazalnikov o stanju v izobraževalnem sistemu in analize rezultatov v povezavi z nacionalnimi dokumenti, ki navajajo, kaj naj bi se učenci naučili oziroma kaj so cilji in standardi izobraževanja, so lahko vir informacij za izboljšave kurikula, učnega procesa in rezultatov izobraževanja.

Pred desetletji so bili kazalniki kakovosti izobraževalnih sistemov oblikovani na primer kot število šol in učencev na posameznih stopnjah izobraževanja ali število učiteljev v razmerju do števila učencev (Poslethwaite, 1994: 26). V zadnjih desetletjih pa ugotavljanje dostopnosti izobraževanja ni več med najpomembnejšimi kazalniki njegove kakovosti, saj je postalo jasno, da zgolj skrb za zagotavljanje dostopnosti ne zagotavlja uspešnosti izobraževanja. V ospredje so prišli kazalniki tako imenovanih rezultatov izobraževanja, na primer ravni doseženega znanja učencev na različnih področjih. Znanje v današnji družbi namreč predstavlja kritični element uspešnosti v globalni gospodarski konkurenci (npr. OECD 2001: 265; Husén in Tuijnman, 1994: 6). To je povzročilo premik politične pozornosti s problemom omogočanja in upravljanja kvantitativne rasti izobraževalnega sistema na spremljanje njegovega delovanja v smislu rezultatov (Husén in Tuijnman, 1994: 11). Dosežki učencev, učinkovitost šol in odgovornost za doseganje zastavljenih ciljev so postali eni najpomembnejših kriterijev

za ugotavljanje kakovosti izobraževalnih sistemov (Bottani in Tuijnman, 1994: 23). Tudi v Sloveniji je eden od ciljev prenovljenega šolskega sistema, da »omogoča doseganje mednarodno primerljivih standardov ob koncu osnovne šole« (*Bela knjiga o vzgoji in izobraževanju v Republiki Sloveniji*, 1995: 71).

Potrebe po informacijah in podatkih o kazalnikih te uspešnosti so tako vedno večje. Eden od načinov pridobivanja teh podatkov je izvajanje nacionalnih in mednarodnih primerjalnih raziskav znanja. Primeri teh raziskav so Program mednarodne primerjave dosežkov učencev PISA (*Programme for International Student Assessment*), ki poteka pod okriljem Organizacije za ekonomsko sodelovanje in razvoj (OECD), Mednarodna raziskava trendov znanja matematike in naravoslovja TIMSS (*Trends in Mathematics and Science Study*) in Mednarodna raziskava bralne pismenosti PIRLS (*Progress in International Reading Literacy Study*), ki potekata pod okriljem Mednarodnega združenja za raziskovanje učinkov izobraževanja IEA (*International Association for Evaluation of Educational Achievement*). Te raziskave dopolnjujejo podatke o ravneh znanja učencev, ki jih nudijo ocene učiteljev in rezultati državnih preizkusov znanja. Vendar so med državnimi preizkusi in raziskavami znanja razlike. Medtem ko je osnovni namen državnega preizkusa obvestilo učencu o njegovem znanju, raziskave znanja nudijo informacije o delovanju izobraževalnega sistema nasploh, na nacionalni ali na mednarodni ravni. Običajno rezultate raziskav znanja podajamo na ravni celotne populacije učencev ali posameznih podskupin.

V članku opisujem vlogo in pomen mednarodnih primerjav znanja za oblikovanje strategije razvoja izobraževalnega sistema. Pomembnost teh raziskav se kaže tudi v številčnosti držav, ki v teh raziskavah sodelujejo. To nakazuje, da se veliko držav, še posebej iz razvitejšega dela sveta, pri oblikovanju strategije razvoja izobraževalnih sistemov naslanja na podatke, zbrane v teh raziskavah, in primerjave stanja v svojem izobraževalnem sistemu s stanjem v drugih državah. V članku opisujem zasnovo in namene mednarodnih primerjalnih raziskav v splošnem, možne uporabe njihovih podatkov in kako lahko iz podatkov teh raziskav oblikujemo veljavne sklepe.

## **Spremljava in evalvacija izobraževalnega sistema skozi mednarodne primerjave znanja**

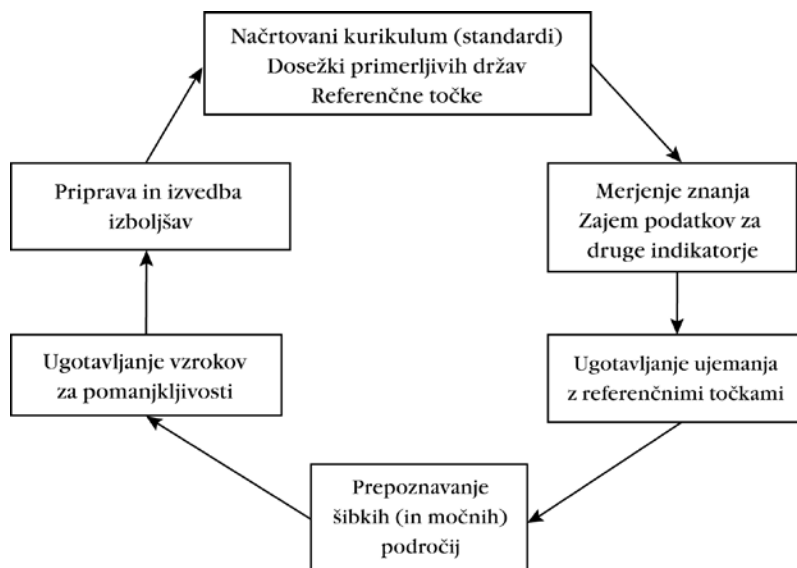
Mednarodne primerjalne raziskave so se pojavile iz interesa več držav za pridobivanje dodatnih informacij o kakovosti njihovega izobraževalnega sistema skozi primerjavo dosežkov učencev z dosežki učencev iz drugih držav. V raziskavah znanja iščemo odgovore na vprašanja, kot so: Kakšno je znanje učencev glede na splošna pričakovanja, dosežke v drugih državah, cilje kurikula ali pripravo na nadaljnje izobraževanje in vključevanje v družbo? Na katerih področjih dosežki učencev ustrezajo pričakovanim ravšem in na katerih ne? Kakšno je znanje posameznih podskupin učencev, na primer učencev iz posameznih regij, ali kakšne so primerjave med dosežki učenk in učencev? Kateri dejavniki v šolskem sistemu ali zunaj njega so povezani z uspešnostjo učencev? Ali se ravni znanja učencev med letom spreminjajo?

Obstajajo pa tudi druga področja primerjanja. Države med seboj lahko primerjajo podatke o kurikulu, številu ur, ki so namenjene posameznim predmetom, organizaciji šolskega sistema na nacionalni ravni in organizaciji življenja in dela na ravni posameznih šol. Ne nazadnje so pomembne primerjave podatkov o mnenjih in stališčih učencev do določenih vprašanj v povezavi z njihovim izobraževanjem. Tako so na primer v povezavi z naravoslovnimi dosežki pomembna stališča učencev do okoljevarstvene problematike, načinov raziskovanja in uporabe znanosti (na primer genško raziskovanje) in zanimanje učencev za poklice, ki so povezani z naravoslovjem.

Prvi korak v mednarodni primerjalni raziskavi je priprava mednarodne definicije področja merjenja znanja in zajete populacije učencev. Na podlagi teh definicij so pripravljene instrumenti za merjenje znanja in zajemanje podatkov o spremljajočih dejavnikih. Preizkuse znanja in spremljajoče vprašalnike izpolnjujejo reprezentativni vzorci učencev, običajno istočasno v vseh sodelujočih državah. Po mednarodno dogovorjeni metodologiji sodelujoče države izvedejo zajeme podatkov na izbranih šolah in učencih in pripravijo podatkovne baze. Po dodatnih mednarodnih preverjanjih so baze združene v skupno mednarodno bazo podatkov raziskave, ki je po objavi njenih rezultatov običajno javno dostopna, na primer prek svetovnega spleta. Veliko mednarodnih raziskav dosežke učencev spremlja s ponavljajočimi zajemi podatkov v razmikih nekaj let za ugotavljanje, kako se dosežki s časom spreminjajo in ali so imele uvedene spremembe v izobraževalnem sistemu želen učinek.

Za oblikovanje sklepov o stanju v izobraževalnem sistemu pa same številke kot rezultati analiz podatkov iz mednarodnih baz niso dovolj. Za vrednotenje rezultatov potrebujemo ustrezne referenčne točke. Te lahko izberemo ali sestavimo na podlagi ciljev izobraževanja, standardov v učnih načrtih, ali pa, kot je najpogosteje v mednarodnih raziskavah, na podlagi dosežkov primerljivih učencev v drugih državah. Zbrane podatke potem lahko interpretiramo v smislu ujemanja z izbranimi referenčnimi točkami. Iz tega lahko izpeljemo izsledke o močnih in šibkih področjih v dosežkih, ugotavljamo, kaj so vzroki za problematična področja in kaj bi bilo treba storiti za izboljšanje stanja. Povezava med temi elementi je prikazana na *sliki 1*.

*Slika 1: Spremljava in evalvacija izobraževalnega sistema*



Vloga raziskav znanja je v izvedbi korakov v prvi polovici kroga, ki omogočajo prepoznavo močnih in šibkih področij v dosežkih učencev glede na izbrane referenčne točke. Za oblikovanje poglobljenih interpretacij rezultatov in odločitve o potrebnih ukrepih za izboljšanje pa so pomembni koraki v drugi polovici kroga, v kateri zajemu podatkovnih baz in osnovnim analizam sledijo kvalitativne ocene in študije rezultatov iz zbranih podatkov s sodelovanjem šolskih strokovnjakov in drugih. Brez tega vsi

zbrani podatki in izračunani dosežki ne morejo učinkovito služiti izboljševanju stanja v izobraževalnem sistemu.

Spremljavo in evalvacijo izobraževalnega sistema lahko izvajamo z nacionalnimi ali z mednarodnimi raziskavami znanja (glej tudi Kellaghan in Greany, 2001). Pri tem so med nacionalnimi in mednarodnimi raziskavami nekatere razlike. Nacionalne raziskave imajo lahko prednost v tem, da se prilagodijo morebitnim specifičnim okoliščinam v državi, še posebej kurikulumu, ki je v izobraževalnem sistemu v uporabi, in tako natančno izmerijo njegove učinke. Poleg tega v nacionalnih raziskavah ni problemov oblikovanja primerljivih definicij ciljne populacije učencev ter prevajanja in prirejanja instrumentarija, ki pogosto otežujejo zagotavljanje veljavnosti primerjav med državami. Kljub temu pa ima sodelovanje v mednarodnih primerjalnih raziskavah tudi nekaj pomembnih prednosti. Očitna prednost je, da mednarodne raziskave nudijo podatke o izobraževalnem sistemu, ki so primerljivi s podatki o drugih sistemih. Ker absolutnih standardov za dosežke v izobraževanju ni, so mednarodne primerjalne raziskave nujne za pridobivanje opisnih podatkov o razponu dosežkov ali drugih podatkov v izobraževalnih sistemih drugih držav (Kellaghan in Greany, 2001: 63; Plomp, 1992: 278). Na ta način mednarodne primerjalne raziskave pripomorejo k postavljanju smiselnih izobraževalnih standardov.

Mednarodne primerjalne raziskave lahko pripomorejo k razumevanju povezav med dejavniki v izobraževalnem sistemu in pojasnjevanju razlik v dosežkih učencev med državami in znotraj njih. Eden izmed pogosto obravnavanih dejavnikov je kurikulum. Beaton, Postlethwaite, Ross, Spearitt in Wolf (1999: 13–14) poudarjajo, da je prednost sodelovanja v mednarodnih primerjalnih raziskavah, da so države prisiljene podrobneje pregledati svoje učne načrte. Ti na posameznem področju morda postanejo zastareli in s tem ne vključujejo vsebin in pristopov, ki v svetu pridobivajo večji pomen. Sodelovanje v nekaterih mednarodnih primerjalnih raziskavah vključuje pregled učnih načrtov držav na področju, ki je predmet raziskovanja. S tem so ministrstva in strokovnjaki za kurikulum lahko opozorjeni na razlike med državami glede vsebin in poudarkov v njih. Države se lahko odločijo za posodobitev učnih načrtov po vzoru primerljivih držav. Lahko pa že same primerjave dosežkov z drugimi državami vodijo do razmišljanja o prenovi učnih načrtov. To še posebej velja za področja, na katerih so ugotovljeni primanjkljaji s konkurenčnimi državami.

Predvsem državam v razvoju sodelovanje v mednarodnih primerjalnih raziskavah lahko pomeni še nekaj dodatnih pridobitev (Plomp, Howie in McGaw, 2003: 954). Zahteva mednarodnih primerjalnih raziskav

po doseganju tehničnih standardov zajemanja podatkov (glej npr. Martin, Rust in Adams, 1999) lahko pripomore k izboljševanju kakovosti samega raziskovanja v državah, ki sicer morda nimajo močne tradicije na tem področju. S sodelovanjem v mednarodnih primerjalnih raziskavah nekatere države zberejo osnovne podatke o dosežkih na posameznih področjih kurikula, kjer morda prej ti niso bili na voljo. Skupno sodelovanje z izkušenejšimi državami omogoča učenje iz njihovih izkušenj. Ne nazadnje pa je pomembno tudi, da je zaradi pozornosti medijev mednarodnim primerjalnim raziskavam izobraževanje napredovalo na prednostni listi področij, ki potrebujejo pozornost političnih vrhov.

### **Definicije znanja in interpretacije rezultatov v mednarodnih raziskavah**

Med najpomembnejšimi elementi zasnove mednarodne primerjalne raziskave je definicija področja znanja, ki ga želimo meriti. V splošnem Cole (1990: 2) omenja dva glavna koncepta: lahko je to diskreten nabor znanj in veščin, ki ga je moč neposredno povezati s specifično zapisanimi cilji in standardi v učnih načrtih, lahko pa so to višje ravni in celovitejša znanja, ki jih ni tako preprosto meriti. Obe definiciji igrata pomembno vlogo v pripravi zasnove raziskave in interpretaciji njenih rezultatov. Povezavo med definicijo merjenja znanja in uporabo rezultatov si pogledjmo na primerih raziskav PISA in TIMSS.

Medtem ko obe raziskavi vključujeta merjenje znanja na področju matematike in naravoslovja in tudi zajemata populacijo učencev ob koncu oziroma tik po koncu obveznega dela izobraževanja, je med raziskavama razlika predvsem v tem, koliko se definicija znanja naslanja na kurikule sodelujočih držav. Definicija področja merjenja znanja matematike in naravoslovja se v TIMSS-u naslanja na kurikule sodelujočih držav v smislu, da je pregled pokritosti predlaganih nalog v mednarodnem preizkusu znanja eden od ključnih elementov, na podlagi katerih oblikujejo končni izbor teh nalog. Domena merjenja in mrežni diagrami za preizkuse znanja so pokriti s kurikuli sodelujočih držav, kolikor je to le mogoče (Plomp in dr., 2003: 968; Mullis, Martin, Gonzalez in Chrostowski, 2004: 397).

V PISI pokritost nalog s kurikuli sodelujočih držav pri izboru nalog za mednarodni preizkus znanja ni odločilni kriterij. Domene merjenja znanja oziroma kompetence, kot jih imenujemo v PISI, se oblikujejo v mednarodnem projektu DeSeCo (*Definition and Selection of Competencies*, OECD, 2002, 2003). Merjene kompetence v PISI so definirane kot sposobnost

uspešnega soočanja z zahtevami na ravni posameznika ali družbe ter uspešnega izvajanja dejavnosti ali nalog. Pri tem projekt DeSeCo razlikuje med pojmom kompetence in veščine. Medtem ko veščine pomenijo izvajanje kompleksnih motoričnih in/ali kognitivnih postopkov natančno in prilagodljivo okoliščinam, kompetence ponazarjajo kompleksni sistem delovanja, ki vključuje tako kognitivne veščine kot tudi stališča in druge nekognitivne komponente. Kompetence so kombinacija kognitivnih in praktičnih spretnosti, znanja, motivacije, vrednostne orientacije, stališč, vedenja, čustev in drugih socialnih in vedenjskih komponent, ki se glede na potrebe vključujejo v učinkovito delovanje posameznika. Čeprav sta kognitivna spretnost in osnovno znanje kritična elementa, je vključitev drugih komponent, na primer motivacije in vrednostne orientacije, pomembna.

Tako je na primer naravoslovna pismenost v PISI 2006 definirana kot znanja, vrednote in zmožnosti, ki naj bi jih 15-letniki izkazovali v okviru ustreznih osebnih, družbenih in globalnih naravoslovnih kontekstov. To se razlikuje od naravoslovnih disciplin v šoli, jih pa vključuje. Naravoslovna pismenost v PISI 2006 se nanaša na posameznikovo naravoslovno znanje in uporabo tega znanja za prepoznavanje vprašanj v raziskovanju, pridobivanje novega znanja, pojasnjevanje naravoslovnih pojavov, izpeljevanje na opažanjih in podatkih temelječih sklepov v naravoslovnem kontekstu, razumevanje karakteristik naravoslovja kot načina pridobivanja znanja in razumevanja sveta, zavedanje, kako naravoslovje in tehnologija vplivata na naše materialno, intelektualno in kulturno okolje, ter pripravljenost angažirati se v problemih, ki so povezani z naravoslovjem.

V PISI je naravoslovna pismenost opredeljena v dveh dimenzijah: znanje naravoslovja in vedenje o naravoslovju. Znanje naravoslovja pomeni znanje pomembnejših področij, kot so fizika, biologija, kemija ter znanje o Zemlji in vesolju. Vedenje o naravoslovju pa pomeni poznavanje načinov naravoslovnega raziskovanja, njegovih ciljev pojasnjevanja naravnih pojavov in poznavanje znanosti in tehnologije ter njune vloge v družbi. K nalogam v preizkusu znanja bodo v PISI 2006 priključena tudi vprašanja o mnenjih in stališčih učencev o pojavih in problemih, ki so povezani z vsebino v nalogah. Ti podatki bodo omogočili pridobitev informacij o stališčih učencev do določenih naravoslovnih vprašanj. Ta stališča učencev v svetu in pri nas razumemo kot pomemben del rezultatov naravoslovnega izobraževanja.

Glede na opisane razlike med definicijami znanja v raziskavah TIMSS in PISA so interpretacije, ki jih iz rezultatov teh dveh raziskav oblikujemo, različne. V TIMSS-u lahko neposredno ugotavljamo, ali učenci izkazujejo

želene ravni dosežkov na področjih, ki so zajeta v kurikulumu. V primeru nezadovoljivih dosežkov tako lahko ugotovljamo, kje so problemi v kurikulumu ali njegovi implementaciji, da že zapisani cilji in standardi niso doseženi. To je mogoče vsaj za večji del področij, ki jih zajema preizkus znanja in so pokrita z nacionalnim kurikulumom. Nepokrita področja, ki jih je po zasnovi raziskave manj, pa morda nakazujejo, kaj bi bilo treba v kurikule še dodati za ustrezne rezultate.

V PISI ta povezava ni tako neposredna. V tej raziskavi ugotovljamo, ali učenci ob koncu obveznega izobraževanja, natančneje v starosti 15 let, izkazujejo znanja in veščine, ki jih na mednarodni ravni razumejo kot pomembne za 15-letnikovo nadaljnje izobraževanje in vključevanje v družbo. Pri tem ni nujno, da so ta znanja in veščine pridobili med izobraževanjem, čeprav naj bi to predstavljalo najpomembnejši vir teh znanj. Naloge v PISI tako niso zastavljene neposredno po vzoru učnih načrtov in učbenikov sodelujočih držav, ampak od učenca zahtevajo uporabo šolskega znanja v širšem kontekstu, na primer v vsakdanjih situacijah ali pri razumevanju pomena in uporabi znanja v znanosti. Če na nacionalni ravni v dosežkih ugotovimo pomanjkljivosti, lahko ugotovljamo, ali kurikulum in njegova implementacija ustrezno podpirata pridobivanje teh znanj, katere vsebine bi bilo treba v kurikulumu dodatno ali bolje obravnavati in kako izboljšati njegovo implementacijo za doseganje splošnega cilja »pripravljenost na življenje«. Po vzoru uspešnejših primerljivih držav se lahko na primer pripravijo predlogi za posodobitve učnih načrtov.

Zaradi razlik v definicijah raziskav torej ni nujno, da bodo rezultati TIMSS-a in PISE enaki. V nobenem primeru jih ne bomo mogli razumeti kot kontradiktorne, ampak kvečjemu kot dopolnjujoče. Različne definicije v raziskavah lahko namreč vodijo do različnih razvrstitev na lestvicah držav. Primer so rezultati Slovenije v Mednarodni raziskavi IAEP (*International Assessment of Educational Progress*, 1991) in v raziskavi TIMSS 1995. Medtem ko je bila razvrstitev Slovenije za populacijo 13-letnih učencev pri matematiki v IAEP 1991 proti dnu lestvice (Lapointe, Mead in Askew, 1992: 16), je TIMSS 1995 za približno leto in pol starejšo populacijo učencev pokazal boljšo sliko (Beaton, Mullis, Martin, Gonzalez, Kelly in Smith, 1996: 22). Mogoči so različni vzroki za te rezultate in med njimi je razlika v metodologiji teh dveh raziskav prav gotovo eden pomembnejših. IAEP 1991 je primerjal 13-letne učence ne glede na razred, ki so ga obiskovali, medtem ko je TIMSS 1995 v večini držav primerjal učence v 7. in 8. razredu. Nadalje je bil v IAEP 1991 uporabljen preizkus znanja, ki je bil sestavljen na podlagi ameriške nacionalne raziskave znanja NAEP (*National*



*Assessment of Educational Progress*), v TIMSS-u 1995 pa je bil mednarodni preizkus znanja sestavljen z upoštevanjem povezave s kurikuli držav za področji matematike in naravoslovja. Ti dve raziskavi nakazujeta, da preden sklepamo o slabih ali dobrih rezultatih glede na razvrstitev države na lestvici sodelujočih držav ali glede na kak drug enoštevilčni kazalnik, je treba podrobno preučiti posamezne elemente raziskav.

### **Možne uporabe podatkov iz mednarodnih raziskav**

Mednarodne primerjalne raziskave lahko služijo vrsti uporab za oblikovanje šolske politike in prakse. Čeprav so medijsko najodmevnejše objave skupnih lestvic držav po povprečnih dosežkih na merjenem področju, lahko iz podatkov teh raziskav izpeljemo vrsto drugih informacij za uporabo v šolski politiki. Poleg neposrednih primerjav dosežkov je eden od pomembnih ciljev mednarodnih raziskav znanja podati opise delovanja različnih dejavnikov v šolskih sistemih in njihovih povezav z dosežki. Spodnji pregled možnih uporab podatkov mednarodnih raziskav temelji na Plomp in dr. (2003: 953).

Posamezna država lahko primerja opis stanja svojega šolskega sistema s stanjem druge države, ki jo posebej zanima, kot so na primer »kulturni« sosedje ali ekonomski konkurenti. Če na primer neka država želi izboljšati svoj kurikulum, lahko primerja stanje v svoji državi s stanjem v drugi državi, in sicer po specifičnih spremenljivkah, ki se povezujejo s kurikulumom (kurikularni materiali, pristopi k poučevanju, usposabljanje učiteljev, dosežki itn.). Take primerjave lahko vodijo do oblikovanja predlogov za spremembe.

Redno sledenje stanju izobraževalnega sistema je potrebno, da bi v izobraževalni sistem lahko vpeljali spremembe takrat, ko bi se to izkazalo za potrebno, in na področju, kjer bi bilo to treba. V ta namen se mnoge mednarodne primerjalne raziskave izvajajo v večletnih ciklih.

Na pridobivanje in izkazovanje znanja vplivajo mnogi dejavniki v izobraževalnem sistemu, zato je treba čim bolj razjasniti njihovo delovanje in vplive. To še posebej velja za tiste dejavnike, na katere je moč vplivati z vpeljavo sprememb v sistemu. Za to so pomembne mednarodne primerjave podatkov o spremljajočih dejavnikih, ki so v raziskavah znanja zajeti z vprašalniki za učence, učitelje, ravnatelje in včasih tudi starše.

Za čim bolj informirano sprejemanje odločitev o potrebnih spremembah v izobraževalnem sistemu je treba pridobiti tudi podatke o možnih vzrokih za razlike med sistemi, prek katerih potem lahko poskušamo oblikovati vzode za izboljšanje stanja.

## **Raziskovanje vpliva različnih dejavnikov v izobraževalnem sistemu na njegovo uspešnost**

Med možnimi uporabami podatkov mednarodnih raziskav je raziskovanje vpliva različnih dejavnikov v izobraževalnem sistemu na njegovo uspešnost verjetno deležno največje pozornosti v tako imenovanih sekundarnih raziskavah. Primerjave povprečnih dosežkov med državami, ugotavljanje trendov v določenih časovnih obdobjih in seznanjanje s procesi v izobraževanju sicer temeljijo na zahtevnih tehničnih in statističnih pripravah, sama uporaba podatkov pa temelji na preprostejšem opisovanju. Raziskovanje dejavnikov poleg ustrezne priprave baz podatkov zahteva kompleksnejše pristope k analizi podatkov. Še pomembneje pa je, da razumemo, kakšne interpretacije rezultatov iz teh analiz lahko oblikujemo.

Raziskave zbirajo podatke o spremljajočih dejavnikih, kot so šolski program, karakteristike učencev, podatke o domačih pogojih, pogojih v šoli in drugih spremenljivkah, kot so motivacija, mnenja in želje za prihodnost. Vendar pa iz rezultatov analiz ne moremo sklepati neposredno o *vzrokih* za uspešnost v šoli oziroma pri mednarodnem preizkusu znanja. Slikovit primer za to so podatki raziskave NAEP 1992 (Mullis, Campbell in Farstrup, 1993, pridobljeno iz Mislevy, 1995: 428). Čeprav vemo, da dodatno branje povečuje uspeh pri branju, je bila v raziskavi korelacija med količino učenja branja in dosežkom pri preizkusu branja negativna. Tako se zdi, da učenje branja zmanjšuje dosežke pri branju. Vendar povprečna razlika v dosežkih pri preizkusu branja med učenci z različnimi časi učenja branja ne meri nujno povprečnega *vzročnega* učinka večje količine branja na dosežke. Drugi dejavniki, ki lahko vplivajo na razlike v dosežkih, namreč niso upoštevani. Negativna korelacija je smiselna, če upoštevamo, da so učenci, ki se dodatno učijo brati, običajno slabši bralci. Zanje se zdi smiselno, da tudi potrebujejo več učenja, hkrati pa imajo učenci, za katere se zdi, da potrebujejo več učenja, običajno že na začetku nižje dosežke. Tudi drugi dejavniki, ki bi jih morda interpretirali kot vzročne vplive na dosežke, so lahko napačni iz povsem podobnih razlogov.

Vzročne vplive bi lahko raziskovali, če bi lahko raziskave znanja izvajali kot znanstvene poskuse s slučajnostnim razvrščanjem opazovanih enot, torej učencev, v kontrolno in eksperimentalno skupino. V tem primeru bi opazovane vplive lahko razumeli tudi kot vzroke, saj so vse druge spremenljivke, tudi tiste, ki se jih ne zavedamo, neodvisne od določevanja pogojev, ki jih hočemo primerjati. Njihovi učinki se izničujejo, ko povečujemo velikost vzorca. Brez slučajnostnega razporejanja se učinki

drugih spremenljivk ne izničuejo nujno in morajo biti predpostavke, ki jih potrebujemo, da lahko sklepamo o vzročnih povezavah opazovanih vplivov, zelo močne: obe primerjani skupini bi morali biti zelo homogeni s podobnimi načini odgovarjanja na vplive, ki jih raziskujemo. Ampak v mednarodnih raziskavah nimamo podobnosti med učenci različnih držav. Tudi ko jih poenotimo glede na starost, spol, dohodke doma in izobrazbo staršev, ostanejo razlike med učenci iz Tajvana, ZDA in Slovenije glede kulture, mnenj, motivacije in vrednot. Ti dejavniki pa lahko vplivajo na učinek izobraževanja in torej na dosežek pri mednarodnem preizkusu znanja.

Vseeno raziskovanje povezav med dosežki in spremljajočimi dejavniki ni povsem nesmiselno. Občasno lahko poenotimo in primerjamo skupine učencev, ko niso slučajnostno razporejeni v primerjalne skupine, in uporabimo statistične pristope, kot je regresija, da izločimo »moteče« vplive. S temi pristopi seveda izločimo le del takšnih motenj in tako sam obstoj opazovanih in dobro razumljenih spremenljivk, ki izkazujejo močne povezave z dosežki, ne more nikoli povsem izločiti možnosti, da so opažene povezave posledica neopazovanih dejavnikov. Raziskovanje povezav med spremljajočimi dejavniki in dosežki je torej uporabno kot podpora hipotezam o vzrokih in pogojih za visoke dosežke. Na podlagi opaženih povezav lahko oblikujemo nove hipoteze, ki jih potem preverjamo v nadaljnjih zajemih podatkov ali na bolj »eksperimentalne« načine.

Poleg izogibanja napačnemu sklepanju o vzročnosti povezav je pri raziskovanju dejavnikov v izobraževalnem sistemu ključnega pomena kakovost zbranih podatkov. Če so spremljajoče spremenljivke slabo definirane ali nezanesljivo merjene, so izsledki o povezavah z dosežki lahko napačni. Običajno je seveda treba uravnotežiti kakovost spremenljivk z njihovo ceno v smislu časa, vloženih sredstev, motivacije in pripravljenosti za sodelovanje. Zbiranje podatkov o spremljajočih dejavnikih od učencev samih je zelo ekonomično v primerjavi z zunanji sodelavci, ki v razredu opazujejo dogajanje. Vendar pa je pri mlajših učencih cena za to natančnost podatkov. Pri učiteljih in ravnateljih pa nikoli ne moremo biti gotovi, koliko se podatki, ki jih poročajo v vprašalnikih, dejansko ujemajo s potekom pouka v razredih in s stanjem na šolah. V nekaterih predraziskavah zato uporabljamo zbiranje istih podatkov iz več virov in tako imenovano trianguliranje, s katerim ugotavljamo kakovost in zanesljivost zbranih podatkov.

## Veljavnost izpeljanih sklepov

Temeljnega pomena pri izvedbi raziskave in uporabi njenih podatkov je veljavnost izpeljanih sklepov. Obstaja veliko nevarnosti v samem izvajanju raziskave in kasnejšem oblikovanju interpretacij, ki lahko to veljavnost zmanjšujejo (glej npr. Crooks, Kane in Cohen, 1996). Na mednarodni ravni je zato uvedena vrsta standardov za izvedbo raziskave v sodelujočih državah, ki jih mednarodni center preverja s tako imenovano kontrolo kakovosti. Poleg ustrezne nacionalne definicije ciljne populacije zajetih učencev morajo sodelujoče države zagotoviti ustreznost prevodov mednarodnih predlog preizkusov znanja in spremljajočih vprašalnikov, vzorčenja šol in učencev, okoliščin poteka zajema podatkov na šolah in kodiranja odprtih odgovorov. Problemi, ki lahko pri teh elementih izvedbe raziskave ogrozijo veljavnost izpeljanih sklepov, so podrobneje opisani v nadaljevanju. Tudi računanje skupnega dosežka pri preizkusu znanja in posploševanje le-tega na celotno področje merjenja znanja lahko vplivata na veljavnost sklepov. Ti vplivi so opisani na koncu razdelka.

Pri prevodu izvirnika mednarodnega preizkusa znanja, ki je pogosto v angleščini in pri PISI še v francoščini, je poleg samega prevajanja treba paziti na »funkcionalno ekvivalentnost« nalog. Funkcionalna ekvivalentnost pomeni, da želimo sestaviti naloge za učence iz različnih držav, ki bodo njihovo znanje merile na primerljiv način. Pogosto je pri prevajanju težko najti ustrezne prevode besed, ki jih učenci v šolah uporabljajo, ob tem pa ohraniti vse merske elemente naloge, kot so težavnost, diskriminativnost, privlačnost distraktorjev in podobno. Mednarodna verifikacija pripravljenih prevodov omogoča zagotavljanje ustreznosti nacionalnega instrumentarija standardom prevajanja in s tem zmanjševanja nevarnosti za neprimerljivost zbranih podatkov.

Mednarodni centri raziskav tudi ugotavljajo ustreznost pripravljenih vzorcev šol in učencev za sodelovanje v raziskavi. Pri tem sta pomembni ujemanje nacionalne definicije ciljne populacije učencev z mednarodno definicijo in reprezentativnost izbranega vzorca za nacionalno definicijo. Pogosto je nemogoče zagotoviti, da bi bili kot ciljna populacija v raziskavi vključeni prav vsi učenci, ki bi načeloma ustrezali mednarodni definiciji. V nekaterih državah iz finančnih ali drugih razlogov ni moč zagotoviti, da bi bili v ciljno populacijo vključeni učenci iz oddaljenih geografskih območij, specifičnega kulturnega ali jezikovnega okolja ali učenci s posebnimi potrebami. Mednarodni standardi v PISI tako na primer določajo,

da na ta način izključenih učencev ne sme biti več kot 2,5 odstotka celotne populacije učencev v državi, ki bi ustrezali mednarodni definiciji.

Po končani izvedbi zajema podatkov mednarodni center preveri tako imenovano odzivnost vzorca. To je odstotek učencev izmed učencev, izbranih v vzorec, ki so se tudi dejansko udeležili raziskave. Če je ta odstotek premajhen ali z drugimi besedami, je veliko manjkajočih podatkov, na podlagi zbranih podatkov ne moremo veljavno sklepati o lastnostih celotne ciljne populacije, ki smo jo želeli zajeti.

Okoliščine, v katerih poteka zajem podatkov o znanju učencev z nalogami v preizkusih znanja, imajo lahko velik vpliv na veljavnost interpretacij rezultatov merjenja znanja. Na primer nizka motivacija za reševanje lahko vpliva na to, da so izsledki iz rezultatov zelo drugačni, kot bi jih dobili iz odgovorov visoko motiviranih učencev. To je lahko še posebej problem v mednarodnih primerjalnih raziskavah, kjer dosežek učencev ne vpliva na njihove ocene v šoli. Vir problema so lahko tudi nezanimive ali pa pretežke naloge.

Dosežki učencev so lahko neustrezno nizki, če niso bili uporabljeni primerni postopki za zajem podatkov. Lahko se zgodi na primer, da ni dovolj časa za reševanje, prostorske razmere niso bile ustrezne ali navodila za reševanje niso vsebovala vseh potrebnih informacij. Dosežki učencev so lahko neustrezno visoki, če so jim bile naloge predstavljene vnaprej, če so imeli na voljo vire, ki običajno niso dovoljeni, jim je kdo pri reševanju nalog pomagal, ali pa so rešitve koga drugega predstavili kot svoje. Nevarnosti »pozitivnega« odstopanja od realnih dosežkov so seveda večje pri preizkusih znanja, ki vplivajo na oceno učencev, kot pri raziskavah znanja.

Če naloge niso ustrezno predstavljene, so lahko slabši odgovori učencev napačno interpretirani kot njihovo neznanje, čeprav je vzrok pomanjkljivega ali napačnega odgovora slabo razumljena naloga ali nezmožnost predstavitve rešitve. Očitni primeri so učenci s posebnimi potrebami, ki potrebujejo posebne prilagoditve, da lahko izkažejo svoje znanje.

Če dejavnikov okoliščin izvajanja preverjanja znanja ne uspemo kontrolirati, lahko to omeji veljavnost interpretacij rezultatov zunaj danih okoliščin. Standardizacija teh okoliščin tako zelo pripomore k izboljšanju veljavnosti interpretacij. Vendar pa lahko toge standardizacije po drugi strani preveč zožijo definicijo področja merjenja znanja, da bi bile interpretacije splošneje uporabne.

Ko so podatki o znanju učencev zajeti, je naslednji korak vrednotenje oziroma, kot rečemo v mednarodnih raziskavah, kodiranje njihovih odgo-

vorov. Pri tem je več nevarnosti za veljavnost interpretacij rezultatov. Pri vrednotenju nalog z vnaprej danimi odgovori je nevarnost, da je vnesena napačna koda ali da je v programu za računanje dosežkov kot pravilen določen v resnici nepravilen odgovor. Pri vrednotenju nalog z odprtimi odgovori je nevarnost poudarjanje nekaterih elementov znanja in ne drugih, ki pa so lahko prav tako pomembni. Problem je na primer, če koderji pozitivno vrednotijo le nekatere stile odgovorov in ne drugih, prav tako pravilnih. Če se recimo pri merjenju znanja naravoslovja koder omeji na pozitivno vrednotenje le tistih odgovorov, ki so tudi slovnično pravilni, rezultati neustrezno prikazujejo slabše naravoslovno znanje učencev, ki so pri izražanju naredili slovnične napake.

Pri vrednotenju nalog z odprtimi odgovori je lahko problem nizko ujemanje kod posameznih koderjev pri istih nalogah ali kod med različnimi koderji. Če koderji spreminjajo vrednotenje istih nalog, različno upoštevajo standarde za vrednotenje ali kodirne sheme, je to velika nevarnost za veljavnost interpretacij rezultatov. Standardizacija vrednotenja med koderji in med nalogami je tako zelo pomembna. V mednarodnih primerjalnih raziskavah poskušamo doseči standardizacijo vrednotenja s pripravo enotnih kodirnih shem in z izvedbo večkratnega vrednotenja dela nalog z odprtimi odgovori. Ujemanje med različnimi koderji tako pomaga ugotavljati stopnjo objektivnosti, ki jo takšno vrednotenje lahko zagotavlja.

### *Računanje skupnega dosežka pri preizkusu znanja*

Preizkusi znanja v mednarodnih raziskavah so sestavljeni iz več samostojnih nalog. Že rezultati pri vsaki posamezni nalogi nam povedo svojo zgodbo. Rezultati pri preizkusu znanja s sto ali več nalogami nam povedo sto ali več zgodb – kar je zlahka prevelika količina informacije še za kurikularne strokovnjake. Določene skupine nalog, ki se povezujejo glede vsebine in znanj, ki jih merijo, običajno povedo podobne zgodbe. Temeljni zakon povzemanja podatkov je, da združevanje poudari skupne lastnosti v nalogah in hkrati zasenči lastnosti, ki so značilne le za posamezne naloge (Mislevy, 1995: 426). Tako je določanje skupin nalog, ki optimizirajo povzemanje informacije za oblikovanje interpretacij rezultatov, ključnega pomena.

Optimalne ravni povzemanja informacije se lahko razlikujejo za različne ravni in namene interpretacij. Skupni dosežki povedo večino zgodbe, če so razlike v dosežkih po posameznih nalogah med skupinami približno podobne. V mednarodnih raziskavah različno uteževanje rezultatov

pri posameznih nalogah ne bi smelo imeti velikega vpliva na primerjave med državami. Vendar pa so načini, kako učenci odgovarjajo na vprašanja, pogosto povezani s poudarki snovi v učnih načrtih in ti se med državami lahko zelo razlikujejo. Čeprav so lahko primerjave podobne med nalogami znotraj posameznih vsebinskih sklopov, se lahko primerjave med državami zelo razlikujejo med posameznimi sklopi v preizkusu znanja. Tako je na primer naravoslovje zelo široko področje, ki ni hierarhično razporejeno, zato si je težko zamišljati, kaj pomeni skupni dosežek glede na to, da so vključena podpodročja, kot so fizika, kemija, biologija in deloma tudi geografija.

Če so primerjave na podpodročjih zelo različne, so interpretacije skupnih dosežkov pri celotnem preizkusu odvisne od števila nalog v posameznem sklopu. Število nalog za posamezno podpodročje v skupnem dosežku naj bi odsevalo njegovo relativno pomembnost znotraj skupnega področja, ki ga merimo. Neprimerno uteževanje pa lahko izvira tudi iz postopkov vrednotenja ali premajhne razpršenosti dosežkov v katerem od posameznih delov preizkusa. Na primer, če je preizkus znanja sestavljen iz nalog z vnaprej danimi odgovori in iz eseja, kjer imata oba enako težo pri končnem dosežku, lahko zaradi nizke razpršenosti dosežkov v enem delu na lestvico vpliva le drugi del.

Če torej želimo delati primerjave med državami na podlagi enoštevličnega skupnega dosežka, moramo poskušati določiti dejavnike v preizkusu znanja, ki vplivajo na ta dosežek, in interpretirati rezultate skupnih dosežkov le, če se vplivi majhni.

Obstajata dva splošna modela za računanje skupnih dosežkov, t. i. klasična teorija in teorija odgovora na postavko (Item Response Theory, IRT). Gonzalez in Beaton (1994: 180) pojasnujeta, da v klasični teoriji predpostavljamo, da so naloge v preizkusu znanja vzorec iz populacije vseh možnih nalog za merjenje znanja na izbranem področju. Skupni dosežek pri preizkusu znanja ali odstotek pravih odgovorov je ocena za učenčev dosežek, če bi uporabili vse možne naloge. Dosežek pri preizkusu znanja predstavlja najboljšo možno oceno za učenčev »resnični« dosežek. Standardna napaka te ocene (napaka zaradi vzorčenja in drugih slučajnostnih vplivov) predstavlja razpršenost ocen za dosežek, ki bi jo dobili, če bi učenec reševal več naborov vzorcev nalog iz množice vseh nalog. Vendar pa ocenjevanje dosežkov na ta način ohrani vezanost na naloge, ki so bile izbrane za merjenje znanja. Če so bile naloge lahke, bodo dosežki učencev visoki, če pa so težke, bodo dosežki istih učencev z istim znanjem nižji.

V zadnjih desetletjih se je pri sestavljanju skupnih dosežkov pri preizkusih znanja uveljavila teorija IRT. V okviru teorije IRT je interpretacija skupnega dosežka pri preizkusu znanja drugačna. Predpostavljamo, da so naloge razporejene na skupni lestvici težavnosti in da je skupni dosežek učenca pri izbranih nalogah najboljša ocena učenčevega znanja, ki jo tudi lahko umestimo na to lestvico. Predpostavljamo, da je učenčevo znanje enako ne glede na težavnost nalog, ki so bile izbrane za merjenje njegovega znanja. Napaka merjenja predstavlja razpršenost ocen učenčevega znanja, če bi uporabili različne nabore nalog za izračun teh ocen.

Osnova za izračun skupnih dosežkov v teoriji IRT so verjetnostni statistični modeli, ki opazovan odgovor učenca na zastavljeno nalogo povezujejo z znanjem učenca in težavnostjo naloge (in lahko še drugimi karakteristikami naloge) prek verjetnosti, da ta učenec na to nalogo odgovori pravilno. Ti modeli omogočajo izločitev neposrednega vpliva težavnosti nalog na dobljeno oceno učenčevega znanja. Prednost teorije IRT za izračun skupnih dosežkov pred klasično teorijo je tako v tem, da omogoča primerjave na skupni lestvici dosežkov in težavnosti nalog tudi v primerih, ko niso vsi učenci reševali povsem istih nalog. To je še posebej pomembno pri raziskovanju trendov v dosežkih, kjer je težko doseči, da bi učenci ob vsakem merjenju znanja reševali iste naloge.

Veljavnost modela IRT je seveda še vedno odvisna od tega, da so v preizkusu znanja naloge, ki merijo isto vrsto znanja. Rezultati na lestvici IRT omogočajo primerjavo tistih elementov znanja, ki so nalogam v preizkusu znanja skupni. S tem modeli IRT sovpadajo s temeljnim zakonom povzemanja podatkov – poudarjanje skupnih značilnosti in zasenčenje posebnosti, ki so lastne le posameznim nalogam. Seveda pa je preverjanje veljavnosti predpostavk modela IRT, ki ga nameravamo uporabiti za zbrane podatke, z dejanskimi značilnostmi v teh podatkih pomemben del postopkov za pripravo ocen skupnih dosežkov pri preizkusu znanja.

### *Posploševanje primerjav skupnih dosežkov na celotno področje merjenja znanja*

Ko smo pridobili ocene skupnih dosežkov učencev pri nalogah v preizkusu znanja, je naslednji korak interpretacija rezultatov zunaj specifičnih okoliščin, nalog in postopkov, ki so bili uporabljeni v samem preverjanju znanja, na celotno področje merjenja znanja, kot je bilo definirano v zasnovi raziskave. Posploševanje pomeni, kako natančno lahko iz učenčevega skupnega dosežka sklepamo na njegovo znanje celotnega področja



merjenja znanja. Splošen okvir konceptov in premislekov pri posploševanju podatkov, zbranih neposredno s preizkusi znanja, na sklepe o vsebini in ravneh znanja učencev je predstavljen v Mislevy, Wilson, Erickan in Chudowsky (2003).

Ko posplošujemo interpretacije o ravneh znanja učencev z izbranih nalog na celotno področje merjenja znanja, je za veljavnost izsledkov pomembno, da je število nalog, iz katerega sklepamo na skupni dosežek, dovolj veliko. Če je število nalog premajhno, področje merjenja znanja lahko ni dovolj široko pokrito, da bi naloge zagotavljale veljavnost interpretacij skupnih dosežkov, saj so lahko posamezna pomembna vsebinska ali procesna podpodročja premalo zastopana v preizkusu znanja.

Še tako dobro izvedeno preverjanje znanja pa bo imelo zelo omejeno uporabnost, če uporabniki rezultatov in interpretacij niso dobro seznanjeni s potekom preverjanja znanja, postopki in nalogami, ki so bili uporabljeni. Nepoznavanje teh elementov lahko vodi do napačnih sklepanj iz rezultatov. Podobno lahko neprimerna pričakovanja za dosežke učencev vodijo do napačnih sklepov. V mnogih državah imajo standardi pomembno vlogo pri poučevanju in učenju ter so pogosto zapisani v učnih načrtih. Če so previsoki, so lahko dosežki učencev (napačno) interpretirani kot slabi, če so prenizki, so lahko dosežki učencev (napačno) interpretirani kot dobri. Nenehna evalvacija standardov in preverjanje načinov njihove uporabe zato igrata pomembno vlogo pri interpretiranju rezultatov preverjanj znanja v povezavi s standardi.

### **Pogoji za veljavnost izpeljanih sklepov**

Kellaghan (1996: 152–156) navaja več pogojev za veljavnost interpretacij mednarodnih primerjalnih raziskav. Pridobljeni podatki morajo dovolj natančno predstavljati dosežke učencev v posameznih državah, kot so ti definirani v raziskavi, dopuščati morajo veljavne primerjave med državami, hkrati pa naj bi jih bilo moč posploševati na oceno »človeškega kapitala« v državi, saj naj bi bil ta pomemben pogoj za ekonomsko konkurenčnost države. Nameni mednarodne raziskave morajo biti jasni. To je pomembno, ker je s tem določeno, katere informacije se bodo zbirale in kako se bo o njih poročalo ter se jih uporabljalo. Toda če je zastavljenih veliko različnih namenov, so lahko neusklajeni ali si celo nasprotujoči. Informacija iz mednarodnih raziskav naj bi služila uporabi pri oblikovanju politike in izbiri strategij za razvoj izobraževalnega sistema; na voljo morajo biti podatki o spremljajočih dejavnikih za interpretacijo razlik v dosežkih. Spemembe,

ki so vpeljane na podlagi rezultatov mednarodnih raziskav, je treba spremljati, raziskovati njihove učinke in posledice. Ne nazadnje je pomembno, da nameni raziskave in izpeljane informacije upravičijo obseg sredstev, ki so bila zanjo porabljena.

Pogoji, ki jih Kellaghan navaja, seveda niso samodejno izpolnjeni. Nekateri so za posamezne uporabe podatkov in rezultatov pomembnejši kot drugi. Dva pogoja sta osnovna: reprezentacija dosežkov in izobraževalnega sistema v državi in veljavnost primerjav. Neizpolnjenost drugih pogojev ne bo onemogočila primerljivosti dosežkov, vendar pa podatki ne bodo optimalno uporabljeni. To velja predvsem, če namen raziskave ni jasen, če informacije ne služijo oblikovanju politike in če podatki o spremljajočih dejavnikih niso na voljo. Ob tem je seveda nekatere od pogojev zelo težko zagotoviti.

## **Sklep**

Odmevnost mednarodnih primerjalnih raziskav se je v zadnjih letih povečala tako v svetu kot pri nas. V članku opisujem vlogo in pomen mednarodnih raziskav znanja pri oblikovanju strategije razvoja izobraževalnega sistema in principe raziskovanja v njih. Sodelujoče države pridobijo iz skupnih izkušenj, tehnik in rezultatov mednarodnih raziskav. Mednarodne primerjave dosežkov na primer pripomorejo k razumevanju in določanju standardov znotraj države. Mednarodne primerjalne raziskave tako lahko imajo pomembno vlogo v vsaki državi, ki poskuša izboljševati svoj izobraževalni sistem.

Čeprav mednarodne raziskave ne morejo nuditi informacije o neposrednih vzrokih za visoke dosežke, ampak največ le posredne informacije o tem, so kazalniki o stanju v šolskem sistemu potrebni pri oblikovanju politike razvoja šolskega sistema. Namesto določanja strategij razvoja v nevednosti in neinformiranosti lahko kazalniki opozarjajo na uspehe in pomanjkljivosti. Kazalniki nudijo sicer nepopolno, vendar dovolj poceni informacijo o poteku izobraževanja in rezultatih učenja v državi. Nudijo neposredne odgovore na »Kakšne so stvari?« in le posredne odgovore na vprašanje »Zakaj so stvari takšne kot so?« oziroma »Kaj moramo spremeniti, da bi jih izboljšali?«. Vloga kazalnikov iz raziskav znanja je podobna vlogi indeksa cen življenjskih potrebščin za ekonomsko politiko (Mislevy, 1995: 433).

Kot smo videli na primeru raziskav IAEP 1991 in TIMSS 1995, lahko različne definicije populacije, izbire nalog in drugih elementov vodijo do

različnega razvrščanja na lestvici dosežkov in torej do različnih interpretacij o kakovosti izobraževalnega sistema. Razprava o tem, katera izbira definicij bi vodila do »pravih« razvrstitev, je brezupna. Čeprav so primerjave dosežkov med državami zaželeno, se izobraževalni sistemi in dosežki učencev v mnogočem razlikujejo in je nemogoče zahtevati, da bi bilo primerjanje tako kompleksnih sistemov, kot so izobraževalni sistemi z vsemi svojimi karakteristikami in okoliščinami, preprosto. En sam kazalnik ne more opisati vseh značilnosti procesa izobraževanja in dosežkov učencev v državi.

## Literatura

- Beaton, A. E., Postlethwaite, T. N., Ross, K. N., Spearitt, D. in Wolf, R. M. (1999). *The benefits and limitations of international educational achievement studies*. Perth: International Academy of Education.
- Beaton, A. E., Mullis, I. V. S., Martin, M. O., Gonzalez, E. J., Kelly, D. L. in Smith, T. A. (1996). *Mathematics achievement in the middle school years: IEA's Third International Mathematics and Science Study (TIMSS)*. Chestnut Hill, MA: Boston College.
- Bela knjiga o vzgoji in izobraževanju v Republiki Sloveniji* (1996). Ljubljana: Ministrstvo za šolstvo in šport.
- Bottani, N. in Tuijnman, A. (1994). International education indicators: framework, development and interpretation. V OECD, *Making education count: developing and using international indicators*, str. 21–35. Paris: OECD.
- Cole, N. S. (1990). Conceptions of educational achievement. *Educational researcher*, 19, 2–7.
- Crooks, T. J., Kane, M. T. in Cohen, A. S. (1996). Threats to the valid use of assessments. *Assessment in education*, 3, 265–285.
- Gonzalez, E. J. in Beaton, A. E. (1994). The determination of cut scores for standards. V A. C. Tuijnman in T. N. Postlethwaite (ur.), *Monitoring the standards of education: papers in honor of John P. Keeves*, str. 171–190. Oxford: Pergamon Press.
- Husén, T. in Tuijnman, A. (1994). Monitoring standards in education: Why and how it came about. V A. C. Tuijnman, & T.N. Postlethwaite (ur.), *Monitoring the standards of education: papers in honor of John P. Keeves*, str. 1–21. Oxford: Pergamon Press.
- Kellaghan, T. (1996). IEA studies and educational policy. *Assessment in education*, 3, 143–160.
- Kellaghan, T. in Greaney, V. (2001) *Using assessment to improve quality of education*. Pariz: UNESCO International Institute for Educational Planning.
- Lapointe A. E., Mead, N. A. in Askew, G. M. (1992). *Learning mathematics*. Princeton, New Jersey: Educational Testing Service.

- Martin, M. O., Rust, K. in Adams, R. J. (ur.) (1999). *Technical standards for IEA studies*. Amsterdam, Nizozemska: IEA.
- Mislevy, R. J. (1995). What can we learn from international assessments? *Educational evaluation and policy analysis*, 17, 419–437.
- Mislevy, R. J., Wilson, M. R., Erickson, K. in Chudowsky, N. (2003). Psychometric principles in student assessment. V Kellaghan, T., Stufflebeam, D. L. in Wingate, L. A. (ur.), *International handbook of educational evaluation*, str. 489–532. Dordrecht, Nizozemska: Kluwer Academic.
- Mullis, I. V. S., Campbell, J. R. in Farstrup, A. E. (1993). *NAEP 1992 reading report card for the nation and the states*. Princeton, New Jersey: Educational Testing Service.
- Mullis, I. V. S., Martin, M. O., Gonzalez, E. J. in Chrostowski, S. J. (2004). *TIMSS 2003 International Mathematics Report: findings from IEAs Trends in International Mathematics and Science study at the fourth and eighth grades*. Chestnut Hill, MA: Boston College.
- OECD (2001). *Education at a glance*. Pariz: Avtor.
- OECD (2002). *Definition and selection of competencies (DeSeCo): theoretical and conceptual foundations, strategy paper*. Pridobljeno v juliju 2005 s spletne strani [http://www.portal-stat.admin.ch/deseco/deseco\\_strategy\\_paper\\_final.pdf](http://www.portal-stat.admin.ch/deseco/deseco_strategy_paper_final.pdf)
- OECD (2003). *Definition and selection of competencies: theoretical and conceptual foundations (DeSeCo), summary of the final report »Key competencies for a successful life and a well-functioning society«*. Pridobljeno v juliju 2005 s spletne strani [http://www.portal-stat.admin.ch/deseco/deseco\\_finalreport\\_summary.pdf](http://www.portal-stat.admin.ch/deseco/deseco_finalreport_summary.pdf)
- Plomp, T. (1992). Conceptualizing a comparative education research framework. *Prospects*, 22, 278–288.
- Plomp, T., Howie, S. in McGaw, B. (2003). International studies of educational achievement. V Kellaghan, T., Stufflebeam, D. L., & Wingate, L. A. (ur.), *International handbook of educational evaluation*, str. 951–978. Dordrecht, Nizozemska: Kluwer Academic.
- Poslethwaite, N. (1994). Monitoring and evaluation in different educational systems. V A. C. Tuijnman, & T. N. Postlethwaite (ur.), *Monitoring the standards of education: papers in honor of John P. Keeves*, str. 23–46. Oxford: Pergamon Press.

## **POMEN IN VLOGA MEDNARODNIH PRIMERJAV ZNANJA V VZGOJI IN IZOBRAŽEVANJU**

*Mojca Štraus*

Kot mnoge druge države se pri oblikovanju strategije razvoja izobraževalnega sistema tudi Slovenija vse bolj naslanja na podatke mednarodnih raziskav znanja. To nakazuje že vključevanje Slovenije v mednarodne raziskave, kot so PISA, TIMSS, PIRLS, razprave o rezultatih in njihovem pomenu. Tema tega članka so pomen in vloga teh mednarodnih raziskav znanja pri oblikovanju strategije razvoja šolskega sistema ter principi raziskovanja, na katerih te raziskave temeljijo. V članku opisujem kakšne informacije lahko iz mednarodnih raziskav znanja pridobimo in kako jih lahko interpretiramo. Za uspešno oblikovanje strategije razvoja izobraževanja je še posebej pomembna veljavnost sklepov, ki jih izpeljemo iz analiz zbranih podatkov. Te veljavnosti ne moremo neizpodbitno dokazati, lahko pa se s skrbno preiščeno zasnovo in natančno izvedbo raziskave izognemo mnogim nevarnostim, ki to veljavnost ogrožajo. V članku so opisani pristopi, ki učinkovito zmanjšujejo nevarnosti za veljavnost uporabe rezultatov.

*Ključne besede:* mednarodne raziskave, kazalniki izobraževanja, merjenje znanja, TIMSS, PISA

## **THE ROLE AND IMPORTANCE OF INTERNATIONAL COMPARATIVE SURVEYS IN EDUCATION**

*Mojca Štraus*

As in many other countries, educational policy makers in Slovenia in the past years take into account the results from the international educational research studies in their decision making. This can be seen through participation of Slovenia in international studies PISA, TIMSS and PIRLS as well as public discussions on their results and importance. The theme of this paper are the role and importance of the international comparative studies for educational policy and the principles of research that are used in these studies. For educational policy on the basis of these studies, the validity of the use of the data is especially important. This validity can never be ultimately proved, however, by using high standards of study design and data collection procedures, many threats to this validity can be avoided. In the paper these procedures and the standards for them are described.

*Key words:* international comparative surveys, educational indicators, knowledge, TIMSS, PISA

## **KNJIŽEVNA VZGOJA KOT PROFESIONALNI IZZIV BODOČIH VZGOJITELJEV/ VZGOJITELJIC IN UČITELJEV/UČITELJIC**

*Vida Medved Udovič*

Začetni študij za pedagoške poklice omogoča akademsko izobrazbo in poklicno kvalifikacijo; v visokošolskem študiju so poudarjene profesionalne kompetence za opravljanje pedagoških poklicev v sodobni družbi, kot sta obvladovanje procesov v vzgoji in izobraževanju ter poglobljeno poznavanje specifičnega (predmetnega) področja.

Članek osvetljuje pomen raziskovanja dialoga današnjih otrok z mladinsko književnostjo v dodiplomskem izobraževanju studentk, bodočih vzgojiteljic in učiteljic, za dvig bralne kulture in pismenosti v otrokovem predšolskem in šolskem obdobju. Pri tem se opira na komunikacijsko vlogo mladinske književnosti in recepcijsko zmožnost mlajših otrok. V