Indeks 363146
Nakład 420 egz.

KWARTALNIK PEDAGOGICZNY

ROK LX: 2015 NUMER 4 (238)

WYDAWNICTWA UNIWERSYTETU WARSZAWSKIEGO
WUW

KWARTALNIK PEDAGOGICZNY

MIĘDZYNARODOWA RADA NAUKOWA,
PROGRAMOWA I KONSULTACYJNA

Oskar Anweiler (Polska Akademia Nauk i Uniwersytet w Bochum, Niemcy), Ewa Czerniawska (Uniwersytet Warszawski), Hans Döbert (Deutsches Institut für Internationale Pädagogische Forschung, Niemcy), Sambor Grucza (Uniwersytet Warszawski), Jürgen Helmchen (Uniwersytet w Münster, Niemcy), Wolfgang Hörner (Uniwersytet w Lipsku, Niemcy), Yoichi Kiuchi (Uniwersytet Naruto-shi, Japonia), Krzysztof Koseła (Uniwersytet Warszawski), Viera Kurincov (Uniwersytet w Nitrze, Słowacja), Josef Malach (Uniwersytet w Ostrawie, Republika Czeska), Wolfgang Mitter (Deutsches Institut für Internationale Pädagogische Forschung, Niemcy), Nella Nyczkało (Narodowa Akademia Nauk Pedagogicznych Ukrainy), Jirzi Prokop (Uniwersytet Karola w Pradze), Mirosław S. Szymański – przewodniczący Rady (Akademia Pedagogiki Specjalnej, Warszawa), Dietmar Waterkamp (Uniwersytet w Dreźnie, Niemcy), Anna Wiłkomirska (Uniwersytet Warszawski)

WOJEWÓDZTWO
KUJAWSKO-POMORSKIE

Publikacja została dofinansowana ze środków budżetu Województwa Kujawsko-Pomorskiego w ramach zadania Strategia Edukacji Województwa Kujawsko-Pomorskiego.

KWARTALNIK PEDAGOGICZNY

ROK LX: 2015 NUMER 4 (238)

Zeszyt tematyczny pod redakcją naukową
Barbary Ciżkowicz

WUW
Warszawa 2015

Rada Redakcyjna: Adam Fijałkowski (redaktor naczelny), **Jan Rutkowski**, **Mirosław S. Szymański**, **Anna Wiłkomirska**, **Piotr Zańko** (sekretarz)

Redaktorzy tematyczni: Adam Fijałkowski – historia wychowania i myśli pedagogicznej, **Janusz Gęsicki** – socjologia oświaty, **Stefan Kwiatkowski** – pedagogika pracy, **Jan Rutkowski** – filozofia edukacji, pedagogika ogólna, **Mirosław S. Szymański** – pedagogika porównawcza, pedagogika ogólna, **Anna Wiłkomirska** – socjologia edukacji, polityka edukacji, **Małgorzata Karwowska-Struczyk** – edukacja wczesnoszkolna

Adres redakcji: „Kwartalnik Pedagogiczny”,
Uniwersytet Warszawski, Wydział Pedagogiczny,
ul. Mokotowska 16/20, 00-561 Warszawa,
e-mail: kwartalnik.pedagogiczny@uw.edu.pl; http://www.kp.edu.pl

Projekt okładki i stron tytułowych: **Jakub Rakusa-Suszczewski**

Redaktor prowadzący: **Dorota Dziedzic**

Redakcja językowa: **Katarzyna Kaczmarek**

Tłumaczenie i redakcja tekstów angielskich: **Paulina Marchlik**

Redaktor statystyczny: **Jan Łaszczyk**

PL ISSN 0023-5938

© Copyright by Wydawnictwa Uniwersytetu Warszawskiego 2015

Pismo dofinansowane przez Wydział Pedagogiczny Uniwersytetu Warszawskiego

Edycja papierowa jest wersją pierwotną czasopisma

Wydawnictwa Uniwersytetu Warszawskiego
00-497 Warszawa, ul. Nowy Świat 4
http:/www.wuw.pl; e-mail: wuw@uw.edu.pl
Dział Handlowy: tel. (48 22) 55-31-333,
e-mail: dz.handlowy@uw.edu.pl

Skład, łamanie i wersja elektroniczna: **Dariusz Dejnarowicz**

Spis treści

ARTYKUŁY

Janusz Trempała – O pułapkach pomiaru osiągnięć szkolnych: „przeciętny uczeń" nie istnieje 9

Maria Groenwald – *Nadmiar* jako problem diagnostyki edukacyjnej 24

Leszek Korporowicz – Ewaluacja jako komunikacja. Od animacji do alienacji 39

Sylwia Jaskuła – Zastosowania ewaluacji edukacyjnej. Trzy rodzaje koncentracji 56

Grażyna Szyling – Ewaluacja szkolnych systemów oceniania uczniów. Poszukiwanie znaczeń 68

Piotr Maciej Skorupiński – Standardy trafności pomiaru 83

Aleksandra Jasińska-Maciążek, Grzegorz Humenny – Wykorzystanie egzaminów o niskiej doniosłości do oceny efektywności nauczania – możliwości i ograniczenia 97

COLLOQUIA

Bolesław Niemierko – Psychologists as educational diagnosticians. On the track of typology 118

Michał Daszkiewicz – Learner's utterance as a foundation of *personal-experience*-oriented classroom diagnostics 130

Paulina Skórska – Gender gap in reading and writing achievements 141

MATERIAŁY I SPRAWOZDANIA Z BADAŃ

Ewa Maria Filipiak – Zdolności wytwarzania konwergencyjnego i dywergencyjnego uczniów szkoły podstawowej i gimnazjum 155

Barbara Ciżkowicz – Zastosowanie SEM do oceny różnic w uwarunkowaniach osiągnięć matematycznych polskich 15-latków w badaniach PISA 2003 i 2012 172

Zofia Lisiecka – Model eksternistyczny jako przykład *uczenia się przez całe życie* 191

R E C E N Z J E

Piotr Maciej Skorupiński – American Educational Research Association, American Psychological Association, National Council on Measurement in Education, *Standards for educational and psychological testing* 201

Noty o Autorach 204

Contents

ARTICLES

Janusz Trempała – On traps in the measurement of school achievement: an "average student" doesn't exist 9

Maria Groenwald – *Surplus* as a problem of educational diagnostics 24

Leszek Korporowicz – Evaluation as communication. From animation to alienation 39

Sylwia Jaskuła – Applications of educational evaluation. Three types of concentration 56

Grażyna Szyling – Evaluation of school student assessment systems. Searching for meanings 68

Piotr Maciej Skorupiński – Standards of measurement validity 83

Aleksandra Jasińska-Maciążek, Grzegorz Humenny – Psychologists in the quality of educational diagnosticians. On the track of typology 97

COLLOQUIA

Bolesław Niemierko – Psychologists as educational diagnosticians. On the track of typology 118

Michał Daszkiewicz – Learner's utterance as a foundation of *personal-experience*-oriented classroom diagnostics 130

Paulina Skórska – Gender gap in reading and writing achievements 141

MATERIALS AND RESEARCH REPORTS

Ewa Maria Filipiak – Convergent and divergent production abilities of primary and secondary school students 155

Barbara Ciżkowicz – The application of SEM to assess differences in the conditions of mathematics achievements of Polish 15-year-olds in PISA 2003 and 2012 172

Zofia Lisiecka – Extramural general education exams as an example of 'lifelong learning' . . 191

REVIEWS

Piotr Maciej Skorupiński – American Educational Research Association, American Psychological Association, National Council on Measurement in Education, *Standards for educational and psychological testing* 201

Contributors to this issue 204

KWARTALNIK PEDAGOGICZNY
2015 NUMER 4(238)

ARTYKUŁY

Janusz Trempała
Uniwersytet Kazimierza Wielkiego w Bydgoszczy

O pułapkach pomiaru osiągnięć szkolnych: „przeciętny uczeń" nie istnieje

Summary

ON TRAPS IN THE MEASUREMENT OF SCHOOL ACHIEVEMENT: AN "AVERAGE STUDENT" DOESN'T EXIST

The thesis of the article is that by focusing attention on the values of the average measurement of school achievement, we ignore variability, thereby reducing significantly the understanding of dynamics in students' behaviour, development and learning (acquisition of knowledge and skills). The article attempts to show that the approach focused on trends in average is not well fitted to the dynamic nature of phenomena diagnosed in education and can be the source of many artefacts. The pitfalls in which in this approach we most often fall include: (a) a static measurement; (b) averaging the results; (c) the treatment of standard deviation as a measurement error and noise information which is not worth dealing with.

Key words: school achievements, stability and variability, traps in the measurement.

red. Paulina Marchlik

Wprowadzenie

Od XIX wieku żyjemy w kulturze „przeciętności". W badaniach naukowych i praktyce społecznej koncentrujemy uwagę na wartościach średnich i tzw. normach, w przekonaniu, że tendencje przeciętne, którym podlegamy, wyrażają stałość i uniwersalność prawidłowości występujących w otaczającym nas świecie. Mówimy o przeciętnych stanach/właściwościach obserwowanych zjawisk: a) fizyko-chemicznych (np. o przeciętnej pogodzie, prędkości, kształtach czy para-

metrach fizjologicznych organizmu); b) społeczno-ekonomicznych (np. o przeciętnym środowisku, mieszkańcu, wynagrodzeniu, jakości życia, a nawet o urodzie ludzi); c) psychicznych (np. o przeciętnych zdolnościach, cechach, inteligencji, zadaniach i osiągnięciach rozwojowych czy edukacyjnych). Średnie tendencje leżą u podstaw standardów (np. postępowania, zdrowia ludzi, ich rozwoju i starzenia się, produkcji przemysłowej, polityki społecznej itd.), które współcześnie regulują coraz szersze obszary praktyki społecznej.

Nie jest inaczej w rozwijanych od co najmniej 200 lat systemach edukacji. Nie pomylę się, jeśli powiem, że współczesna szkoła jest nastawiona na przeciętność: począwszy od programów nauczania, treści podręczników i ich układu (dostosowanych do możliwości przeciętnego ucznia i standardowych wymagań stanowiska pracy, do której jest przygotowywany), aż do oceny osiągnięć edukacyjnych ucznia/studenta (opartej na średniej jego ocen cząstkowych zgromadzonych w różnym czasie pomiaru, oraz ich porównaniach ze średnią grupową: na tle klasy, szkoły czy kraju). Na podstawie średniej oceniamy i promujemy uczniów oraz podejmujemy działania korekcyjne (zob. Ross, Fischer 2013; Kosicova i in. 2014), ale także opracowujemy rozmaite rankingi, wyłaniające najlepszych uczniów, najlepsze szkoły w kraju, a nawet najlepsze na świecie systemy edukacji (np. słynna PISA). W ostatnim czasie „grę" liczbami przeciętnymi wykorzystuje się nawet do oceny efektywności pracy naukowej uczonych i uczelni, w których pracują[1].

W prezentowanym opracowaniu chcę wykazać, że koncentrując uwagę na wartościach średnich pomiaru osiągnięć uczniów, nie dostrzegamy lub ignorujemy ich zmienność w czasie i różnice indywidualne w tym zakresie. Dwa są tego powody. Po pierwsze, zgodnie z obowiązującą metodologią badań naukowych, przyjmujemy założenie o istnieniu stałych i uniwersalnych zasad zachowania, rozwoju i uczenia się. W tym podejściu ważniejsze jest to, co powtarzalne i predyktywne, niż to, co zmienne i nieprzewidywalne. Po drugie, zwykle stosujemy standardowe modele statystycznej analizy wyników pomiaru, skoncentrowane na tendencjach przeciętnych, które są „nastawione" na identyfikowanie tego, co raczej jest stałe i powszechne, niż zmienne i wyjątkowe/specyficzne. Koło się zamyka. Mimo iż osiągnięcia szkolne nazywamy zmiennymi, to w ich analizie zwykle stosujemy narzędzia statystyczne, które poprzez uśrednianie danych „stabilizują" zmienność. Wydaje się, że nie są one dobrze dopasowane do dynamicznej natury tego, co badamy. Gra liczbami skoncentrowana na tendencjach przeciętnych w pomiarze osiągnięć szkolnych może być więc źródłem

[1] Jest to jednak odrębny problem, który wymaga omówienia w innym miejscu (zob. *numbers game*, Parnas 2007).

wielu artefaktów. Coraz lepiej uświadamiamy sobie, że podejście skoncentrowane na miarach średnich ma ograniczone zastosowanie w zrozumieniu zachowania, rozwoju i uczenia się (nabywania wiedzy i umiejętności) uczniów.

U podstaw prezentowanych rozważań leżą trzy kluczowe założenia: a) istotą zachowania, rozwoju i uczenia się[2], nie jest określony stan, lecz jego zmiana – niezaistnienie zmiany jest niemożliwe; b) podstawową kategorią analizy zmian jest czas – zmiany zachodzą w czasie i ich pomiar nie jest możliwy w oderwaniu od upływu czasu; c) wraz z upływem czasu zmienia się kontekst pomiaru – w różnych jego punktach inny jest nie tylko dany stan jednostki, ale także kontekst (stany otoczenia), który nadaje znaczenie osiągniętym wynikom pomiaru.

Przedstawione założenia wskazują na trzy podstawowe kategorie analizy zachowania, rozwoju i uczenia się człowieka: zmiana, czas pomiaru i jego kontekst. Uprzedzając dalszy wywód, w tym miejscu chciałbym zaznaczyć, że wprawdzie pomiar powtarzany w kolejnych punktach czasu (podłużny) dostarcza dane lepiej dopasowane do dynamicznej natury badanych przez nas zjawisk, ale nie gwarantuje dobrego zrozumienia tego, co badamy. Nie mniej ważnym jest sposób analizy tak zgromadzonych danych.

Podejście dynamiczne

Gdy mówię moim studentom o dynamiczności, często pytają na czym ona polega, albo co znaczy, że badana przez nas rzeczywistość jest dynamiczna? Najprostszą jest odpowiedź, że wszystko to, co badamy stale się zmienia: jednostka, ale także jej otoczenie. Odpowiedź taka zwykle ich zadowala i zwykle godzą się z poglądem, że nauka o zachowaniu i rozwoju musi opisać zasady, wedle których zmiany zachodzą. Dalsze dyskusje na temat tych zasad nie są już jednak tak proste. Badane zjawiska są bowiem funkcją ko-akcji (wielokrotnie powtarzalnego w czasie współoddziaływania) różnych czynników, tworzących całościowo rożne, ale powiązane systemy złożonego mechanizmu/mechanizmów zmiany, którego – trzeba to przyznać – do końca nie rozumiemy. Chcąc wyjaśnić tę złożoność uciekamy się często do autorytetów, przywołujemy powszechnie akceptowane w literaturze teorie albo manifestujemy sceptycyzm naukowy, mówiąc, że obecnie nie wszystko daje się dobrze zbadać i zrozumieć wszystkie zasady, wedle których dokonują się obserwowane zmiany.

[2] Z formalnego punktu widzenia ogólne prawa zachowania, rozwoju i uczenia się są tożsame.

Od co najmniej dwóch dekad wielu badaczy przywiązuje coraz większą wagę do zmienności. Aby uchwycić w spójne ramy analizę stabilności i zmienności złożonych zjawisk, które obserwujemy w naturze, odwołują się do teorii dynamicznych systemów. Wzbudza ona ostatnio zainteresowanie w tak rożnych dyscyplinach jak fizyka, biologia, psychologia czy pedagogika (zob. Ross, Fisher 2013). W psychologii rozwoju pionierami jej stosowania są Esther Thelen (Thelen, Smith 2006), Paul van Geert (Van Geert, Steenbeek 2005) i Kurt Fischer (Fisher, Bidell 2006). W podejściu tym akcentowane są dwie zasady: a) każde zachowanie jest efektem wielu stanów/właściwości osoby, jak i kontekstu; b) zmienność wykonania zadań jest ważną informacją dla zrozumienia zachowania i rozwoju człowieka. Coraz częściej postulowany jest także dynamiczny model interakcji między stale zmieniającą się osobą i stale zmieniającym się otoczeniem, jako dwoma złożonymi i otwartymi systemami (Lerner, Hultsch 1983; Magnusson, Stattin 2006). W ujęciu tym wszystkie komponenty funkcjonowania i rozwoju są włączone w pętlę wieloczynnikowych i wielokrotnie powtarzających się w czasie współdeterminacji (ko-akcji) (zob. Gotlieb 2003).

Można zaryzykować stwierdzenie, że założenie o dynamiczności zachowania i rozwoju jest wśród badaczy współcześnie prawie powszechnie akceptowane, przynajmniej w sferze deklaracji. Gorzej z praktyką gromadzenia i analizy danych empirycznych. Popełniamy tu w sposób mniej lub bardziej świadomy wiele błędów związanych z upływem czasu, w którym dane zjawisko badamy[3]. Prezentowaną poniżej dyskusję ograniczam do trzech pułapek, w jakie najczęściej wpadamy, koncentrując uwagę na wartościach średnich w poszukiwaniu stałych i uniwersalnych prawidłowości. Podejście to, jak postaram się pokazać, ogranicza w sposób istotny naszą wiedzę na temat dynamiczności badanych przez nas zjawisk, w tym osiągnięć szkolnych ucznia.

Pułapka 1: Wnioskowanie o dynamiczności na podstawie statycznej fotografii

Rozpocznę od pułapki jednokrotnego pomiaru interesującej nas „zmienności", typowej dla dominującego w naukach społecznych podejścia korelacyjnego, gdy chcemy poznać mechanizm kształtowania się jakiegoś zachowania

[3] O pułapkach związanych z czasem na różnych etapach procesu badawczego w psychologii rozwojowej zob. Trempała (2007).

i rozwoju czy uczenia się jednostki na podstawie danych zgromadzonych w danym czasie badania t_1. W takich przypadkach zwykle budujemy model teoretyczny badanego zjawiska, przyjmując, że czynniki „tworzące" dany mechanizm „coś robią", na przykład aktywizują (np. bodźce, środowisko), ukierunkowują (np. cele, wartości, potrzeby) czy tworzą warunki (np. geny, drogi nerwowe) przebiegu aktywności obserwowanej w zachowaniu, rozwoju i uczeniu się jednostki. Czynniki te traktujemy jako zmienne, i zgodnie z planem badań poszukujemy między nimi związków/różnic, które zostały przyjęte w założonym przez nas modelu danego zjawiska. Stwierdzone w pomiarze relacje między miarami średnimi tych czynników interpretujemy w kategoriach przyczynowo-skutkowych, wyjaśniając genezę obserwowanego zachowania zgodnie z założeniami przyjętego modelu teoretycznego. W postępowaniu tym zachowujemy się jak fotografowie „zatrzymujący w kadrze" dane zjawisko/zdarzenie, które post factum próbujemy wyjaśnić w oparciu o przyjęte założenia teoretyczne na temat relacji między badanymi zmiennymi.

Na przykład, na podstawie danych publikowanych w literaturze, często przyjmujemy, że w porównaniu z domami rodzinnymi, uczniowie zamieszkujący w domach dziecka, doświadczają ograniczonej stymulacji środowiskowej. Na tej podstawie różnice w zakresie interesującej nas zmiennej (np. w osiągnięciach szkolnych) obserwowane w danym czasie pomiaru, wyjaśniamy efektem miejsca zamieszkania badanych uczniów. Czynimy tak, chociaż wiemy, że nie każdy dom rodzinny jest źródłem stymulacji aktywizującej rozwój dziecka, a domy dziecka mogą dostarczać sprzyjających jego rozwojowi doświadczeń, których nie jest w stanie zapewnić rodzina. Gorzej, w badaniach tego rodzaju często w ogóle nie bierzemy pod uwagę ko-akcji, tj. tego, że interesujące nas czynniki wielokrotnie współoddziałują na siebie w sposób powtarzalny, np. w ciągłych interakcjach nie tylko rodzice mają wpływ na zachowanie dziecka, ale i odwrotnie, zachowania dziecka wpływają na postawy rodziców. Zastanawia również łatwość, z jaką na podstawie statycznych korelacji między np. wykształceniem rodziców lub szerzej statusem socjo-ekonomicznym rodziny (SES), a ocenami szkolnymi dziecka, wnioskujemy o uwarunkowaniach rodzinnych, mimo że nadal nie mamy wyczerpującej wiedzy o roli innych środowisk, w których ono uczestniczy i interakcji między tymi różnymi środowiskami w kształtowaniu się zachowania i rozwoju jednostki (zob. np. Rutter 1990).

Pojawia się pytanie: czy wnioskowanie o dynamice zmian jest uprawnione na podstawie statycznej „fotografii" zjawiska, gdy dysponujemy „fotografią" stanu/właściwości osoby i jego otoczenia, ale w danym, jednym punkcie czasu ich

pomiaru. Skąd bierze się pewność, że te stany/właściwości i relacje między nimi są stałe, powtarzalne? Przedstawione tu przykłady pokazują, że wnioskowanie o funkcjonowaniu czy mechanizmach kształtowania się cech, zdolności czy umiejętności ucznia na podstawie jednokrotnego pomiaru interesujących nas zmiennych i statycznych (statystycznych) między nimi korelacji/różnic, nie jest uprawnione. Można powiedzieć, że w takich przypadkach mamy do czynienia z „nierozwojowymi" lub „pseudogenetycznymi"[4] badaniami rozwojowymi, których wyniki mogą być obciążone błędami pomiaru, związanymi z przyjętymi założeniami teoretycznymi.

Wielu badaczy uważa, że ograniczenia te znikają lub znacząco redukujemy je, gdy zastosujemy plan badań podłużnych (zob. np. Bijleveld i in. 1998; Fisher, Bidell 2006 i in.), tj. będziemy powtarzali pomiar interesujących nas zmiennych w kilku kolejnych punktach czasu. Niestety, nawet wówczas można mówić o nierozwojowych czy pseudogenetycznych badaniach. W ich ramach otrzymujemy zwykle dane „migawkowe" badanego zjawiska, które nie gwarantują jego zrozumienia przy zastosowaniu tradycyjnie stosowanych modeli analizy statystycznej skoncentrowanej na wartościach średnich pomiaru. Nie radzą sobie one bowiem z upływem czasu w pomiarze zmian w zachowaniu i rozwoju człowieka (zob. np. Bergman 2002; Trempała, Olejnik 2011 i in.). Dochodzimy tu do zasygnalizowanej w tytule opracowania pułapki: uśredniania wyników pomiaru zjawisk, które z natury są dynamiczne[5].

Pułapka 2: Wnioskowanie o dynamiczności na podstawie miar średnich

W pułapkę tę wpadamy, gdy wnioskujemy, na przykład, o prawach/zasadach uczenia się czy nauczania (lub wychowania), na podstawie uśrednionych wyników pomiaru interesującej nas zmiennej (np. wiedzy, zdolności, umiejętności), w różnych grupach uczniów, i/lub w różnych sytuacjach, albo w rożnym czasie

[4] Psychologia rozwoju jako dyscyplina, której przedmiotem jest geneza przejawów życia psychicznego, od czasów Piageta, nazywana jest także psychologią genetyczną.

[5] Jeden z młodszych moich kolegów, doskonały psychometra i uznany w literaturze międzynarodowej znawca wielozmiennowych modeli analizy statystycznej, po wysłuchaniu mojego wykładu o ograniczeniach podejścia skoncentrowanego na tendencjach przeciętnych w psychologii, określił to podejście jako „przekleństwo psychologii".

pomiaru (tj. okazjach, zgodnie z wymiarami „skrzynki danych" Cattella). Zakładając liniowość i powtarzalność badanych zjawisk, stosujemy, ostatnio coraz bardziej zaawansowane, wielozmiennowe modele analizy statystycznej, której wyniki traktujemy jako podstawę wnioskowania o stałych i uniwersalnych prawidłowościach.

Na przykład, gdy interesuje nas izolowana funkcja (np. spostrzeganie, rozumienie, empatia) lub umiejętność (np. czytanie, pisanie, liczenie), zwykle poszukujemy przeciętnej tendencji w poziomie wykonania zadań mierzących daną zmienną (np. testem umiejętności czytania) w danej grupie/grupach uczniów (klasie/klasach), danej sytuacji/sytuacjach (szkolnych, pozaszkolnych) i w danym punkcie/punktach czasu pomiaru dydaktycznego (okazjach). Gdy chcemy poznać wzory powiązań między różnymi zmiennymi, to porównujemy ze sobą średnie ich miary w różnych grupach, w różnych sytuacjach i różnych punktach pomiaru. Stosując to podejście, wychodzimy z założenia, że liczby „nie kłamią", a obliczenia komputerowe danych są dokładne i pewne.

W związku z tym, wyobraźmy sobie hipotetyczny przykład badania podłużnego nad efektami nauczania, którego wzory możemy odnaleźć w literaturze naukowej. Załóżmy, że interesują nas postępy uczniów klas początkowych w nauce czytania. Osiągnięcia w tym zakresie mierzymy zwykle testami umiejętności czytania i porównujemy ze sobą średnie grupowe wyników w kolejnych punktach czasu badania. Na podstawie tendencji „przeciętnych" wnioskujemy o postępach grupy badanych uczniów lub/i o skuteczności stosowania określonej metody nauczania, jeśli porównujemy te tendencje w różnych grupach uczniów (klasach), poddanych różnemu treningowi tej umiejętności. Generalnie, tego rodzaju schemat wnioskowania jest poprawny z punktu widzenia obowiązującej metodologii badań naukowych. Stosując się do jej zasad możemy jednak czasami uzyskać dziwne wyniki, nie dające się dobrze wyjaśnić. Wskazują na to m.in. wyniki badań grupy fińskich badaczy skupionych wokół J.-E. Nurmiego (Aunola i in. 2002). Badając poziom sprawności w czytaniu i samoocenę tej umiejętności w trzech punktach czasu pomiaru w okresie półtora roku początkowej nauki szkolnej, stwierdzili zaskakującą prawidłowość: w okresie tym, postęp w nauce czytania nie był istotny statystycznie, chociaż w tym samym czasie wzrosła w sposób istotny samoocena w tym zakresie przez uczniów. W analizie tej pominięto różnice indywidualne w pomiarze początkowym (*baseline measures*) poziomu czytania uczniów rozpoczynających naukę szkolną i skoncentrowano uwagę na stopniu wzrostu miar średnich w kolejnych punktach czasu badania. Błąd ten jednak

naprawiono[6], stosując analizę skupień „skoncentrowaną na osobach" (*by cases*), stwierdzili, że prawidłowość ta dotyczy tylko około 1/3 badanych dzieci (zob. ryc. 5.2, w: Trempała, Olejnik 2011: 131). W analizie tej wyodrębnili dwa dodatkowe skupienia uczniów, przejawiających inne, przeciwstawne i tym samym, znoszące się matematycznie tendencje podłużne (zależne od czasu pomiaru) w zakresie badanych efektów nauki czytania. W konsekwencji wykazali, że prawidłowość dotycząca tylko części badanych może ujawnić się błędnie jako „przeciętna".

Podany przykład pokazuje, że standardowe modele statystycznej analizy średnich grupowych, nie zawsze dobrze radzą sobie z różnicami międzyosobniczymi wyników pomiaru osiągnięć w różnym czasie badania. W konsekwencji wyniki analizy tendencji przeciętnych mogą wskazywać pozorne prawidłowości. Koncentrując uwagę na średniej grupowej, nie możemy być pewni, że badamy rzeczywiste zmiany w osiągnięciach badanych.

Wątpliwości te narastają, gdy przyjrzymy się uważniej jakiemuś rozkładowi uzyskanych wyników pomiaru zmienności w badanej grupie. Stosując metodę analizy „po palcu", dostrzegamy przede wszystkim, że badani różnią się między sobą w sposób mniej lub bardziej istotny, że wprawdzie istnieją jakieś skupienia osób osiągających podobne wyniki, ale nie takie same. Nasuwa się wówczas myśl, zaskakująca z punktu widzenia podejścia skoncentrowanego na miarach średnich, że w badanej przez nas próbie nie istnieje „przeciętna osoba". Na abstrakcyjność koncepcji „przeciętnej osoby" zwrócił uwagę ponad 70 lat temu Kurt Lewin (1936/1966). Dysponując współcześnie narzędziami modelowania komputerowego, okazuje się, że w grupie badanych często nie znajdujemy nawet jednej osoby, której wyniki byłyby zgodne z „przeciętnym" wzorem interesującej nas zmienności, obliczonej przez komputer (zob. np. Molenaar i in. 2002).

W ślad za tymi badawczymi spostrzeżeniami, możemy powiedzieć, że „przeciętny uczeń" nie istnieje. Jeśli nie istnieje „przeciętny uczeń", to trudno jest także utrzymać w mocy przekonanie, że istnieje jakiś „przeciętny" wzór mechanizmu, kształtowania się czy uczenia się. Ludzie różnią się między sobą nie tylko stanami/właściwościami zachowania w danym punkcie czasu, ale także wzorami jego rozwoju. Wydaje się więc, że koncepcja przeciętnej tendencji jest abstraktem bezużytecznym w badaniu dynamiczności.

[6] W gruncie rzeczy, błędu tego badacze nie popełnili, ale wskazali na potencjalne zagrożenia tego rodzaju wnioskowania statystycznego.

Pułapka 3: Wnioskowanie o dynamiczności skoncentrowane na średnich miarach zmian wewnątrz-osobniczych w czasie badania

Trzecia z pułapek, na którą chcę w tym miejscu zwrócić uwagę, jest szczególnym przypadkiem sytuacji poprzednio omówionej, tj. uśredniania wyników pomiaru. Polega na traktowaniu różnic wewnątrzosobniczych, zachodzących w czasie badania, jako błędu pomiaru. Zwykle jest on utożsamiany z „szumem informacyjnym", którym nie warto się zajmować. W ten sposób tracimy jednak ważne dane na temat zmian wewnątrzosobniczych, zachodzących w trakcie badania. Tym samym, tracimy dane na temat specyficznych efektów procesu uczenia się (tj. gradient wprawy, transfer itp.). Nie wnikając tu w szczegóły, rozważmy dwa przypadki, które rodzą wątpliwości.

Wartość predyktywna średnich wartości pomiaru. Średnie wyniki pomiaru interesującej nas zmiennej traktujemy w psychologii jako dobry predyktor wykonania przez jednostkę tego samego rodzaju zadań lub podobnych. Analogicznie, średnie osiągnięcia szkolne są traktowane jako miary dyspozycji ucznia, gwarantujące jego osiągnięcia w przyszłej edukacji. Czynimy tak w przekonaniu, że istnieją stałe zasady, regulujące zachowanie i rozwój człowieka, na podstawie których potrafimy przewidywać osiągnięcia jednostki. Równocześnie jednak, jesteśmy świadomi ograniczeń tego założenia, co często możemy „wyczytać" z dyskusji wyników uzyskanych przez autorów tego rodzaju badań.

Na przykład, panuje powszechna zgoda, że na podstawie średniej odległości rzutów oszczepem przez daną osobę, trudno jest przewidzieć, jaki wynik osiągnie ona w kolejnym rzucie. Oszczep może polecieć dalej lub bliżej niż przeciętnie, czasami nawet w sposób znaczący. Zależy to od wielu czynników, zarówno od stanu/stanów oszczepnika, jak i otoczenia. Odchylenia te nie zawsze jednak są przypadkowe. Duża część zmienności osiągnięć może wykazywać regularność, specyficzną dla danego zawodnika. Na przykład, jeden z zawodników zwykle posyła oszczep dalej w pierwszych próbach, a drugi w ostatnich. Pierwszy poprawia rzuty po niepowodzeniu, drugi – odwrotnie. Pierwszy rzuca dalej, gdy ma więcej „widzów/kibiców", drugi – na zawodach niższej rangi, itd. Mimo podobnej średniej osiągnięć w serii rzutów tych dwóch zawodników, różnią się oni między sobą. Dynamiczne procesy kształtują każdą aktywność, nie tylko sportową. Trener sportowy, który nie uwzględni specyficznych dla danego zawodnika dynamicznych prawidłowości jej kształtowania się w czasie, nie odnosi sukcesów, jest złym trenerem.

W tym kontekście dziwi, dlaczego szkoła, jej programy i diagnostyka, oparta jest na ocenie przeciętnych osiągnięć edukacyjnych ucznia. Gorzej, średnia ocena z danego przedmiotu lub ogólna średnia z ocen w danym roku szkolnym, jest traktowana jako predyktor dalszych postępów edukacyjnych ucznia. Praktyka pokazuje, że wybitni uczniowie szkół podstawowych, często nie są nimi w szkołach wyższego szczebla (średniej i wyższej), a wybitni absolwenci szkół wyższych, nie zawsze pomyślnie radzą sobie w rozwoju kariery zawodowej.

Wartość średnia wcześniejszych osiągnięć szkolnych ucznia jest więc mało użyteczna w przewidywaniu jego dalszych losów edukacyjnych i zawodowych.

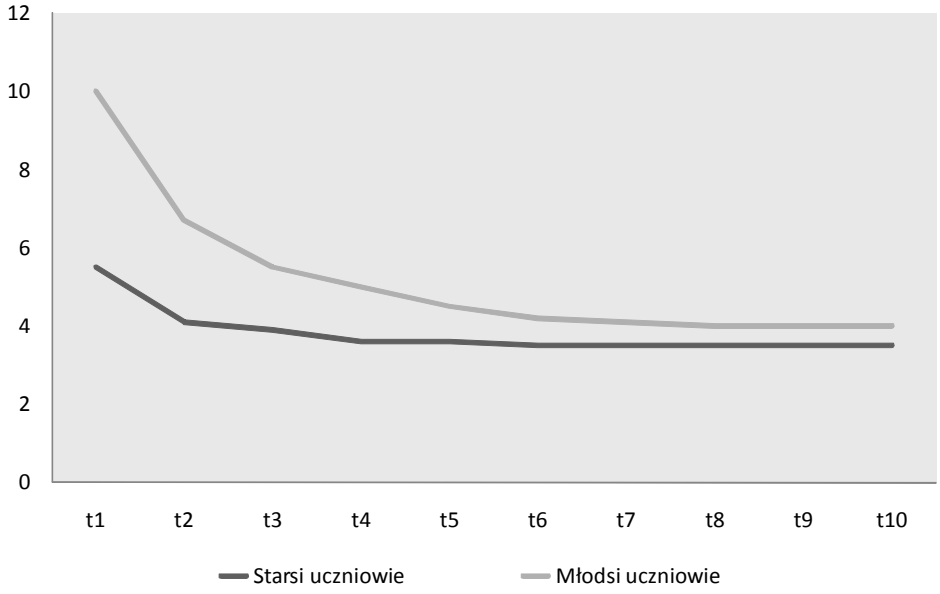
Trafność oceny różnic między średnimi grupowymi. Z problemem tym często mamy do czynienia w badaniach laboratoryjnych lub tzw. studiach treningowych (*training studies*), w których badani uczniowie w różnym wieku życia lub w różnych warunkach pomiaru (np. różnych programów edukacyjnych), wykonują w kolejnych próbach takie same zadania. W takich przypadkach, badacze zwykle ograniczają poszukiwanie zasad uczenia się do średnich osiągnięć w grupach uczniów, np. ze względu na wiek życia lub warunki pomiaru[7].

W tym kontekście wyobraźmy sobie hipotetyczny przykład prostego badania treningowego, którego wzór, niestety, możemy odnaleźć w literaturze uznawanej za wpływową. Na przykład, młodsi i starsi uczniowie wykonują takie same lub podobne zadania (np. uczą się rozpoznawania właściwej odpowiedzi na bodziec, rozpoznawania zgłoski w wyrazie lub recytacji wiersza, itp.) w 10 kolejnych próbach/powtórzeniach. Z rozkładu wyników pomiaru osiągnięć w kolejnych próbach wynika, że początkowo młodsi uczniowie osiągają zdecydowanie gorsze wyniki, ale pod koniec serii powtarzanego pomiaru, różnice ze starszymi zbliżają się do nieistotnych. Jednak przeciętna różnica, między średnimi wynikami grupowymi w serii powtarzanych pomiarów, jest istotna statystycznie na korzyść uczniów starszych. Na tej podstawie badacz zwykle dochodzi do wniosku, że młodsi uczniowie gorzej radzą sobie z tego rodzaju zadaniami niż starsi.

Generalnie, wniosek ten jest poprawny z punktu widzenia założeń teoretycznych i metodologicznych, przyjętych przez autora badań. Zwróćmy jednak uwagę na to, że chociaż schemat tego badania ma dynamiczny charakter, to uwzględniając ograniczenia podejścia skoncentrowanego na średnich wartościach, o których była mowa powyżej, wnioskowanie ma statyczny charakter. Ignorujemy efekty powtarzania wykonania zadań w czasie, redukując wiedzę

[7] W schemacie eksperymentalnym wprowadzamy dodatkowo grupy kontrolne: a) osób biernie oczekujących w czasie między *pretestem* i *posttestem* (kontrola nieaktywna), oraz b) osób wykonujących w tym czasie zadania *placebo* (aktywna kontrola).

o uczeniu się badanych w różnym wieku życia do „jednej różnicy", tj. między wartościami średnimi osiągnięć w grupach wiekowych, pomijając zmiany związane z czasem badania. W badanej rzeczywistości mamy jednak do czynienia z 10-cioma malejącymi w czasie badania różnicami międzygrupowymi, od bardzo istotnych na początku, do nieistotnych (lub prawie nieistotnych) na końcu powtarzanego pomiaru. Mamy w takich przypadkach do czynienia z ważną dla wnioskowania o zasadach uczenia się tendencją: zmniejszania się średnich różnic wiekowych wykonania danego rodzaju zadań wraz z liczbą powtórzeń pomiaru osiągnięć (zob. ryc. 1). Analogiczne przykłady odnajdziemy w badaniach eksperymentalnych, w których poszukuje się różnic między funkcjonowaniem poznawczym młodych dorosłych i ludzi starszych (zob. np. Salthouse 1994).



Ryc. 1. Liczba błędów w wykonaniu hipotetycznego zadania przez uczniów młodszych i starszych w kolejnych punktach czasu pomiaru t_1–t_{10}

Wykres przedstawiony na ryc. 1 obrazuje dwie ważne prawidłowości, które możemy przeoczyć, koncentrując uwagę na przeciętnych wartościach osiągnięć badanych grupach uczniów w powtarzanym pomiarze. Mianowicie pokazuje, że: a) ćwiczenie tego, co mierzymy powoduje poprawę w obu grupach wiekowych; b) gradient poprawy wykonania kolejnych zadań jest większy w grupie uczniów młodszych niż starszych, mimo że obliczona średnia osiągnięć pod wpływem przeprowadzonego treningu (uczenia się) jest wyższa w grupie uczniów

starszych w porównaniu z młodszymi[8]. Nikt nie zaprzeczy, że pominięcie tych informacji jest szkodliwe w poszukiwaniu zasad zachowania, rozwoju, uczenia się. Nie chodzi tylko o to, że powtarzając pomiar, aktywizujemy to, co mierzymy, że zachodzi poprawa w wyniku uczenia się tego, co mierzymy. Wiemy o tym, od czasów Wygotskiego. Prace publikowane w latach 60. i 70. XX wieku przez Schaie, Baltesa czy Nesselroade'a pokazały jak można oddzielać od siebie efekty powtarzania pomiaru od efektów wieku życia badanych we wnioskowaniu o zmianach w ludzkim zachowaniu i rozwoju. Przedstawiony przykład uświadamia kwestię nie mniej ważną. Pominięcie w tej analizie, np. różnic w gradiencie poprawy, powoduje, że tracimy okazję przedyskutowania wielu hipotez dotyczących osiągnięć badanych uczniów, tj. roli poziomu początkowego pomiaru (*baseline measures*), czy plastyczności rozwojowej (podatności na oddziaływania i zakresu modyfikowalności zachowania, zob. Baltes, Schaie 1977; Baltes 1987 i in.) powiązanych z wiekiem życia.

Uogólniając, możemy powiedzieć, że jeśli w rzeczywistości nie istnieje „przeciętny uczeń", ani „przeciętna grupa uczniów", to tym bardziej nie istnieje „przeciętna różnica", której koncepcja także wydaje się być abstraktem o ograniczonym zastosowaniu w analizie zachowania, rozwoju i uczenia się.

Zamiast zakończenia

W prezentowanym opracowaniu w zasadzie nie powiedziałem niczego nowego, biorąc pod uwagę dyskusje na ten temat, które od dawna są obecne w literaturze. Podsumowując najważniejsze dokonania psychologii w XX wieku, wskazałem dwa istotne „przesunięcia", jakie w poprzednim stuleciu dokonały się w badaniach nad zachowaniem i rozwojem człowieka (Trempała 2001a, 2001b). Pierwsze, dokonały się w sferze badań nad naturą badanych przez nas zjawisk: a) z tego, co jest **stałe**, w kierunku tego, co jest **zmienne** i **plastyczne** oraz b) z tego, co jest **uniwersalne**, na to, co **wyjątkowe** i **specyficzne**. Natomiast drugie przesunięcie, w sferze mechanizmów zachowania i rozwoju następuje c) od badań nad **reaktywnością** jednostki w kierunku badań nad jej **aktywnością**

[8] Należy podkreślić, że w innych grupach wiekowych, innych sytuacjach pomiaru lub w badaniu innych funkcji czy procesów, prawidłowość ta może być odwrotna, ale mogą ujawnić się także inne zasady, zamaskowane w analizie tendencji przeciętnych, które dla jasności tu prezentowanych treści pomijam.

własną, jaką podejmuje ona w interakcjach z otoczeniem. Powyższe przesunięcia zainteresowań badawczych wyrosły głównie z dostrzeżenia wielowymiarowej, systemowej organizacji świata oraz dynamicznych powiązań między różnymi elementami całości, której człowiek jest częścią.

„Przesunięcia" te spowodowały, że obserwowana przez nas rzeczywistość zaczęła jawić się nam jako złożona kompozycja aktywnych, „dziejących się" zdarzeń, w ciągłym procesie ich powstawania, rodząc wrażenie płynności, nieogarnioności oraz niepewności w przewidywaniu ludzkiego zachowania i rozwoju. Nacisk w tym ujęciu na stale zmieniającą rzeczywistość ostro kontrastuje z dążeniem nauki do wiedzy o niezmiennym, dającym się określić porządku. Istnienie stałych i uniwersalnych praw wydaje się wątpliwe, ale nie dają się one wykluczyć (zob. Jaeger, Rosnow 1988), szczególnie wtedy, gdy uznamy, że wprawdzie niektóre z obserwowanych zmian wyrażają nowość, nieokreśloność i wyjątkowość, ale inne – kompletność, spójność i liniowy porządek (zob. koncepcja totalnej zmiany Overton 2006).

W prezentowanym opracowaniu starałem się pokazać, że standardowe metody pomiaru i analizy statystycznej skoncentrowane na miarach średnich nie są dobrze dopasowane do dynamicznej natury badanych przez nas zjawisk i mogą prowadzić do artefaktów w ich badaniu. Wydaje się, że niedoskonałość tych metod oraz nasze przyzwyczajenie do skupiania uwagi na miarach średnich, to dwa główne powody tego, że nadal „błądzimy" w płynności, nieogarnioności i niepewności obserwowanych zdarzeń. Na szczęście, wzrost zainteresowania badaczy problemem zmienności poszerzył w naukach o życiu wiedzę o ograniczeniach podejścia skoncentrowanego na tendencjach przeciętnych, co przynosi skutek w postaci doskonalenia planów gromadzenia danych oraz poszukiwania matematycznych modeli ich analizy, które są coraz lepiej dopasowane do dynamicznej natury badanych zjawisk (Trempała, Ciechuch, w przyg.).

Bibliografia

Aunola K., Leskinen E., Onatsu-Arvilommi T., Nurmi J.-E. 2002. *Three methods for studying developmental changes: A case of reading skills and self-concept*, „British Journal of Educational Psychology", 72, s. 343–364.

Baltes P.B., Schaie K.W. 1977. *On the plasticity of intelligence in adulthood and old age: When Horn and Donaldson fail*, „American Psychologist", 31, s. 720–725.

Baltes P.B. 1987. *Theoretical propositions of life-span developmental psychology. On the dynamics between growth and decline*, „Developmental Psychology", 23, s. 611–626.

Bergman L.R. 2002. *Studying processes: Some methodological considerations*, [w:] *Paths to the successful development. Personality in the life course*, red. L. Pulkinnen, A. Caspi, Cambridge University Press, Cambridge, s. 177–202.

Bijleveld C.C.J., van der Kamp L., J.Th. *with* Mooijaart A., van der Kloot W.A., van der Leeden R., van der Burg E. 1998. *Longitudinal data analysis: Designs, models and methods*, Sage Publications, London–Thousand Oaks–New Delhi.

Fischer K.W., Bidell T.R. 2006. *Dynamic development of action and thought*, [w:] *Theoretical models of human development. Handbook of child psychology*, t. 1, red. W. Damon, R.M. Lerner, Wiley, New York, s. 313–399.

Gottlieb G. 2003. *Probabilistic epigenesis of development. The learning brain*, [w:] *Handbook of developmental psychology*, red. J. Valisner, K.J. Connoly, Sage, Thousand Oaks, s. 3–17.

Geert von P., Steenbeek H. 2005. *Explaining after by before: Basic aspects of a dynamic systems approach to the study of development*, „Developmental Review", 25, s. 408–442.

Jaeger M., Rosnow R. 1988. *Contextualism and its implications for psychological inquiry*, „British Journal of Psychology", 79, s. 63–75.

Kosicova V., Liberska H., Farnicka M. 2014. *Formative function of assessment in a psycho-didactical context*, „Polskie Forum Psychologiczne", 19, 1, s. 716.

Lerner R., Hultsch D.F. 1983. *Human development*, McGraw-Hill Book Company, New York.

Lewin K. 1936/1966. *Principles of topological psychology*, McGraw-Hill Book Company, New York.

Magnusson D., Stattin H. 2006. *The person in context: a holistic-interactionistic approach*, [w:] *Theoretical models of human development. Handbook of child psychology*, t. 1, red. W. Damon, R.M. Lerner, Wiley, New York, s. 45–61.

Molenaar P.C.M., Huizenga H.M., Nesselroade J.R. 2002. *The relationship between the structure of inter-individual and intra-individual variability: A theoretical and empirical indication of developmental system theory*, [w:] *Understanding human development*, red. U.M. Staudinger, U. Lindberger, Kluwer Press, Dordrecht, s. 339–360.

Overton W.F. 2006. *Developmental psychology: philosophy, concepts, methodology*, [w:] *Theoretical models of human development. Handbook of child psychology*, t. 1, red. W. Damon, R.M. Lerner, Wiley, New York, s. 18–88.

Parnas D.L. 2007. *Stop the numbers game. Counting papers slaws the rate of scientific progress*, „Communications of the ACM", 50, 1, s. 19–21.

Ross L.T., Fischer K.W. 2013. *Dynamic systems theory*, [w:] *The child: An encyclopedic companion*, red. R.A. Shweder i in., University of Chicago Press, Chicago.

Rutter M. 1990. *Psychosocial resilience and protective mechanisms*, [w:] *Risk and protective factors in the developmental psychopathology*, red. J.E. Rolf, A.S. Masten, D. Cicchetti, K.H. Neuchterlein, S. Weintroub, Cambridge University Press, New York, s. 181–214.

Sternberg R.J., Roediger III H.L., Halpierin D.F. red. 2007. *Critical thinking in psychology*, Cambridge University Press, Cambridge.

Thelen E., Smith L.B. 2006. *Dynamic systems theories*, [w:] *Theoretical models of human development. Handbook of child psychology*, red. R.M. Damon, R.M. Lerner, t. 1, Wiley, New York, s. 313–399.

Trempała J. 2001a. *Dwa przełomy w badaniach nad rozwojem psychicznym człowieka*, „Przegląd Psychologiczny", 44, s. 85–92.

Trempała J. 2001b. *Najważniejsze dokonania psychologii rozwojowej w mijającym wieku*, „Studia z Psychologii KUL", 10, s. 369–382.

Trempała J. 2007. *Methodological dilemmas in the use of the time as a variable in developmental psychology*, „Polish Psychological Bulletin", 38, s. 117–125.

Trempała J., Olejnik M. 2010. *Badanie rozwoju psychicznego człowieka*, [w:] *Psychologia rozwoju człowieka*, red. J. Trempała, Wydawnictwo Naukowe PWN, Warszawa, s. 127–143.

Trempała J., Cieciuch J. (w przyg.). *Analiza zmian w zachowaniu i rozwoju człowieka. O błędach w badaniu dynamiczności*.

KWARTALNIK PEDAGOGICZNY
2015 NUMER 4(238)

Maria Groenwald
Uniwersytet Gdański

Nadmiar jako problem diagnostyki edukacyjnej

Summary

SURPLUS AS A PROBLEM OF EDUCATIONAL DIAGNOSTICS

The article presents an analysis of the phenomenon of *surplus* in educational diagnostics and strategies of coping with it adopted by diagnosticians (and teachers). An application of analysis of interpretative practice allowed for a characterisation of *surplus* on several strata in terms of its: location, possession, type of understanding and valid prescriptivism. The analysis led to the recognition of strategies employed by diagnosticians with a view to combatting it, not perceiving it, using it as a new area of diagnostic search carried out within the educational reality.

Key words: *surplus* in educational diagnostics, strategies of diagnosticians.

red. Paulina Marchlik

Na przestrzeni zaledwie kilkunastu ostatnich lat można zaobserwować, jak niedostatek wiedzy o postępach lub osiągnięciach uczniów, jakości pracy szkół oraz efektywności innowacji edukacyjnych, ulega sukcesywnemu zastępowaniu przez jego przeciwieństwo, czyli wielość informacji, opinii i formułowanych na tej podstawie zaleceń, kierowanych głównie do nauczycieli jako wykonawców. Zmieniają się też umiejętności diagnostów w pozyskiwaniu danych, bowiem miejsce intuicyjności zajmuje poprawne i zróżnicowane metodologicznie diagnozowanie, pozwalające na opracowywanie rozlicznych diagnoz; ich wielorakość w pewnych sytuacjach może wręcz implikować wrażenie *nadmiaru*. Czym jest ów *nadmiar* oraz jakie są strategie radzenia sobie z nim w codziennych praktykach diagnostycznych – to zagadnienia, które podejmuję w niniejszym opracowaniu.

1. Założenia metodologiczne

Zgodnie z literaturą przedmiotu *nadmiar* oznacza jednostkowe doświadczenie podmiotu w relacji do otaczającego świata. Poprzez „wytwarzanie" lub „obcowanie" z nim, określa się go jako:

– nadwyżkę, czyli to, co jest czymś innym i czymś więcej, co jednak niekoniecznie musi być bytem policzalnym lub mierzalnym;
– „nadprogramową" pozostałość, resztę z czegoś;
– przewidywanie tego, czego może okazać się za wiele, choć mogłoby być mniej (Wolińska 2014: 59).

Bernard Waldenfels (2009: 122) *nadmiarem* nazywa to, co przerasta zastane warunki sensu[1] i prawidłowości oraz odbiega od nich i jest doznawane w kategoriach różnicy. Opisuje *nadmiar* przez pryzmat znaczeniowych kontrastów:

1) ze względu na usytuowanie (miejsce), ukazuje go jako to, co jest zewnętrzne wobec wewnętrznego (zewnętrzne *vs* wewnętrzne),
2) z uwagi na posiadanie, jako to co należy do innego w przeciwieństwie do własnego (obce/wyobcowane *vs* własne),
3) ze względu na rodzaj rozumienia, jako to, co jest innego rodzaju, odmienne, osobliwe w przeciwieństwie do swojskiego (odmienne *vs* swojskie) (tamże: 109–110),
4) ze względu na obowiązującą normatywność, jako to co nieprzewidywalne i odbiegające od standardu (nienormowane *vs* reguły, porządki, prawa) (tamże: 23).

Przyjęte w niniejszym opracowaniu postępowanie badawcze poprowadzę pod kątem *nadmiaru*: a) ujawnianego w doniesieniach naukowych i raportach z badań diagnostów; b) przywoływanego podczas paneli dyskusyjnych prowadzonych w ramach Konferencji Diagnostyki Edukacyjnej. Takie badanie wpisuje się w proponowane przez Jamesa A. Holsteina i Jabera F. Gubriuma (2009: 678–694) oraz Harolda Garfinkla (2007: 99–102) podejście określane mianem analizy praktyki interpretacyjnej, umożliwiającej śledzenie edukacyjnej empirii przez pryzmat stosowanych praktyk diagnostycznych. Cenną własnością tego ujęcia jest jego otwartość na modele i procedury badawcze wykorzystywane zarówno w ujęciu

[1] Przy czym sens, może tu być rozumiany wielorako, jako: (a) sens ontologiczny, czyli przyjmowany porządek świata; (b) sens aksjologiczny i agatologiczny; (c) ze względu na „znaczenie", „treść nazwy"; (d) sens teoriopoznawczy, np. odzwierciedlony w paradygmatach i przyjmowanej w nich normatywności (za: Stróżewski 2005: 465–466).

ilościowym, jak i jakościowym. Zaznaczam, że użycie analizy praktyki interpretacyjnej nie oznacza próby łączenia mikro z makroteorią, ani diagnozowania z bliska z diagnozowaniem z daleka, ale odzwierciedla dążenie do pełniejszego zrozumienia problemów stojących przed współczesną diagnostyką edukacyjną. Ich złożoność – czego przykładem jest właśnie zjawisko *nadmiaru* – implikuje potrzebę sięgania po nowe, zhybrydyzowane i suplementarne analizy edukacyjnej rzeczywistości, prowadzone na przecięciu instytucji, kultury i społecznych interakcji oraz koncepcji teoretycznych, jak również łączenia dyscyplin i dziedzin naukowych. Zaletą takiego postępowania badawczego staje się otwarcie na komplementarność podejść, umożliwiające wychwytywanie swoistych znaczeń nadawanych zjawiskom społecznym i edukacyjnym przez diagnostów jakościowych, ale dopełnianych diagnozami zobiektywizowanymi, i odwrotnie. W praktyce obrane podejście jest realizowane w formie naprzemiennie przeplatających się etapów analizy: diagnoz z bliska (wpisujących się w paradygmat interpretatywny) z diagnozami z daleka (paradygmat normatywny, pozytywistyczny), gdzie cechą tego przeplatania jest dialektyka dwóch pól działania, przy równoczesnym odrzuceniu reprodukcji jednego przez drugie.

Jako materiał do analiz wybrałam najnowsze prace diagnostów opublikowane w ciągu ostatnich dwu lat w ramach Seminarium Diagnozy i Ewaluacji (2013 rok) oraz Konferencji Diagnostyki Edukacyjnej (2013 rok i 2014 rok), w celu zobrazowania i opisania z ich pomocą zjawiska *nadmiaru* ujawnianego w badaniach edukacyjnych. W doborze tych publikacji kierowałam się aktualnością podejmowanych problemów z zakresu diagnostyki edukacyjnej oraz zawartością treściową doniesień wskazujących, między innymi, na obecność zjawiska *nadmiaru* w rozpatrywanych zagadnieniach (Rubacha 2008: 160). Analiza tych opracowań, prowadzona przez pryzmat ramy pojęciowej Waldenfelsowskiej definicji *nadmiaru*, umożliwi odczytanie znaczeń nadawanych temu zjawisku przez diagnostów prowadzących diagnozowanie z bliska i z daleka oraz pozwoli wyłonić strategie radzenia sobie z nim. Dodam, że przywołane materiały źródłowe nie wyczerpują wszystkich dostępnych danych i są jedynie przykładami ilustrującymi prowadzony wywód.

2. Aspekty *nadmiaru* w diagnostyce edukacyjnej

Zgodnie z przyjętym założeniem, zjawisko *nadmiaru* w diagnostyce edukacyjnej rozpatruję uwzględniając kilka jego aspektów: 1) usytuowanie, 2) posiadanie, 3) rodzaj rozumienia, 4) obowiązującą normatywność. Choć nie są one

rozłączne i często wykazują powiązanie, jednak na użytek analizy dokonam ich umownego rozdzielenia.

2.1. *Nadmiar* wynikający z usytuowania diagnozowania: zewnętrzne *vs* wewnętrzne

Ze względu na umiejscowienie, *nadmiar* w diagnozowaniu może przejawiać się jako to, co jest zewnętrzne wobec własnego, wewnętrznego. Reakcją na doświadczenie tego, patrząc z perspektywy osób prowadzących diagnozowanie z bliska, np. nauczycieli, może być określona postawa wobec wyników badań zewnętrznych: egzaminów, sprawdzianów i ewaluacji oraz implikowane nią konsekwencje. Zazwyczaj traktują je jako rodzaj oceny zewnętrznej, z którą jednak nie zawsze w pełni się zgadzają, co prowadzi – jak pisze Grażyna Szyling – do konfrontacji diagnoz z bliska z diagnozami z daleka. W jej wyniku pedagodzy obierają jedną z dwu strategii: przetrwania „pozwalającego w miarę bezpiecznie doczekać następnych zmian polityki oświatowej" (2013: 64), bądź adaptacji. Przykładu praktyk adaptacyjnych dostarcza Małgorzata Żytko opisująca badania OBUT (Ogólnopolskie Badania Umiejętności Trzecioklasistów); prowadzone w celu „stopniowego wyposażania nauczycieli w instrumenty badawcze pozwalające im samodzielnie opisywać efekty własnej pracy i modyfikować praktykę edukacyjną" (2013: 115), są przez nich odbierane jako specjalny sprawdzian weryfikujący umiejętności dzieci, który w kolejnych latach może skutkować ukierunkowaniem kształcenia trzecioklasistów pod test.

Strategię adaptowania egzaminów obiera też większość nauczycieli i uczniów, zwłaszcza z gimnazjów i szkół ponadgimnazjalnych. Zofia Lisiecka i Wioletta Kozak (2014: 88–89) przytaczają przypadek szkolnej polonistyki jako przedmiotu w przeszłości broniącego się przed ocenianiem analitycznym silniej niż tzw. przedmioty ścisłe. Zamierzała bowiem zachować w ten sposób swoją wielowymiarowość w zakresie treści kształcenia, ukierunkowanego także na rozwój postaw uczniów, wspieranie ich indywidualności i kreatywności, przy zachowaniu różnorodności metodologicznej. Jednak z czasem sprzeciw słabł, wręcz dostrzegano korzyści, jakie daje kształcenie pod test, osiągane nawet za cenę oszustwa (Groenwald 2014: 44–45). Obecnie, w wyniku wieloletniego podporządkowywania kształcenia wymogom egzaminacyjnym, wiedza i umiejętności wykraczające poza standard są w tych przedmiotach uznawane za zbędne, traktowane w kategorii *nadmiaru* niewartego nauczania-uczenia się i diagnozowania.

Podsumowując: W opisach działań podejmowanych wobec *nadmiaru* ukazanego ze względu na jego usytuowanie, pojawiają się określenia właściwe językowi rozwiązań siłowych: konfrontacja, opór, obrona, przetrwanie, itd. Wynika z nich, że to, co wewnętrzne broni się przed przymusem przebudowy dotychczasowego kształcenia pod wpływem diagnozy z zewnątrz. Jednak ostatecznie ugina się wobec niej i – paradoksalnie – to co własne czyni zbędnym naddatkiem, stosując mimikrę bądź adaptację, jako taktyki biernego oczekiwania na dalszy bieg zdarzeń. Z perspektywy nauczycieli diagnozujących z bliska, realizowane z daleka diagnozowanie instytucjonalnie, zazwyczaj odbierane jest jako wrogie, szkodliwe, nietrafne, a przy tym opiniotwórcze i ważkie z uwagi na długofalowe konsekwencje. Warto pamiętać, że podobnie, czyli jako szkodliwe i niedoskonałe (zwłaszcza psychometrycznie), diagnozowanie z bliska jest postrzegane przez diagnozujących z daleka.

2.2. *Nadmiar* ze względu na posiadanie: należące do innego, „obce" *vs* własne

Rozdźwięk między diagnostami prowadzącymi badania z bliska (np. nauczycielami) a diagnozującymi z daleka, wyraźnie uwypuklają te opracowania, w których badania własne (realizowane w obrębie klasy) stają się punktem odniesienia do diagnozowania przez innych, odbieranego jako „obce", z trudem oswajane i z tego też powodu obdarzane mianem zbędnego *nadmiaru*, najchętniej eliminowanego ze szkolnej codzienności. Zdaniem Jerzego Chodnickiego (2014: 94), dla nauczycieli typowe jest dystansowanie się wobec wniosków formułowanych w ramach diagnozowania zewnętrznego, gdyż nie w pełni odpowiadają one potrzebom ich warsztatu pracy. W związku z tym z diagnoz zewnętrznych, jako „obcych", jedynie wybiórczo niektóre ich fragmenty włączają do pracy z uczniami, co można interpretować: (po pierwsze) jako próbę obrony „siebie i uczniów przed zostaniem zakładnikami jednej, niezmiennej koncepcji lub wizji edukacji" (Szyling 2013: 56); (po drugie) jako milczący protest przeciwko pomijaniu nauczycieli w decyzjach dotyczących zakresu badań i ich w nim uczestnictwa (Żytko 2013: 115).

Co ciekawe, podobne znaczenie jako bezwartościowych, głównie z racji niepoprawności metodologicznej, diagności profesjonalni nadają diagnozom nauczycielskim. Tymczasem ich postępowanie, istotnie niejednokrotnie oparte na wiedzy zdroworozsądkowej i doświadczeniu pedagogicznym, Bolesław Niemierko (2013a: 13–15) tłumaczy koniecznością podejmowania w klasie licznych (około

tysiąca dziennie), a więc i szybkich decyzji przy informacjach niepełnych, gdyż nie poprzedzanych poprawnym metodologicznie diagnozowaniem (nierealnym w ferworze szkolnych zdarzeń). Bywa, że w pracy kierują się w tzw. mądrością pokoju nauczycielskiego, a kiedy już skłaniają ku zewnętrznym wynikom badań, bądź racjom politycznym czy ekonomicznym, czynią to nierefleksyjnie (Niemierko 2014: 88). Zarazem traktowanie tego, co „obce" jako zbędnego balastu lub przeszkody w kształceniu, utrudnia wyjście poza rutynę, wszczęcie poszukiwań nowych rozwiązań problemów szkolnej codzienności, oswojenie odmienności i uczynienie jej „własną".

Podsumowując, jako diagnostyczny *nadmiar* mogą być traktowane diagnozy z daleka, doświadczane przez nauczycieli w kategoriach: a) przeciwieństwa własnych diagnoz; b) czynnika zagrażającego zburzeniem poukładanego świata szkolnej codzienności. Podobnie, czyli jako „obce", ale i też mało wartościowe i z tego powodu nieprzydatne, badacze z daleka uznają diagnozy nauczycielskie. Pojmowanie przez każdą ze stron tego, co „obce" jako niewartościowe czy dekonstruujące dobrą edukację, może być właściwą strategią wówczas, gdy „obce" istotnie raczej będzie szkodziło kształceniu niż wspomagało je. Jednak traktowanie „obcego" jako *nadmiaru*, może prowadzić na diagnostyczne i dydaktyczne bezdroża, sprzyjające zamknięciu się w rutynie niczym w twierdzy i odpieraniu nowych możliwości problematyzowania własnych doświadczeń, interpretowania ich na nowo, przekładania oraz włączania innowacyjnych metod badań do własnego warsztatu diagnostycznego.

2.3. *Nadmiar* ze względu na rodzaj rozumienia: swojskie, oswojone *vs* nieswoje, osobliwe

Konfrontacja diagnostów preferujących badania z bliska z diagnozującymi z daleka ujawnia *nadmiar* jako szkodliwą cechę diagnoz formułowanych poza własnym środowiskiem diagnostycznym. W związku z tym diagnozujący z daleka postrzegają diagnozy z bliska jako subiektywne, ukierunkowane na przypadki lub zanurzone w etnografii, a więc jako całkowicie odmienne od ilościowych, zobiektywizowanych i pretendujących do uogólnień. I odwrotnie, prowadzący diagnozowanie z bliska prezentują dystans wobec diagnoz z daleka. Jako przykład Z. Lisiecka i W. Kozak (2014: 83) przywołują opinie nauczycieli o egzaminach zewnętrznych, zazwyczaj określanych jako niszczące i szkodliwe. Autorki zwracają przy tym uwagę, że poglądy te są formułowane wyłącznie w oparciu

o „powszechne mniemania", „intuicję pedagogiczną", „doświadczenie dydaktyczne", czy „przyczynkarskie doniesienia z mikrobadań szkolnych bądź regionalnych", zatem nie są opiniami opartymi na rzetelnych badaniach naukowych. Z perspektywy diagnostyki edukacyjnej, wyłaniające się tu nieporozumienie, stanowi pochodną zastosowanych podczas diagnozowania metod badań, dobieranych ze względu na przyjmowane założenia ontologiczne i epistemologiczne, leżące u podstaw konstruowanej rzeczywistości (Kincheloe, McLaren 2009: 455).

Wydawać by się mogło, że w pracy nauczyciela niezbędne jest uwzględnianie diagnoz i z bliska, i z daleka, gdyż – z uwagi na odmienne perspektywy poznawcze – dostarczają one danych komplementarnie dopełniających się. Jednak w praktyce uzyskane dzięki nim informacje są przez pedagogów doświadczane jako problematyczny *nadmiar* danych. Z analiz G. Szyling (2014) wynika, że niezależnie od tego, czy pochodzą one z badań jakościowych, czy z ilościowych, wykorzystywane są tylko te, które wykazują spójność z ich własnymi oczekiwaniami. W obliczu nadwyżki wiedzy, doświadczanej jako trudna percepcyjnie i dydaktycznie wielość niespójnych danych, przyjmują oni postawę bezradności, lub podejmują próbę zredukowania jej i dopasowania do założeń modelu teoretycznego obranego na użytek prowadzonej analizy, bądź też akceptują złożoność szkolnej codzienności, dostrzegając wiele możliwości jej interpretowania.

Podsumowując, z dotychczasowych analiz wynika, że *nadmiar* rozpatrywany ze względu na rodzaj rozumienia, jest kategorią opisującą nie tylko nadmierną ilość, czy naddatek. Znaczenie to bowiem bywa nadawane wszelkim diagnozom odbiegającym od spodziewanych. I choć dokonanie redukcji nadwyżki wiedzy okazuje się strategią stosowaną najczęściej, w dalszej części pracy pokażę, iż nie jest ona strategią stosowaną wyłącznie.

2.4. *Nadmiar* ze względu na obowiązującą normatywność: nienormowane *vs* regulowane

Tego rodzaju *nadmiaru* w diagnozowaniu doświadcza zwłaszcza diagnosta prowadzący badania ilościowe. Jedną z własności tego podejścia jest normatywność, odzwierciedlana w apriorycznych założeniach o zawartości treściowej wzorca, będącego potem kryterium formułowanych osądów. W istocie są one założeniami ontologicznymi dotyczącymi natury rzeczywistości, poznawanej i ocenianej w odniesieniu do obieranego konstruktu oraz wiedzy posiadanej przez diagnostę. W świetle tegoż, diagnoza dostarcza co najwyżej prawdy lokalnej:

arbitralnie zdefiniowanej, zmierzonej, poznanej i ulokowanej na osi norm (Klus-Stańska 2013: 85). W edukacji normatywność diagnozowania ilościowego widoczna jest, między innymi, podczas egzaminów zewnętrznych, stosujących testy sprawdzające wyłącznie wiedzę i umiejętności w zakresie określonym standardami wymagań egzaminacyjnych, sprawdzane poprzez przyrównywanie uczniowskich odpowiedzi do klucza. To, co nie wpisuje się w jego zakres treściowy, nie podlega ocenie, gdyż jest traktowane jako *nadmiar* wiedzy „zbędnej". Liczne protesty przeciwko traktowaniu jako *nadmiaru* tego, co odbiega od klucza, doprowadziło ostatnio do zmian w ocenianiu prac egzaminacyjnych i przejściu z analitycznego na holistyczne, wymagające wszakże, między innymi, sprawdzania każdej pracy przez co najmniej dwóch egzaminatorów. Niestety, wysoki koszt takiego przedsięwzięcia skutecznie uniemożliwia jego realizację (Szaleniec 2015: A6).

Stosowanie w badaniach edukacyjnych wyłącznie podejścia ilościowego oraz normalizacji, oznaczałoby selektywne zaprowadzanie porządku i eliminowanie kreatywnych uczniów, których nieszablonowe podejścia są traktowane jako *nadmiar* odbiegający od klucza. Co skądinąd byłoby sprzeczne z postrzeganiem kształcenia jako procesu polimorficznego, będącego pochodną wiązania i wzajemnego przenikania się rozporządzeń urzędowych, prawno-dyrektywnych, diagnoz i ewaluacji zewnętrznych z oddolnymi inicjatywami uczniów i nauczycieli, ich imperatywami moralnymi, realizowanym diagnozowaniem powszednim czy intuicją w działaniu (Udzik 2014: 144). Choć tak rozumiana wielość eliminuje zjawisko *nadmiaru*, z wypowiedzi nauczycieli wynika, że nad zgiełk „nowej wieży Babel" przedkładają przejrzystość oraz proste i jednoznaczne wykładnie normatywne (Sadoń-Osowiecka 2014: 108), nawet za cenę: a) redukcjonizmu w badaniach; b) zagrożenia wykluczeniem z analiz kulturowego tła edukacji (Korporowicz 2013).

Podsumowując, diagnozowanie rozpatrywane ze względu na zakres, w jakim wyniki badań odbiegają od przyjętych wzorców i norm, ujawnia *nadmiar* jako kategorię obecną przede wszystkim w diagnozowaniu ilościowym, realizowanym w paradygmacie pozytywistycznym, z wykorzystaniem ankiet i pomiaru dydaktycznego, oraz dostarczającym danych, które umożliwiają formułowanie uogólnionych wniosków. Jednak wprowadzony w to diagnozowanie swoisty rygor, czyni je działaniem wyłącznie technicznym, wyobcowanym etycznie i normatywnie kontrowersyjnym (Wolińska 2014: 34), gdyż realizowanym jako selektywne zaprowadzanie porządku (Waldenfels 2009: 23). Mimo ostrzeżeń diagnostów

przed nazbyt radykalnymi próbami wyeliminowania *nadmiaru* z badań edukacyjnych, obawa przed możliwym chaosem rodzi pragnienie porządkowania i kontrolowania szkolnej codzienności, m.in. poprzez egzaminy zewnętrzne i zewnętrzną ewaluację.

3. Strategie radzenia sobie z *nadmiarem* w diagnozowaniu

Z analizy prac z zakresu diagnostyki edukacyjnej wyłania się kilka sposobów radzenia sobie przez diagnostów z *nadmiarem* w diagnozowaniu. Przedstawionymi w tym opracowaniu będą: a) walka z nim; b) świadome niedostrzeganie go; c) traktowanie go jako sytuacji zmuszającej do szukania nowych rozwiązań.

Nadmiar jako impuls do walki

Retoryka walki z nadmiarem jest wyraźnie słyszalna choćby w pracach diagnostów przywołanych w tym opracowaniu. Pojawiające się w nich sformułowania takie jak: „obrona przed", „konfrontacja", „wpadanie w pułapki", „przytłoczenie", „niszczenie" itd., są stosowane przez badaczy usytuowanych zarówno na stanowisku paradygmatu interpretatywnego, jak i pozytywistycznego (normatywnego), niewspółmiernych ontologicznie, epistemologicznie i metodologicznie (Konecki 2000: 20–22). Toczony między nimi spór, Barbara Ciżkowicz tak opisuje: „przeciwnicy stosowania rozległych badań ilościowych podkreślają ułomność makrowskaźników, wskazując brak możliwości uwzględniania w nich danych kontekstowych, które wpływają na wyniki kształcenia. Zwolennicy badań ilościowych zwracają uwagę na możliwości generalizacji i porównań, jakie stwarza ten rodzaj badań" (2013: 128).

Zamiana walki z *nadmiarem* na konstruktywny dialog nawiązywany między diagnostami z bliska a diagnostami z daleka (np. Groenwald 2013; Klus-Stańska 2013; Jaskuła, Korporowicz 2014), zdaniem Yvonn S. Lincoln i Normana K. Denzina wydaje się mało prawdopodobny w najbliższej przyszłości, bowiem trudno wyobrazić sobie, by stojący po przeciwnych stronach podziału badacze mogli i chcieli się usłyszeć, rzetelnie przeanalizować odmiennie ukierunkowane rezultaty badań i różnie konstruowane poglądy na rzeczywistość społeczną. Choć sami przyznają, że „wszelkie potyczki, konflikty i zobowiązania mają niewielkie szanse by nas gdziekolwiek zaprowadzić" (2009: 674), tym nie mniej w obrębie diagnostyki edukacyjnej bezsensowny spór między nimi wciąż jest kontynuowany. Stanisław Rainko (2011: 96) wręcz sugeruje, że skonfliktowanym (głównie

na płaszczyźnie metodologii) badaczom zależy wręcz na podtrzymywaniu go, bowiem w nim znajdują rację własnego bycia w roli naukowców-badawczy. W ostatnich latach zaostrzeniu utarczek diagnostów na płaszczyźnie metodologii, sprzyja autentyczne pragnienie odniesienia zwycięstwa w walce „o" i „przy" udziale zasobów takich jak środki na granty badawcze. Walka o nie toczy się w otoczeniu, a czasem przy udziale, kręgów politycznych i administracyjnych (Lincoln, Denzin 2009: 674). Nauczyciele natomiast, zazwyczaj uchylają się przed bezpośrednimi konfrontacjami i zamiast nich stosują wymuszoną adaptację lub strategie przetrwania (Szyling 2013: 61–64), które wszakże nie są obce również badaczom profesjonalnym.

Świadome niedostrzeganie *nadmiaru*

Jednym ze stanowisk diagnostów (zarówno z bliska, jak i z daleka) wobec problemów z *nadmiarem* jest niedostrzeganie go, możliwe dzięki przyjmowaniu postaw: „nieświadomości", „obojętności", „bierności" i „niewinności". „Nieświadomość", autentyczna bądź pozorowana, prowadzi diagnostę do zamknięcia się w obrębie już oswojonych i wręcz rutynowo stosowanych metod diagnozowania, sprzyja budowaniu niepodważalnej wiary we własne umiejętności oraz umacnia przeświadczenie o diagnostycznej nieomylności, w istocie będącej złudnym samooszukiwaniem się.

„Przydatna" dla zachowania bezproblemowości w praktykach edukacyjnych może okazać się postawa „obojętności" wobec diagnostycznego *nadmiaru*, przyjmowana przez diagnozujących z bliska nauczycieli wobec wyników egzaminów zewnętrznych czy ewaluacji zewnętrznej, zaś przez badaczy zewnętrznych – wobec diagnoz nauczycielskich. Różni się od postawy „nieświadomości" tym, że diagnosta dopuszcza myśl o *nadmiarze*, po czym natychmiast wypiera ją z świadomości, dzięki czemu nie ma potrzeby mierzenia się z nim.

„Obojętność" koresponduje z „biernością", zazwyczaj wynikającą z braku wiary w możliwość wprowadzenia zmian, przy czym poczucie bezsilności bywa tłumaczone brakiem decyzyjności w kwestiach oświatowych, podejmowanych poza diagnostą (nauczycielem czy badaczem). Powołując się na publiczne debaty nauczycieli prowadzone przez Piotra Zamojskiego na Uniwersytecie Gdańskim (2011), G. Szyling (2013: 63–64) zwraca uwagę na zgłaszane przez debatujących systemowe uwiązanie i zredukowanie do roli bezrefleksyjnych wykonawców, zobligowanych do wykazywania się osiągnięciami. Nie mogąc sprostać nierealnym wymogom, realizują je w ograniczonym zakresie, nie zawsze ujawnianym podczas badań zewnętrznych, co sprzyja budowaniu fikcyjnego wizerunku stanu edukacji.

„Poczucie niewinności" wynika z przeświadczenia nauczycieli-diagnostów z bliska o braku możliwości wpływania na konstrukcję narzędzi i przebieg badań zewnętrznych. Wyeliminowani oraz samoeliminujący się z nich, nie identyfikują się ani z nimi, ani z diagnozami, ani z ich konsekwencjami. Jeśli ponadto te dane odbiegają od osobistych lub potocznych poglądów na dany temat, w celu zniwelowania efektu kontrastu redukują je, traktując jako „obce" i zbędne (tamże: 60). To samo „poczucie niewinności" jest prezentowane w procesie redukcji takich wartości społecznych jak mądrość, prawda, odpowiedzialność; uznane za zbędne w ewaluacji zewnętrznej, są zamieniane na utylitarność, pragmatyzm i skuteczność (Korporowicz 2013: 36–37).

Charakterystyczna dla przywołanych przejawów świadomego i dobrowolnego niedostrzegania *nadmiaru* w diagnostyce jest skłonność do samousprawiedliwiania się badaczy za postępowanie, którego winnymi czynią innych: uczniów, ich rodziców, nauczycieli, zwierzchników, badaczy z odmiennej opcji metodologicznej, ewaluatorów, egzaminatorów lub nieodpowiednie programy edukacyjne. Choć ten sposób postępowania podważa zaufanie społeczne wobec diagnoz, opracowujący je czują się zwolnieni z ponoszenia odpowiedzialności za przyjmowane założenia metodologiczne i postępowanie diagnostyczne.

Nadmiar jako impuls do poszukiwania nowych sposobów poznawania i wyjaśniania rzeczywistości edukacyjnej

Z analizowanych publikacji diagnostów, prócz podejmowania walki z *nadmiarem* bądź świadomego niedostrzegania go, istnieje jeszcze inne podejście do niego; myślę tu o potraktowaniu go jako impulsu do poszukiwania nowych rozwiązań. Jednym z nich może być dialog paradygmatów (Klus-Stańska 2013: 76). Przy stworzeniu warunków sprzyjających takiemu dialogowi, ich niewspółmierność nie stoi na przeszkodzie prowadzenia badań rozumiejących, które mogą być szansą dla tego nurtu diagnostyki, uwzględniającego wielowymiarowość edukacyjnych doświadczeń. Okazją po temu są odbywające się od dwudziestu lat Konferencje Diagnostyki Edukacyjnej, podczas których dochodzi do wymiany poglądów podczas dyskusji w czasie sesji oraz paneli, bywa – dość gwałtownych. Najczęściej zderzają się podczas nich racje diagnostów prezentujących badania z bliska z argumentami diagnostów z daleka, wewnątrzszkolne diagnozy lekcyjnej codzienności z profesjonalnymi diagnozami ilościowymi. Co istotne, prezentowana podczas nich wielość podejść nie jest świadectwem diagnostycznego chaosu, lecz przejawem poznawczego porządkowania kierującego się różnymi zasadami i regułami. I właśnie te różnice okazują się cenne bardziej niż jednomyślność,

gdyż wraz z zaciekawieniem innością i gotowością do namysłu nad nią, rodzi się szansa na dialog (Klus-Stańska 2002: 22).

Inną strategią sprowadzającą się ostatecznie do ograniczenia bądź nawet wyeliminowania *nadmiaru* są „przekłady", dokonywane np. przez nauczyciela wówczas, gdy na użytek pracy z uczniami sięga po wyniki egzaminów. Ich wykorzystanie adekwatne do potrzeb, wymaga dokonania przekładu międzyparadygmatycznego, polegającego na zinterpretowaniu (czyli swoistym „przetłumaczeniu") danych ilościowych na język jakościowej praktyki dydaktyczno-wychowawczej. Warto zauważyć, że ma tu miejsce podwójne tłumaczenie, gdyż pierwotny przekład powstaje podczas ilościowej analizy arkuszy przez egzaminatorów, a kolejny – gdy po wyniki tych badań sięga nauczyciel, zanurzony w szkolną codzienność i komunikujący się w niej za pomocą języka odmiennego od języka egzaminacyjnego, a więc wymagającego przełożenia. Konferencyjne prace donoszą o przykładach takich translacji, do których dochodzi przy korzystaniu przez nauczycieli z diagnoz sformułowanych w ramach Gimnazjalnego Programu Kształcenia Kompetencji Kluczowych. Uczestniczący w tych badaniach pedagodzy deklarowali, że dla nich „istotny był zwłaszcza komunikat, jak konkretne zadania były rozwiązywane przez konkretnych uczniów, klasy w kontekście wyników uzyskiwanych w całej szkole" (Kondratek i in. 2013: 141). Jeszcze inne opracowania konferencyjne donoszą o przypadkach:

– przekładania (przez gimnazjalne zespoły ds. edukacyjnej wartości dodanej (EWD), ds. diagnoz i analiz) uzyskiwanych przez uczniów wyników egzaminów i wskaźników EWD na działania modyfikujące kształcenie (Szmigel 2013: 63),
– wykorzystania arkusza egzaminacyjnego z języka polskiego jako nauczycielskiego narzędzia diagnostycznego (Kozak 2014: 375–378).

Podobnie mogą nauczycielskimi diagnozami inspirować się profesjonalni badacze, przyjmując je za punkt wyjścia do diagnozowania z daleka. W tej kwestii B. Niemierko (2013b: 36) proponuje przenoszenie problematyki osiągnięć twórczych ucznia i kształcenia wielostronnego z poziomu indywidualnego na poziom instytucji, aczkolwiek przyznaje, że ten kierunek przepływu diagnoz wydaje się wciąż niewystarczająco efektywny.

Nadmiar – między problemem a dobrodziejstwem. Zakończenie

Nadmiar, rozumiany najogólniej jako „pewnego rodzaju mnogość rzeczy, stanów rzeczy, procesów, zdarzeń itp." (Stróżewski 2005: 473), w opracowaniu został ukazany – z jednej strony – jako problem diagnostyki edukacyjnej do-

świadczany wszakże i przez diagnostów: a) w obrębie teorii, konstruowanej w odniesieniu do zróżnicowanych ujęć paradygmatycznych i metodologicznych; b) w realizowanych praktykach, zarówno podczas diagnozowania z bliska, jak i z daleka; c) w diagnozach, które dostarczają kłopotliwej dla praktyków nadwyżki wiedzy i zamiast pomagać w pracy z uczniami, utrudniają ją. Doznawany jako to, co zewnętrzne, „obce" i nieoswojone, nieprzewidywalne, bo wymykające się znanym regułom i zasadom, uruchamia przyjmowanie wobec niego postaw: walki bądź – przeciwnie – świadomego niedostrzegania problemu, oddziałując destrukcyjnie na diagnostykę edukacyjną.

Równocześnie analiza prac odsłania także inne oblicze *nadmiaru*, wskazując na relacyjny charakter atrybutów: własny-obcy, wewnętrzny-zewnętrzny, normatywny-nieuregulowany, itd. Z tej perspektywy pojawia się on jako wyzwanie dla diagnostów, zmusza ich bowiem do rezygnacji z przyzwyczajeń i do odchodzenia od utartych schematów postępowania badawczego. Przypomina, że stosowane teorie i metody nie są dane raz na zawsze, a tym co stanowi o sensie badań – również tych prowadzonych w ramach diagnostyki edukacyjnej – jest „tworzenie własnego warsztatu badawczego, własnej teorii i metodologii za każdym razem, kiedy podejmowany jest nowy temat" (Domańska 2012: 176). Tym samym staje się *nadmiar* inspiracją do podejmowania działania odbiegającego od rutynowego, w czym pomocne może być dialogowanie adwersarzy, albo wzajemne przekładanie języków odmiennych metodologii, wyrastających z opozycyjnego usytuowania leżących u ich podstaw paradygmatów.

Przywołane w tym artykule sposoby radzenia sobie z *nadmiarem* są jednymi z wielu, natomiast istotne jest to, że niezależnie od tego, czy będzie on rozumiany jako problem czy jako inspiracja diagnostyki edukacyjnej, może stać się impulsem do poszukiwania rozwiązań trudności i otwarciem przed diagnostami nowych pól poszukiwań. Niewykluczone, iż dostarczy dowodów potwierdzających tezę Michaela Kane'a (Niemierko 2009: 147), że im większa będzie wiedza diagnostów o prowadzonych badaniach i ich skutkach, tym obszerniejsze odkrywać się będą przed nimi obszary ich niepewności i niewiedzy.

Bibliografia

Chodnicki J. 2014. *Dylematy dydaktyki matematyki czy dylematy diagnostyki?*, „Ruch Pedagogiczny", nr 2, Wyższa Szkoła Pedagogiczna ZNP, Warszawa.

Ciżkowicz B. 2013. *Obraz polskiej szkoły w badaniach 15-latków*, „Teraźniejszość – Człowiek – Edukacja", nr 4 (64), Wydawnictwo Dolnośląskiej Szkoły Wyższej, Wrocław.

Domańska E. 2012. *Historia egzystencjalna*, Wydawnictwo Naukowe PWN, Warszawa.

Garfinkel H. 2007. *Studia z etnometodologii*, tłum. A. Sulżycka, Wydawnictwo Naukowe PWN, Warszawa.

Groenwald M. 2013. *Sens spotkań diagnostyki z ewaluacją w szkole*, „Teraźniejszość – Człowiek – Edukacja", nr 4 (64), Wydawnictwo Dolnośląskiej Szkoły Wyższej, Wrocław.

Groenwald M. 2014. *Diagnozowanie – między etyką bliskiego a dalekiego zasięgu*, „Ruch Pedagogiczny", nr 2, Wyższa Szkoła Pedagogiczna ZNP, Warszawa.

Holstein J.A., Gubrium J.F. 2009. *Praktyki interpretacyjne a działanie społeczne*, tłum. K. Iwińska, [w:] *Metody badań jakościowych*, t. 1, red. N.K. Denzin, Y.S. Lincoln, Wydawnictwo Naukowe PWN, Warszawa.

Jaskuła S., Korporowicz L. 2014. *Odrębność i współgranie. Polskie relacje diagnostyki i ewaluacji edukacyjnej*, „Ruch Pedagogiczny", nr 2, Wyższa Szkoła Pedagogiczna ZNP, Warszawa.

Kincheloe J., McLaren P. 2009. *Teoria krytyczna i badania jakościowe*, tłum. P. Pluciński, [w:] *Metody badań jakościowych*, t. 1, red. N.K. Denzin, Y.S. Lincoln, Wydawnictwo Naukowe PWN, Warszawa.

Klus-Stańska D. 2002. *Konstruowanie wiedzy w szkole*, Wydawnictwo Uniwersytetu Warmińsko-Mazurskiego, Olsztyn.

Klus-Stańska D. 2013. *Myślenie poza diagnozą i ewaluacją. Pytanie o możliwość zmiany logiki tworzącej edukację*, „Teraźniejszość – Człowiek – Edukacja", nr 4 (64), Wydawnictwo Dolnośląskiej Szkoły Wyższej, Wrocław.

Kondratek B., Szaleniec H., Legutko M. 2013. *Rola diagnozy w Gimnazjalnym Programie Kształcenia Kompetencji Kluczowych*, [w:] *Polska edukacja w świetle diagnoz prowadzonych z różnych perspektyw badawczych*, red. B. Niemierko, M.K. Szmigel, Grupa Tomami, Kraków.

Konecki K. 2000. *Studia z metodologii badań jakościowych. Teoria ugruntowana*, Wydawnictwo Naukowe PWN, Warszawa.

Korporowicz L. 2013. *Społeczna etyka ewaluacji*, „Teraźniejszość – Człowiek – Edukacja", nr 4 (64), Wydawnictwo Dolnośląskiej Szkoły Wyższej, Wrocław.

Kozak B. 2014. *W jakim stopniu nowy arkusz egzaminacyjny z języka polskiego może stać się przydatnym narzędziem diagnostycznym dla nauczyciela?*, [w:] *Diagnozy edukacyjne. Dorobek i nowe zadania*, red. B. Niemierko, M.K. Szmigel, Grupa Tomami, Kraków.

Lincoln Y.S., Denzin N.K. 2009. *Ósma i dziewiąta faza. Badania jakościowe (w) przełomowej przyszłości*, tłum. K. Miciukiewicz, [w:] *Metody badań jakościowych*, t. 2, red. N.K. Denzin, Y.S. Lincoln, Wydawnictwo Naukowe PWN, Warszawa.

Lisiecka Z., Kozak W. 2014. *O potrzebie badań nad efektem zwrotnym egzaminu maturalnego z języka polskiego*, „Ruch Pedagogiczny", nr 2, Wyższa Szkoła Pedagogiczna ZNP, Warszawa.

Niemierko B. 2009. *Diagnostyka edukacyjna. Podręcznik akademicki*, Wydawnictwo Naukowe PWN, Warszawa.

Niemierko B. 2013a. *Czy diagnostyka edukacyjna i ewaluacja w edukacji zachowają odrębność?*, „Teraźniejszość – Człowiek – Edukacja", nr 4 (64), Wydawnictwo Dolnośląskiej Szkoły Wyższej, Wrocław.

Niemierko B. 2013b. *Diagnostyka edukacyjna duża i mała*, red. B. Niemierko, M.K. Szmigel, Grupa Tomami, Kraków.

Niemierko B. 2014. *Dwie dekady dojrzewania diagnostyki edukacyjnej w Polsce*, „Ruch Pedagogiczny", nr 2, Wyższa Szkoła Pedagogiczna ZNP, Warszawa.

Rainko S. 2011. *Dwa paradygmaty: esej z teorii wiedzy*, Państwowy Instytut Wydawniczy, Warszawa.

Rubacha K. 2008. *Metodologia badań nad edukacją*, Wydawnictwa Akademickie i Profesjonalne, Warszawa.

Sadoń-Osowiecka T. 2014. *Dydaktyczne dyskursy w diagnostycznych rozważaniach o przedmiotach przyrodniczych na Konferencjach Diagnostyki Edukacyjnej*, „Ruch Pedagogiczny", nr 2, Wyższa Szkoła Pedagogiczna ZNP, Warszawa.

Stróżewski W. 2005. *Istnienie i sens*, Znak, Kraków.

Szaleniec H. 2015. *Każdą maturę powinny sprawdzać dwie osoby*, „Gazeta Prawna", nr 49 (3942).

Szmigel M.K. 2013. *Gimnazjum niejedno ma imię – na podstawie wywiadów z dyrektorami*, red. B. Niemierko, M.K. Szmigel, Grupa Tomami, Kraków.

Szyling G. 2013. *Diagnoza i ewaluacja a teoretyczne horyzonty refleksji nad działaniem*, „Teraźniejszość – Człowiek – Edukacja", nr 4 (64), Wydawnictwo Dolnośląskiej Szkoły Wyższej, Wrocław.

Szyling G. 2014. *Problemy z „nadwyżką wiedzy" w diagnozie prowadzonej z bliskiej perspektywy*, „Teraźniejszość – Człowiek – Edukacja", nr 3 (67).

Udzik B. 2014. *Dyskurs edukacyjny wpisany w komunikowanie ewaluacji (wokół zagadnień pragmatyczno-lingwistycznych)*, „Ruch Pedagogiczny", nr 2, Wyższa Szkoła Pedagogiczna ZNP, Warszawa.

Waldenfels B. 2009. *Podstawowe motywy fenomenologii obcego*, tłum. J. Sidorej, Oficyna Naukowa, Warszawa.

Wolińska A. 2014. *Zdziwienie / oczekiwanie / bezczynność. Estetyka wobec nadmiaru*, Wydawnictwo Naukowe Semper, Warszawa.

Zamojski P. 2011. *Pomyśleć szkołę inaczej... Edukacja matematyczno-przyrodnicza w toku debat publicznych*, Fundacja Uniwersytetu Gdańskiego, Gdańsk.

Żytko M. 2013. *Ogólnopolskie Badania Umiejętności Trzecioklasistów (OBUT) – pedagogiczny eksperyment z „politycznymi konsekwencjami"*, „Teraźniejszość – Człowiek – Edukacja", nr 4 (64), Wydawnictwo Dolnośląskiej Szkoły Wyższej, Wrocław.

KWARTALNIK PEDAGOGICZNY
2015 NUMER 4(238)

Leszek Korporowicz
Uniwersytet Jagielloński

Ewaluacja jako komunikacja. Od animacji do alienacji

Summary

EVALUATION AS COMMUNICATION. FROM ANIMATION TO ALIENATION

The evaluation process generates not only procedures of assessment, but also negotiation of evaluation questions, criteria, indicators and, first of all, purposes. All participants in the process have their specific roles, rights and elaborate design of evaluation. Such approach to evaluation studies makes this type of research deeply involved in personal and social interaction of the participants and animates developmental resources of particular communities. That is why evaluation as an "invitation to development" is an important component of organizational and educational culture of each school. At the same time the process could be used in a very instrumental way to control and even manipulate the procedures and its results as a strategy of bureaucratic management.

Key words: evaluation process, social communication, interaction, institutional reflexivity, intercultural evaluation, participation.

red. Paulina Marchlik

Wstęp

Na horyzoncie wielu rodzajów badań społecznych, w tym także pedagogicznych, studia ewaluacyjne wykazują się jednym z najwyższych poziomów interdyscyplinarności, ale także postulowanym zastosowaniem, a nawet formą zaangażowania. To ono sprawia, że pomimo faktu, iż każdy rodzaj badania ma charakter procesualny, to badania ewaluacyjne eksponują i wykorzystują go

jako jedno z kluczowych swoich „narzędzi", które urasta wręcz do świadomie praktykowanych wartości. Choć w pragmatyce utylitarnych celów tego procesu jego efekt przysłania często świadomość rzeczywistych możliwości, jak i uwikłań jego realizacji, metodologia badań ewaluacyjnych w pełni podkreśla interakcyjny charakter czynności jej projektowania i przebiegu. Interakcyjność procesu badawczego pozwala z kolei widzieć go jako konkretny rodzaj komunikacji społecznej jaka angażuje wszystkich swoich uczestników.

Potraktowanie badań ewaluacyjnych jako procesu komunikacji społecznej umożliwia odkrycie wielu jej aspektów, które uchodzą obserwacji w tradycyjnym jej postrzeganiu zdominowanych coraz częściej przez administracyjne i technokratyczne wymogi funkcjonowania współczesnych instytucji edukacyjnych. Co istotne rozumienie procesu ewaluacyjnego zyskuje na przywołaniu elementarnych, choć równie często ignorowanych cech procesu komunikowania. Wynikają one z istoty zarówno tego, co określamy jako „inter", jak i równie ważnego segmentu pojęcia jakim jest „akcja". To pierwsze odsyła nas bowiem do tak istotnych atrybutów komunikacji jak wzajemność i wymiana, to drugie, do atrybutu podmiotowości i sprawstwa. Komunikacja nie jest więc jednokierunkowym transferem, przekazem lub transmisją. Nie jest także jednostronnym procesem percepcji, który czyni odbiorcę komunikatu biernym przedmiotem oddziaływania, wpływu, a czasami nawet manipulacji. Ewaluacja postrzegana, a co ważniejsze praktykowana jako komunikacja zakłada jej wysoce interakcyjny, a więc uspołeczniony, rozwojowy i demokratyzujący charakter.

Walory komunikacyjnego traktowania ewaluacji nie ujawniają się dopiero w fazie ujawniania, czy także dyskusji nad jej wnioskami, dokonanymi analizami, więc w fazie ustalania wartości ewaluowanych działań. Proces komunikacyjny rozpoczyna się znacznie wcześniej, już w fazie samego zamysłu nad potrzebą badań ewaluacyjnych, jej celami, elementami projektu oraz jego podmiotami. Ogromna wiedza jaką dysponujemy o różnorodnych aspektach dynamiki procesów komunikowania może w ten sposób służyć strategom i realizatorom badań ewaluacyjnych wskazując na ich efekty daleko wcześniej, ale i później niż to się na ogół uważa. To także na gruncie nauk o komunikowaniu wypracowano koncepcje „komunikacji powierzchniowej" i komunikacji głębokiej, która doskonale opisuje nie tylko poziom zaangażowania, ale rzeczywisty charakter procesu ewaluacyjnego, który może być twórczą animacją, ale także popaść w patologię wyalienowania.

Ewaluacja w kontekście globalizacji procesów komunikowania

Znane z wielu opisów cztery generacje ewaluacji już niebawem poszerzone będą o szybko rozwijający się, kolejny etap jej rozwoju, to znaczy czwartą generację badań ewaluacyjnych wkomponowaną w specyfikę globalizującego się „społeczeństwa sieci". Powstająca w ten sposób piąta generacja ewaluacji nie będzie jednak zwykłą rekontekstualizacją poprzedniej, ale istotną jej zmianą, o nowych dynamizmach rozwojowych. Warto zwrócić uwagę na kilka z nich, które mają szczególny wpływ nie tylko na konkretne procedury, ale cały proces ewaluacyjny.

A. Medializacja przestrzeni społecznych interakcji

Gwałtowny postęp w dziedzinie technologii informacyjnych „społeczeństwa sieci", istotnie poszerzył społeczną przestrzeń informacji, wygenerował konieczność ich ustawicznej wymiany, a w ślad za tym komunikacji jako codziennej i funkcjonalnie niezbędnej sfery nowego społeczeństwa wiedzy (Castells 2007; Barney 2004). Wszechobecność mediów powoduje zmianę wartości przestrzeni komunikowania jako fundamentalnego kontekstu realizacji, ale i ewaluacji większości działań społecznych. Ich realizacja nie jest możliwa bez daleko idącej medializacji wielu istotnych czynności, jak: wyszukiwanie, przetwarzanie, analiza i przekazywanie danych poprzez złożone często systemy informatyczne. Efektem ubocznym stało się jednak funkcjonalne uzależnienie tak powstałej przestrzeni od posiadania i rozwoju infrastruktury medialnej, poza którą przekaz ten nie jest możliwy. Funkcjonowanie nowej przestrzeni wymaga równie nowych kompetencji informacyjno-komunikacyjnych, których brak lub niewystarczający poziom rozwoju staje się przyczyną poważnego zepchnięcia na margines społeczności komunikacyjnej (Korporowicz 2007). Zespoły ewaluacyjne muszą w tej sytuacji nie tylko posiąść wspomniane umiejętności informacyjno-komunikacyjne, ale same stać się formą społeczności informacyjnej, umiejętnie wkomponowanej w „sieć" komunikacyjnych interakcji. Od samego początku planowania procesu ewaluacyjnego strategia badawcza stać się musi nade wszystko strategią komunikacyjną, ponieważ skonstruuje stosowny system komunikacji wraz z kardynalnymi zasadami jego zarządzania. Muszą one opanować i wykorzystać nie tylko tworzące się w szybkim tempie zbiory danych, ale także linie i kanały powiązań komunikacyjnych, które mogą demokratyzować, ale także marginalizować i wykluczać jego uczestników.

B. Hybrydalna przestrzeń ewaluacji

Społeczeństwo sieci dokonało nie tylko istotnego poszerzenia, ale przede wszystkim przekroczenia dotychczasowych granic przestrzeni komunikacyjnej wyznaczanych „klasycznymi" środkami przekazu, włączając w nie radio i telewizję. Ta swoista de-terytorializacja przestrzeni informacyjnej, komunikacyjnej i przestrzeni wiedzy wytwarza nowy rodzaj kompetencji, która wiążę się także z umiejętnością łączenia wielu mediów i osiągania efektu synergetycznego w „przestrzeni hybrydalnej", a więc takiej, która łączy świat realny i wirtualny. Procesy te bezpośrednio rzutują na charakter projektowanych badań ewaluacyjnych, które stają się informacyjnie „hybrydalne", przyjmują realność światów wirtualnych, a ewaluatorów uczą przemieszczania się między nimi. Ze względu na fakt realizacji wielu projektów i działań społecznych w przestrzeni wirtualnej, podobnemu rozszerzeniu uległa przestrzeń ewaluacji. Naturalnym rezultatem nakładania się przestrzeni informacyjnych i komunikacyjnych jest ich inter-akcja (Jaskuła 2011a, 2011b). Wiedza ewaluacyjna powstaje wówczas w środowisku nie tylko porównań, ale przede wszystkim wymiany i rekonfiguracji jej treści. Bardzo to różni perspektywę wielodyscyplinarną od międzydyscyplinarnej, tworząc bardziej „iloczyn" niż „sumę" wiedzy. Ewaluatorzy, którzy nie rozwijają umiejętności interakcyjnych pozostają zredukowani do obszaru jednej dziedziny lub zasobów jednego rodzaju przestrzeni (Jaskuła 2012). Ogranicza to konieczną innowacyjność badań ewaluacyjnych i umiejętność działania w ramach różnych jej strategii, podejmowania takich jej wyzwań, które wymagają przemieszczania się z jednej przestrzeni do innej oraz łączenia odmiennych procedur organizacji i projektowania badań. Współczesny ewaluator staje się sam w sobie medium spotykania różnych wrażliwości, wyobraźni ale i kompetencji społecznych.

C. Wzrastająca ranga relacji między-kulturowych

Spośród wielu interakcji, które wykazują cechy nakładania się, transgresji i wymiany coraz ważniejsza staje się przestrzeń relacji międzykulturowych (Korporowicz 2010). Stwarza ona możliwość innego wartościowania i rozumienia wiedzy powstałej w kontekście odmiennych tradycji jej powstawania. Jest to sytuacja tworzenia nowego typu ewaluacji międzykulturowej, która moderuje i negocjuje każdy z respektowanych systemów wartości. Gwałtowny rozwój międzynarodowych programów działalności w każdej niemal sferze życia społecznego i gospodarczego wymaga wypracowania czegoś więcej niż ujednoliconych standardów wartościowania lub też narzucenia systemu dominującego.

Ewaluatorzy pozbawieni kompetencji międzykulturowych nie mogą pracować w środowiskach wielokulturowych, organizacjach transnarodowych oraz międzynarodowych systemach wiedzy. W kontekście współczesnych procesów globalizacyjnych, które generują procesy kulturowego różnicowania, jest to ważna kompetencja zawodowa i społeczna.

D. Obligatoryjny charakter umiejętności technokratycznych

Współczesne społeczeństwo wiedzy w sposób narastający modeluje ją w sposób wymagany poprzez postęp cywilizacji technicznej. Wiedza ta daleka jest od wartości, jaką przypisywano mądrości i zaletom ludzkiego charakteru. W sposób konieczny musi jej towarzyszyć umiejętność wykorzystania coraz bardziej zaawansowanych technologicznie systemów technologii informacyjnej. Ich stosowanie konsekwentnie uzależnia od tej właśnie umiejętności bezpośrednio rzutując na możliwość użycia wspomagających procedur badawczych. Procedury badań ewaluacyjnych, jak i cały proces ewaluacji głęboko uzależnia się od technologicznych komponentów każdej jego fazy, szczególnie etapu zbierania, przetwarzania, ale i komunikowania wyników ewaluacji. Umiejętność ta przenosi się ze sfery nauk technicznych, w nauki przyrodnicze, społeczne, a nawet humanistyczne i niezależnie od jej oceny różnicuje szanse na skuteczne uczestnictwo w społeczeństwie wiedzy.

Sens ewaluacyjnych partycypacji

W istocie różne mogą być formy społecznych partycypacji. Uczestnictwo jako zasada współudziału nie gwarantuje właściwie niczego. Uczestniczyć można w procedurach korupcji, dewastacji, przeróżnych form barbarzyństwa, ludobójstwa i demagogii. Uczestnictwo tak praktykowane daje wiele korzyści, zapewnia różnorodne atrakcje, a czasami gwarantuje szybkie i spektakularne kariery. Zawsze powstaje jednak pytanie, czy taka forma partycypacji jest sposobem samorealizacji i opiera się na faktycznym poszanowaniu podmiotowości konkretnych osób i wspólnot? Współczesne systemy zarządzania coraz częściej zmuszają nas do stawiania takich pytań, aby uniknąć zagrożenia pozornym stawaniem się sprawnymi „trybikami w maszynie", a wszystkie działania poddawane procedurom wartościowania uchronić przed realnym zagrożeniem alienacji.

Komunikacyjny aspekt procesu ewaluacyjnego zostaje jednak poważnie zredukowany w sytuacji zewnętrznego ustalenia nie tylko celów, ale zasadniczego przedmiotu, obszarów, pytań i kryteriów ewaluacyjnych, a więc zasadniczych elementów projektu ewaluacyjnego. Dzieje się tak poprzez systematyczne upodobnianie w realiach polskich praktyk zarządczych strategii ewaluacyjnej do komponentów i strategii badań akredytacyjnych, ale także poprzez swoistą desocjalizację jej zastosowań. Interakcja społecznych podmiotów procesu ewaluacyjnego stanowi o istocie jego tożsamości, której nie zastępują inne formy diagnostyki społecznej lub edukacyjnej. W jeszcze większym stopniu nie mogą jej zastąpić różnorodne formy monitoringu, kontrolingu czy nadzoru.

Nie bez przyczyny odkrywanie i przywracanie sensu społecznych partycypacji, jakie miały charakteryzować polskie procesy demokratyzacji przemian społecznych realizowanych od początku lat 90. ubiegłego wieku, nakierowane były na obronę, wspieranie i rozwój realnej podmiotowości zaangażowanych w nie ludzi, niezależnie od stopnia ich formalnego zorganizowania. W praktyce oznaczać ma to taką formę ich udziału, która traktuje ich nie jako „petentów", a więc nie jako „przedmioty", ale podmioty konkretnych rozwiązań w sposobie planowania, projektowania i realizacji procesu ewaluacji. Tylko wówczas partycypacja oznacza uczenie się i odkrywanie, wzajemność i wymianę, przekazywanie i dzielenie się doświadczeniem każdego uczestnika jako wartego zauważenia i respektu, wartego uważnego pochylenia się i wsłuchania w jego doświadczenie i intencje. Kierunek tak pojętej partycypacji wyznacza przeciwieństwo społecznego uprzedmiotowienia, które staje się niestety cechą wielu strategii „zarządzania poprzez parametry", a nie poprzez wartości, która wkrada się w inwazję rozporządzeń, a nie w przewodzenie świadomie kierowaną zmianą. Prosta obserwacja sposobu zarządzania współczesnym systemem szkolnictwa wyższego w postaci niespotykanego do tej pory technokratyzmu Krajowych Ram Kwalifikacji, czy też coraz powszechniejsza zamiana strategii uspołecznionych ewaluacji na gotowe projekty akredytacji i audytu w wszystkich praktycznie dziedzinach życia społecznego pokazuje skalę zagrożenia. Zagłębieni w rozpoznawanie wymogów pozyskiwania i funkcjonowania grantów badawczych i sterowani koniecznością pozyskiwania niezbędnych środków instytucjonalnego „być albo nie być", tracimy krytyczny zmysł dostrzegania tego, co niszczy podstawowe przesłanie realnej demokratyzacji i przestajemy wskazywać na konieczność pytania o wartości istotnie przeżywane, projektowane i respektowane. Wiele fundacji, stowarzyszeń, ale i uczelni z pokorą przyjmuje wymuszenia proceduralne, a procedury wartościowania opiera na znikomym uspołecznieniu najistotniejszych jej kryteriów.

Pojawienie się krytycznej refleksji demaskującej ukryte praktyki o anty-dialogicznym, i hipertechnokratycznym charakterze staje się deską ratunku dla obrony podmiotowych aspektów partycypacji społecznych jako zapomnianego komponentu zmian demokratycznych społeczeństwa i państwa polskiego, które przyjmuje się jako dokonane i nie weryfikuje w realnej praktyce działań ewaluacyjnych.

Edukacja to nie przystosowanie, ewaluacja to nie interwencja

To nie kryzys ukraiński uświadamia współczesnym Europejczykom, iż nie ma wartości danych raz na zawsze, obecnych w sposób oczywisty, reprodukowanych samoczynnie, a więc takich jakie wziąć można w nawias i powrócić do rutyny codzienności. W polskiej rzeczywistości edukacyjnej, wszystkich zresztą szczebli pytania o jej udział w tworzeniu wolnego i demokratycznego społeczeństwa zastępuje gonitwa za coraz bardziej wyalienowanymi systemami zapewniania jakości. Kształcenie na potrzeby rynku stało się niemal celem strategii edukacyjnych, i bardzo trudno wyjaśnić, że procesy wartościowania oraz refleksji instytucjonalnej szkoły nie mogą być komponentem techniki nadzoru. Niestety równie trudno przekonać, że właściwie pojęta ewaluacja to w istocie pytanie, które wynika z pytania o podmiotowość konkretnej społeczności, o jej potencjał rozwojowy, o godność w dochodzeniu do własnej drogi rozwoju, a nie przyśpieszone i ukryte techniki kontroli.

Warto więc zauważyć, że edukacja w służbie społecznych partycypacji to nie trening współczesnych sposobów zarządzania. Nie jest to także zwykły przekaz wiedzy, a nawet umiejętności. Po wielu latach niepowodzeń, przynajmniej teoretycznie, coraz więcej środowisk zaczyna postrzegać edukację szerzej, uzupełnioną o trzeci jej ważki element, jakim jest kształtowanie postaw, a więc również o systemy wartości. Daje to szansę na widzenie w edukacji przynajmniej perspektywy na dotarcie do osobowości, wrażliwości, wyobraźni i odpowiedzialności jej uczestników, to znaczy nie tyle „klientów", odbiorców, a nawet „wychowanków", ile interaktywnie traktowanych osób i wspólnot. To poprzez wartości możemy odkryć ich podmiotowość, która ujawnia najistotniejszy sens i niejako busolę uczestnictwa. To podmiotowość staje się uzasadnieniem jej promocji, ale też celem przyświecającym wysiłkom wdrażającym demokratyczne formy społecznych partycypacji. Jeśli więc tracimy często z oczu busolę podmiotowości, działania społeczne zamieniają się w poligon stosunków władzy,

giełdę technik sterowania i strategii przystosowania. Polityka publiczna pozbawiona tak pojętej edukacji i podobnie rozumianej ewaluacji, traci swoją misję i zamienia się w mało skuteczną dystrybucję niedoborów, metodykę podziałów społecznych oraz sztukę usankcjonowanych marginalizacji.

Edukacja publiczna ma odwrócić opisane powyżej formy społecznych deprywacji. Ona stwarza z przestrzeni i polityki partycypacji strefę rozwoju, stawania się, budowania tożsamości jako godności, dopełnia, a nie poniża człowieka zarówno wtedy, gdy jest on przedszkolakiem, uczniem, nauczycielem, obywatelem, a nawet petentem. Takie myślenie o edukacji, w której staje się ona przestrzenią odkrywania ludzkiej podmiotowości sięga polskiego Renesansu, zostaje ubogacone oświeconym kształceniem uczciwego obywatela, gdy nadchodzą już czasy narodowego upokorzenia, po czym ponownie odradza się, gdy powracamy do działania i myślenia opartego o atrybut odzyskanej wolności. Edukacja i partycypacja publiczna, podobnie jak wyrastająca z tych samych korzeni świadomość obywatelska, stanowią w ten sposób podstawowe przesłanie w budowaniu komunikacyjnych strategii procesu ewaluacyjnego. Wówczas obronić się można przed eurokratycznymi zapędami administracyjnych technik sterowania, myśleniem według parametrów i rytualnych procedur mniej lub bardziej ukrytej kontroli, a w konsekwencji ewaluacji jako interwencji, niemal manipulacji. To współczesny język zarządzania opisuje ewaluację właśnie jako „interwencję" w procedury zarządcze. Takie myślenie o ewaluacji traktuje ją jako „instrument", stosowany zresztą nie jako rozwojowy komponent procesu zmian instytucjonalnych, kulturowych i mentalnych, ale działanie o charakterze zewnętrznego nacisku.

Warto przy tym pamiętać, że tak jak działania edukacyjne, tak i ewaluacyjne realizują się poprzez konkretne formy komunikacji, co więcej nie są one jedynie ich komponentem, ale same w sobie tworzą ewaluacyjną rzeczywistość, współokreślają jej charakter i dookreślają sposób zastosowania.

Ewaluacja jako komunikacja międzykulturowa

Procesy globalizacyjne powodują narastanie liczby międzynarodowych badań ewaluacyjnych realizowanych w odmiennych kontekstach kulturowych. Jest to konsekwencja równie intensywnego rozwoju programów międzynarodowych we wszystkich obszarach życia gospodarczego, społecznego i kulturowego. Ewaluacja międzynarodowa i międzykulturowa jest także efektem narastającego

umiędzynarodowienia i współpracy wielu instytutów badawczych, doradczych i szkoleniowych realizujących projekty ewaluacyjne na zapotrzebowanie globalizujących się działań we wszystkich sferach życia. Bardzo często jest ona inicjowana przez instytucje i urzędy międzynarodowe, które doskonalić chcą swoje funkcje technokratyczne, kontrolne i korporacyjne. Wymaga wszak ona nowych kompetencji ze strony ewaluatorów, którzy muszą być świadomi ogromnych różnic w systemach wartości, które stanowią dla nich obszar odniesienia w procesie konceptualizacji badań.

Istniejące programy ewaluacji międzynarodowych nie zawsze jednak uwzględniają istotne różnice w systemach wartości i wzorów kultury organizacyjnej ewaluowanych środowisk, unifikując przyjmowane kryteria ewaluacji wyznaczane przez dominujące ośrodki ewaluacyjnych decyzji, podporządkowanych ośrodkom decyzji menedżerskich i politycznych. Istniejący stan rzeczy w tak unifikowanych projektach nie uwzględnia faktu, iż procesy globalizacyjne, ale także umiędzynarodowienie działania wielu instytucji prowadzą nie tylko do unifikacji, ale także dywersyfikacji społecznych kontekstów ewaluowanych działań (Rekłajtis i in. 2010). Sytuacja ta powoduje konieczny wzrost zainteresowania problemem adekwatności społecznej, środowiskowej i kulturowej wielu badań ewaluacyjnych – jeśli nie mają być one przejawem instytucjonalnej dominacji, a często i uprzedmiotowienia badań ewaluacyjnych. Wrażliwość i kompetencja międzykulturowa ewaluatorów staje się w tych warunkach coraz bardziej potrzebną umiejętnością.

Dominacja poprzez unifikację kryteriów ewaluacji przenika w sferę jej projektowania, metodologii, a następnie wnioskowania. Nie bez racji mówi się o swoistym „ewaluacyjnym imperializmie", który przejawia się nie tylko w narzucaniu celów i przedmiotu ewaluacji, ale także, metod, procedur a nawet mentalności ewaluatorów traktowanych jako „funkcjonariuszy systemu". Bez uwzględnienia różnorodności kontekstualnej ewaluowanych programów, a następnie uspołecznienia procesu projektowania i adaptacji kulturowej kryteriów ewaluacyjnych ewaluacja międzynarodowa zagrożona jest społecznym wyalienowaniem lub wkomponowaniem w mechanizmy reprodukowania istniejących stosunków władzy, często dominacji o ukrytym metodologicznie i społecznie charakterze.

Istnieje kilka ważnych przyczyn, które wymuszają konieczność zdobywania nowych kompetencji społecznych, w tym także międzykulturowych, we współczesnych procesach ewaluacyjnych o nasilającej się skali umiędzynarodowienia.

Po pierwsze, jest to **mobilność kultur**.

Przekroczenie granic geograficznych, administracyjnych lub etnicznych wytworów kultury symbolicznej, a więc także istotnych dla ewaluacji systemów wartości ewaluowanych podmiotów. Konwencjonalny związek miejsca, ludzi i ich wytworów zostaje uwolniony i może być skonfigurowany w innych miejscach, ale nade wszystko przemieszczony i zmieniany. Ewaluator lub zespół ewaluacyjny, musi wykazać się szczególnie daleko posuniętą kompetencją, która pozwala mu widzieć odmienne znaczenia, ale i wartości przemieszczanych elementów dorobku kulturowego.

Po drugie, jest to **deterytorializacja przestrzeni komunikowania**.

Jest ona bezpośrednią konsekwencją mobilności kultur i polega na oderwaniu od wspomnianego już miejsca jej wytworów. Deterytorializacja bezpośrednio powoduje nasilenie zróżnicowania przestrzeni kulturowych, ich wieloskładnikowość zarówno ze względu na ich treści, ale także ze względu na ogarniające ich procesy. Przestrzeń taka przekracza wiele barier, co w istotny sposób wpływa na zasady tworzenia się wspólnot kulturowych, nastawionych znacznie bardziej na ich treści, niż ich terytorialną lokalizację. Możemy powiedzieć, iż „oś krystalizacji" wartości kulturowych w dobie ich deterytorializacji ma charakter coraz bardziej abstrakcyjny, wybrany i refleksyjny, w mniejszym stopniu rutynowy, konwencjonalny i sytuacyjny. Ewaluacja międzykulturowa podkreśla znaczenie nowej przestrzeni komunikacyjnej, analizuje rolę nowych kompetencji i dyspozycji do rozpoznawania i konfigurowania wartości w nowych obszarach interakcji o własnych regułach konsolidacji.

Po trzecie, są to **konwergencje i interakcje**.

Nakładanie się na siebie intensywnie „poszerzających się" przestrzeni musi prowadzić do wspominanych już inter-akcji. Dzieje się to zarówno w przestrzeni o silnym korelacie środowiskowym, o własnym etosie i tożsamości miejsca, jak i przestrzeni symbolicznej, w tym przestrzeni wirtualnej, ciągle trudnej do opisania, pomiaru i badania. Nikt dzisiaj nie jest w stanie określić, na czym dokładnie polegać będą kompetencje społeczne w wirtualnej przestrzeni interakcji, ewaluacji oraz tworzenia nowych strategii badawczych. Interakcje przestrzeni i wartości stanowią względnie obiektywny czynnik charakteryzujący wyzwania ewaluacji międzykulturowej. Umiejętność nie tylko funkcjonowania, ale i skutecznego rozwoju w dynamice owych interakcji stanowić będzie o kluczowych

kompetencjach najbliższych nam, a prawdopodobnie jeszcze w większym stopniu, przyszłych strategii ewaluacji.

Konwergencje i interakcje różnego typu przestrzeni w jeszcze większym stopniu niż opisane powyżej mobilności i deterytorializacje stymulują potrzebę ujawnienia odmiennych konfiguracji wartości, kompetencji i komunikacji międzykulturowej ewaluatorów, ale także swoisty rodzaj kulturowej odpowiedzialności badań ewaluacyjnych.

Po czwarte, jest to **ożywienie znaczenia „małych ojczyzn"**.

Dywersyfikujące działanie procesów globalizacyjnych prowadzi do pogłębienia, a nawet rewitalizacji tożsamości konkretnych grup oraz ich wartości rdzennych, potrzeby podmiotowości interpretowanej w kategoriach nieredukowalnych praw kulturowych i praw człowieka. Ich konceptualizacja sięga w warunkach europejskich aż do wielu XV, w którym na niezbywalne, podmiotowe prawa narodów, ale także najmniejszych nawet grup etnicznych wskazał wybitny polski myśliciel, rektor Akademii Krakowskiej, Paweł Włodkowic. Potencjał ten ujawnia się współcześnie w postaci ożywienia atrakcyjności i aksjologii nie tylko wielkich, ale także małych wspólnot kulturowych traktowanych w kategoriach małych ojczyzn, które dają poczucie zakorzenienia, ale także wzmocnienia i dowartościowania ludzi zagubionych i przytłoczonych instrumentalnością współczesnej cywilizacji. Wyrastające z tego źródła wartości nie mogą być ignorowane w projektach zestandaryzowanych koncepcji ewaluacji, realizowanej w perspektywie międzynarodowej i międzyśrodowiskowej. Perspektywa ta, w odróżnieniu od transkulturowej, wrażliwa jest na rozumienie wartości zarówno dziedziczonych, ale i nowych, a jedynie konstruowanych dla konkretnych społeczności, które postrzegane są jako ważkie środowisko działania, a nawet rodzaj „małej ojczyzny" ewaluowanej społeczności. W tym sensie współczesne badania ewaluacyjne nie muszą mieć funkcji wyłącznie unifikujących, ale wręcz przeciwnie, uwrażliwiających na kontekstualny i podmiotowy wymiar badanych wartości.

Ewaluacja w perspektywie socjologii kulturowej

Socjologia kultury, a szczególnie jej wersja w postaci socjologii kulturowej eksponuje znaczenie i wielorakie funkcje kultury symbolicznej (Kłoskowska 1985, 2005). Przywołanie tej perspektywy najlepiej charakteryzuje istotę opierającej

się technokratycznej redukcji i nie sprowadzonej do poziomu informacji, wielopostaciowej treściowo wiedzy, jaką stać się może dojrzała i odpowiedzialna społecznie ewaluacja. Jej wartość polega bowiem nie tylko na tworzeniu instrumentarium praktycznych działań – o utylitarnym i dającym się wymierzyć efekcie, ale przede wszystkim na odkrywaniu ich szeroko rozumianego sensu wyznaczonego przyjętymi celami, intencjami oraz systemem motywacji. Społeczeństwo wiedzy *jest typem organizacji społecznej, w której tworzenie i wykorzystanie wiedzy staje się strategicznym czynnikiem jego funkcjonowania i rozwoju, a sama wiedza kluczowym, dynamicznie rosnącym zasobem określającym warunki życia, system wartości oraz sposoby uczestnictwa w kulturze symbolicznej i życiu społecznym*. Badania ewaluacyjne, traktowane jako działania społeczne i jako istotny komponent współczesnego społeczeństwa wiedzy, mieszczą się w pełni w przytoczonej powyżej definicji. Warto przy tym zauważyć, iż rozumienie pojęcia społeczeństwa wiedzy, samej wiedzy, ale i ewaluacji postrzeganej jako rodzaj „zasobu społecznego", odwołuje się zarówno do jego pragmatyczno-funkcjonalnego, jak i do aksjologicznego i logotwórczego wymiaru. Ten aspekt procesów ewaluacyjnych ginie niestety w technokratycznych koncepcjach ewaluacji, pojmowanej jako narzędzie parametrycznego pomiaru skuteczności programów zarządzania. W tak okrojonych projektach ewaluacji jest ona bardziej techniką menedżerskich „interwencji" niż strategią animowania zintegrowanego rozwoju.

Ze względu na rozliczne konsekwencje uczestniczenia w społeczeństwie wiedzy, równie istotne są konsekwencje procesów wykluczenia lub choćby marginalizacji, łącznie z marginalizacją i wykluczeniem z procesów ewaluacji. Wykluczenie ewaluacyjne staje się niestety coraz częściej spotykanym, ubocznym skutkiem desocjalizacji praktyk ewaluacyjnych zamienionych, jak wspomniano o tym wcześniej, na procedury akredytacji, audytu lub nadzoru. Wykluczenie ewaluacyjne dotyczy wówczas nie tylko możliwości wpływania na wybór strategii wartościujących konkretne działania oraz techniki pozyskiwania wiedzy, ale także jest to wykluczenie samej możliwości wchodzenia w ewaluacyjne interakcje, a więc dzielenia się i wymiany doświadczeń, wyrażania własnego systemu wartości. Marginalizacja w procesie ewaluacyjnym dokonuje się poprzez wykluczenie komunikacyjne, łącznie z zepchnięciem do pozycji ewaluacyjnego audytorium, ale także poprzez brak umiejętności odczytywania, interpretacji i rozumienia wartości wyrosłych na współczesnych zasobach wiedzy oraz procesach jej tworzenia. W praktyce oznacza to nie tylko odsunięcie od czynności projektowania ewaluacji, ale także szerzej pojęty brak kompetencji do krytycznej oceny zbieranych informacji, ich przydatności oraz widzenia ich realnego

znaczenia, po drugie, brak umiejętności wyznaczania celu procedurom zdobywania i gromadzenia wiedzy i po trzecie, brak umiejętności widzenia jej roli w wartościowaniu konkretnych przedsięwzięć i ich rezultatów. Wykluczenie ewaluacyjne widziane w perspektywie socjologii kulturowej oznacza poddanie się pełnej funkcjonalizacji ról poddanych instrumentalnym wymogom konieczności technokratycznych, biurokratycznych i organizacyjnych.

Socjologia kulturowa wyróżnia cztery kategorie, poprzez które pozwala dostrzec wykluczenie ewaluacyjne i dokonać analizy zasadniczego dla człowieka problemu kultury (Korporowicz 2011), w tym kultury ewaluacyjnej.

1) Procesy odczytywania, tworzenia i przekazu znaczeń. Są one podstawą zaistnienia i rozwoju każdej kultury symbolicznej. To nie systemy zachowań tworzą właściwie rozumianą kulturę. Behawioralne rozumienie kultury gubi elementarną prawdę o człowieku, jaka zawiera się w umiejętności nadawania znaczeń ludzkim zachowaniom i tylko poprzez ten fakt możemy zrozumieć istotę jego człowieczeństwa, ale także wiele jego atrybutów, jakim jest tworzenie i wykorzystanie wiedzy. Bardzo często zapominamy, że wiedza ma dla człowieka określone „znaczenie", które konstytuuje wszystkie jego działania i pozwala porozumiewać się z innymi ludźmi. Wykluczenie ewaluacyjne, z tego punktu widzenia, oznacza opisaną powyżej nieumiejętność postrzegania i uczestniczenia w procesach określania informacji znaczących, tworzonej z nich wiedzy i definiowania ich praktycznego znaczenia.

2) Procesy tworzenia i doświadczania wartości. Stanowią one drugi, fundamentalny wymiar kultury, który nadaje sferze znaczeń fundamentalny wymiar sensu, nadaje im wagę lub też do nich dystansuje. Sfera ta określa faktyczną rolę zachowań, ale i symboli, umożliwia zaistnienie kryteriów ich hierarchizacji i wyboru, różnicowanie i ocenę motywów, celów i sposobów działania. Wykluczenie ewaluacyjne oznacza w tym przypadku – utratę umiejętności wartościowania znaczeń, w praktyce utratę umiejętności analizy funkcjonalnej wszystkich elementów kultury symbolicznej w realnej sferze życia społecznego, szczególnie w działaniach organizacji oraz innych zinstytucjonalizowanych formach aktywności.

3) Wzory interakcji społecznej. Wzory interakcji wyznaczają sposoby wzajemnych oddziaływań osób i stanowią podstawę kultury społecznej grupy (Kłoskowska 1981: 111, 126, 199, 239, 255, 331, 427, 429). Wzorce te określają w typowych sytuacjach życia zbiorowego pakiet zachowań adekwatnych, dopuszczalnych, wzbronionych lub koniecznych, określają przebieg ich realizacji i z różną siłą łączą się z kulturą symboliczną grupy. Ich wydzielenie jako odręb-

nej sfery zjawisk może mieć charakter jedynie aspektowo-analityczny, jest jednak z różnych powodów konieczne. Te same „wzory interakcyjne" powodowane być mogą różnymi motywami, odmiennym służyć celom, inne wyrażać treści. Zmierzający do obiektywizacji, nastawiony na obserwowalny wymiar zachowań opis pominąć może to, co jest w nich najistotniejsza, co leży u ich podstaw (Korporowicz 1983, 1996).

4) Dynamizmy i potencjały rozwojowe kultury. Gdy mówimy o kulturze jako zasobie i dynamizmie wartości, jako drodze ku świadomym i opartym na wiedzy wyborom, nie jest to możliwe bez właściwego im, swoistego „wektora sensu", a więc **teleologicznego** i **powinnościowego porządku**. „Wektory" te proponowałbym określić używanym tu terminem logotwórczych dynamizmów kultury, a więc tymi, które tworzą jej logos, które nadają naszym działaniom konkretny kierunek i wyznaczają ich cele.

Stwierdzenie ich istnienia nie może jednak konstruować prostej, zerojedynkowej analityki sensu, nie może również ograniczać się do konstatacji ich wielości Z rożnych względów, o których trzeba będzie powiedzieć możliwie wyraźnie i precyzyjnie, ujawnienia wymaga ich jakościowo odmienny i zróżnicowany charakter. Analiza logotwórczych dynamizmów kultury nie może być w swej istocie opisem jednorazowego aktu nadania, ale raczej nadawania sensu konkretnym przeżyciom i działaniom, wyjaśnieniem procesu, a nie stanu, i to dalekiego od liniowego, ewolucyjnego przebiegu.

Viktor Frankl stwierdził, że w kontekście tworzonej przez niego logoterapii „[...] logos znaczy tyle co sens. Byt ludzki rzeczywiście zawsze wykracza poza siebie, zawsze wskazuje pewien sens. W tym sensie człowiekowi w jego bytowaniu nie chodzi o przyjemność..., ale też nie o samourzeczywistnienie. Chodzi mu raczej o wypełnienie sensu. W logoterapii mówimy tu o «woli sensu»" (Frankl 1978: 80).

Podobnie też jak nie możemy myśleć nie zakładając jakiegoś wyobrażenia przestrzeni i czasu, „[...] tak też byt ludzki jest zawsze bytem skierowanym ku sensowi, choćby go prawie nie znał. Mamy tu coś takiego, jak przedwiedza o sensie, a poczucie sensu leży u podstaw «woli sensu» [...] nawet samobójca wierzy w sens, jeśli nie życia, dalszego trwania, to przecież wierzy w sens umierania [...]" (Frankl 1978: 80–81).

Jeżeli więc człowiek „wypełnia sens", skierowany jest ku czemuś lub też posiada „wolę sensu", to czynić to może na różne, tyleż wyzwalające, co zniewalające sposoby. Kategoria sensu, podobnie jak szersza kategoria logosu – rozpatrywana być musi w perspektywie rozwoju, aby zarówno omawianej wcześniej kategorii znaczeń, wartości i wzorców interakcji, nadać realny, indywidualny

i społeczny wymiar, jaki staje się zadaniem uczestników społeczeństwa wiedzy. Rekonstrukcja tego procesu – to ważne zadanie ewaluacji ze współczynnikiem humanistycznym, ewaluacji, która zawsze osadzona być musi w konkretnym świecie kultury, i która nie jest poza nią w ogóle możliwa.

Wykluczenie ewaluacyjne przejawia się, w tym przypadku, w zamknięciu postrzegania potencjałów rozwojowych przez uczestników ewaluacji i na preferowaniu w praktyce całego procesu komunikacyjnego schematycznych i niereflekcyjnych automatyzmów wartościowania. Odzwierciedlają one procesy uprzedmiotowienia, a czasami destrukcji w sytuacjach wymagających działań twórczych i transgresyjnych. Wykluczenie ewaluacyjne analizowane poprzez redukcję logotwórczych dynamizmów kultury stanowić będzie coraz bardziej konieczny komponent nie tylko socjologicznych, ale także antropologicznych i etycznych badań nad procesem ewaluacyjnym w warunkach społeczeństwa współczesnego.

Zakończenie

Współczesne wizje ewaluacji dzielą się w istocie na dwa niewspółmierne nurty, które posługują się odmiennymi językami, określają istotę badań ewaluacyjnych w odmienny, niemal sprzeczny sposób, a sam proces ewaluacyjny kierują ku innym celom. Z jednej strony, jest to tendencja do ujmowania ewaluacji jako technokratycznego instrumentu, który podporządkowany jest równie technokratycznym celom optymalizacji decyzji poprzez doskonalenie, ale czasem niestety manipulowanie narzędziami iluzorycznie pojmowanego pomiaru. Z drugiej strony, jest to zaawansowana antropologizacja wizerunku działań, w obrębie procesu ewaluacyjnego – eksponująca jego zamierzone i nie zamierzone rezultaty społeczne, twórcze lub destrukcyjne, zawsze jednak analizowane w kontekście ich wielorakich wartości. Między tymi dwoma tendencjami istnieje strefa, która umożliwia ujawnienie komunikacyjnego charakteru procesu jako wielorakiej interakcji między wszystkimi jego uczestnikami. Realne czynności planowania, realizacji oraz upowszechnienia wyników, w tym długi okres ustalania jego celów pozwala lepiej zrozumieć rzeczywisty sens przedsięwzięcia, który ukazuje zarówno jego twórcze, animacyjne, jak i przedmiotowe, czasem zniewalające wręcz rezultaty. Interakcyjne rozumienie badań ewaluacyjnych pozwala wszak zauważyć wszystkie możliwości, jakie stoją przed teorią, ale i praktyką procesu ewaluacyjnego, zarówno w odniesieniu do szerszych, jak

i węższych pól jego aplikacji, łącznie z pracą małych społeczności klasowych, a nawet indywidualnej pracy konkretnych osób. We wszystkich tych przypadkach podstawowy cel ewaluacji jako świadomej analizy wartości nie może być zredukowany do technik zarządzania i zapominać o swojej własnej, równie świadomej meta-wartości. To ona zdecyduje, czy chcemy i w jakim kierunku powinniśmy rozwijać studia ewaluacyjne oraz jakiej praktyce je powierzać.

Zarówno społeczeństwo informacyjne, społeczeństwo sieci lub analizowane właśnie społeczeństwo wiedzy zawiera w sobie immanentne sprzeczności (Giddens 2007; Beck i in. 2009; Sztompka 2005). Sprawiają one, iż to co poszerza (globalizacja) jednocześnie zawęża (społeczeństwo wewnątrz sieci), to co otwiera (społeczeństwo otwarte) jednocześnie zamyka (uzależnienia). Z tych to powodów wszyscy, którzy skutecznie pozyskali umiejętności i doświadczenie (kompetencje) funkcjonowania w systemie społeczeństwa wiedzy, stanowią naturalną konkurencję i marginalizują wszystkich, którzy kompetencji tych nie posiadają, spychają ich w obszar wykluczenia, daleko od procesów decyzyjnych, możliwości rozwoju zawodowego i partycypacji społecznych, eliminują ich także z procesu interakcyjnego w procesach ewaluacyjnych (Brislin i in. 1994; Lusting i in. 2006). Ci, którzy w procesie tym „coś znaczą", i potrafią komunikować swoje interesy, potrzeby i wartości, lub którzy potrafią okazać swoją władzę, przenoszą swoje działania na tych, których proces ewaluacyjny nie eksponuje (Korporowicz 2009). Ta swoista dwoistość zaawansowanych technologicznie społeczeństw „ery konceptualnej" zawiera mało demokratyczne i drastycznie ograniczające zasady, które eliminują wszystkich, jeśli tylko nie chcą lub nie zdołali się do nich dostosować.

Bibliografia

Barney D. 2007. *Społeczeństwo sieci*, Wydawnictwo Sic!, Warszawa.

Beck U., Giddens A., Lash S. 2009. *Modernizacja refleksyjna*, Wydawnictwo Naukowe PWN, Warszawa.

Brislin R., Yoshida T. 1994. *Intercultural Communication Training. An Introduction*, Sage Publ., London.

Castells M. 2007. *Społeczeństwo sieci*, Wydawnictwo Naukowe PWN, Warszawa.

Frankl V. 1978. *Nieuświadomiony Bóg*, Instytut Wydawniczy PAX, Warszawa.

Giddens A. 2007. *Nowoczesność i tożsamość*, Wydawnictwo Naukowe PWN, Warszawa.

Jaskuła S. 2011a. *Intercultural space. Contemporary challenges and perspectives*, [w:] *Poland – Jordan – European Union: Future Aspects*, red. S. Abudayeh, K. Bojko, Jordan University Press, Amman, Jordan.

Jaskuła S. 2011b. *Informacyjna przestrzeń tożsamości*, [w:] *Tożsamość i komunikacja*, red. L. Dyczewski, J. Szulich, R. Szwed, Wydawnictwo KUL, Lublin.

Jaskuła S. 2012. *Symbioza i odrębność dwóch światów*, [w:] *Współczesna przestrzeń tożsamości*, red. S. Jaskuła, L. Korporowicz, „Politeja" – Jagiellonian Cultural Studies, Pismo Wydziału Studiów Międzynarodowych i Politycznych UJ, nr 2/20, Kraków.

Jedność i różnorodność. Kultura vs. kultury. 2010, red. E. Rekłajtis, R. Wiśniewski, J. Zdanowski, Wydawnictwo Uniwersytetu Kardynała Wyszyńskiego, Warszawa.

Korporowicz L. 1983. *Kompetencja kulturowa jako problem badawczy*, „Kultura i Społeczeństwo", nr 1.

Korporowicz L. 1996. *Tworzenie sensu. Język. Kultura. Komunikacja*, Oficyna Naukowa, Warszawa.

Korporowicz L. 2007. *Komunikacja jako rzeczywistość interakcyjna człowieka*, [w:] *Kultura – Media – Społeczeństwo*, red. D. Wadowski, Wydawnictwo KUL, Lublin.

Korporowicz L. 2008. *Cultures in Interaction. Towards a Community of Heritage*, „Politeja" Pismo Wydziału Studiów Międzynarodowych i Politycznych UJ, nr 2/1 (10/1), Kraków.

Korporowicz L. 2009. *Komunikacja międzykulturowa jako transgresja*, [w:] *Dialog na pograniczach kultur i cywilizacji*, red. T. Paleczny, M. Banaś, Wydawnictwo Uniwersytetu Jagiellońskiego, Kraków.

Korporowicz L. 2010. *Podmioty przestrzeni międzykulturowej*, [w:] *Jedność vs. różnorodność*, red. J. Zdanowski, E. Rekłajtis, R. Wiśniewski, Wydawnictwo Uniwersytetu Kardynała Stefana Wyszyńskiego, Warszawa.

Korporowicz L. 2011. *Socjologia kulturowa. Poszukiwania i kontynuacje*, Wydawnictwo Uniwersytetu Jagiellońskiego, Kraków.

Kłoskowska A. 1981. *Socjologia kultury*, PWN, Warszawa.

Kłoskowska A. 2005. *Kultury narodowe u korzeni*, Wydawnictwo Naukowe PWN, Warszawa.

Lusting M.W., Koster J. 2006. *Intercultural Competence. Interpersonal Communication Across Cultures*, Pearson Education, Boston.

Sztompka P. 2005. *Socjologia zmian społecznych*, Znak, Kraków.

KWARTALNIK PEDAGOGICZNY
2015 NUMER 4(238)

Sylwia Jaskuła
Państwowa Wyższa Szkoła Informatyki i Przedsiębiorczości w Łomży

Zastosowania ewaluacji edukacyjnej. Trzy rodzaje koncentracji

Summary

APPLICATIONS OF EDUCATIONAL EVALUATION. THREE TYPES OF CONCENTRATION

Contemporary applications of evaluation in education should be accompanied by an educational diagnosis, i.e. the use of educational diagnosis should be supported by evaluation. The use of evaluation in educational practice, as well as diagnosis, should precede proper understanding of these notions and strategies. The level of recognition of these functions, roles, goals and needs of evaluation influences its proper implementation and application. Usability of evaluation in education can be seen in the context of three concentrations: sociological, educational and socio-technical. The effectiveness of a particular concentration depends on the individual needs of a specific educational environment.

Key words: educational evaluation, sociological concentration of evaluation, pedagogical concentration of evaluation, socio-technical concentration of evaluation, evolution of evaluation.

red. Paulina Marchlik

W polskim systemie oświaty, diagnostyka edukacyjna istotnie zdystansowała badania ewaluacyjne. Choć początki ewaluacji w edukacji polskiej sięgają lat 90., to w tamtym okresie nie przyjęła się ona w pełni ze względu brak odpowiednio przygotowanych zarówno wykonawców, jak i odbiorców. Za to czas „egzaminów zewnętrznych" skutecznie rozwinął diagnozę edukacyjną, nadając jej obecnie już status dojrzałości i świadomości zarówno metodologicznej, jak i metodycznej. Teoria i praktyka diagnostyki edukacyjnej w realiach polskiego systemu

edukacji w inicjatywnym okresie rozwoju ewaluacji była nieporównywalnie znacznie bardziej rozwinięta, świadoma swych zadań, dojrzała akademicko, a prace i działania jej najwybitniejszego animatora, Bolesława Niemierki stwarzały grunt właściwego kierunku dojrzewania (Niemierko 1998, 2009, 2010).

We współczesnej rzeczywistości szkolnej, gdzie ewaluacja została wprowadzona jako obligatoryjne badanie, jej zastosowaniom winna towarzyszyć diagnoza edukacyjna, i odwrotnie: zastosowania diagnozy edukacyjnej powinny być wspierane przez ewaluację. Ewaluacja i diagnoza mogą, a nawet powinny w procesach badania, analizowania i rozpoznawania wartości edukacyjnych wzajemnie korzystać ze swoich zasobów (Jaskuła 2012b). Należy przy tym pamiętać, że są to dwie różne metody badawcze, nie tylko na poziomie metodologicznym, ale również zakresowym, planistycznym, decyzyjnym oraz w zakresie wykorzystania ich wyników. W szerokim spektrum używanych współcześnie w edukacji metod badawczych, wręcz nierozsądne wydaje się stosowanie ich niezależnie od siebie, nie wykorzystując ich wzajemnego zwielokrotnionego potencjału, w szczególności na lokalnym poziomie funkcjonowania instytucji edukacyjnych. Obowiązująca w każdej placówce, na podstawie rozporządzenia o nadzorze pedagogicznym, „ewaluacja wewnętrzna" ma szansę pełnego wykorzystania rozwojowych walorów ewaluacji, a diagnoza edukacyjna może objąć swym zakresem obszary, które faktycznie wymagają wewnętrznej, pogłębionej analizy. Ponadto w takiej mikroskali zwiększa się szansa, ale i możliwość zaprojektowania tych badań w taki sposób, żeby z jednej strony wzajemnie się uzupełniały, a z drugiej strony wykorzystywały swoje zasoby.

Zmienne losy ewaluacji w edukacji, poczynając od fazy fascynacji jej możliwościami, poprzez fazę obaw i negacji, do fazy obligacji i obowiązkowego jej stosowania wymuszonego przepisami prawnymi lub procedurami[1], nie sprzyjały rozwojowi jej rozumienia i zastosowania. Liczne działania edukacyjne mające na celu przybliżenie sensu i wartości oraz możliwości wykorzystywania ewaluacji w oświacie, pozbawione często jednorodnego i spójnego zorientowania, nie poprawiały wszechobecnego stanu chaosu (Jaskuła 2011, 2014a). Jej masowość i często wymuszone formy stosowania sprowadziły ją często do elementów odprawianego „rytuału", a tym samym odchodząc od celowego wartościowania działań w kierunku wymuszonej konieczności. Jednym z największych problemów, z którym się obecnie zmagamy, jest pozbawienie ewaluacji jej społecznych i podmiotowych wartości, i ukierunkowanie jej na biurokratyzację i instrumen-

[1] Mowa tu o rozporządzeniach o nadzorze pedagogicznym obowiązującym w oświacie od 2009 roku.

talizację, a także brak właściwego i celowego wykorzystania wyników badań ewaluacyjnych.

Ewolucja ewaluacji

Współczesne rozumienie ewaluacji ulegało na przestrzeni lat modyfikacji i należy pamiętać, że definiowanie ewaluacji w początkach jej istnienia i po 90 latach przemian jest odmienne. Jeśli obecnie ewaluacja ma być podstawą uczciwej praktyki, to musi brać swój początek z analiz tego, czym jest ewaluacja współcześnie i jakim przemianom uległa od początków jej istnienia. Odniesienie obecnych przemian badań ewaluacyjnych do jej korzeni wydaje się zabiegiem niezbędnym w procesie poznawania, rozumienia, oceny i oszacowania jej wartości. Tym bardziej, że kolejne generacje ewaluacji, od pierwszej do czwartej (a nawet piątej) (Jaskuła 2010), kładły nacisk na odmienne elementy, a co się z tym wiąże inne funkcje ewaluacji stanowiły podstawę praktycznego jej stosowania. Ewolucja ewaluacji od pierwszej do czwartej generacji, pokazuję drogę odchodzenia od aspektów pomiarowych czy oceniających w kierunku wartości skoncentrowanych na: procesualności, dialogiczności, kontekstualności, uspołecznieniu, interakcji, otwartości, animacyjności czy integralności.

W początkowym czasie stosowania zinstytucjonalizowanych studiów ewaluacyjnych orientowano je głównie na uzyskiwanie mierzalnych rezultatów swoich analiz. To upraszczanie ewaluacji, poprzez utożsamianie jej z procedurami pomiarowymi, wpisywało się w ogólną tendencję badawczą tego okresu związaną z dominacją technik pomiaru dydaktycznego, psychometrycznego i socjometrycznego. Redukcjonizm strategii metodologicznych ewaluacji, powodował zarówno uproszczenie samego jej rozumienia, jak również jej funkcji i celów. Nadmierna symplifikacja problematyki ewaluacyjnej wiązała się również z ograniczeniem wymiarów analiz przedmiotu ewaluacji, pozbawiając je aspektów przede wszystkim społecznych, ale również organizacyjnych czy instytucjonalnych. Takie technokratyczne podejście nie uwzględniało złożoności zjawisk i procesów rzeczywistości społecznej. Stosowanie prostych procedur pomiarowych nie dawało wystarczających odpowiedzi na temat natury zjawisk społecznych, ani też nie umożliwiało zaobserwowania dynamiki zmiany społecznej. Już w tym początkowym okresie powstawania badań ewaluacyjnych, uznawany za jej ojca – Ralph Tyler, widział potrzebę rozszerzenia celów i funkcji ewaluacji, wychodzących poza stosowanie prostych procedur pomiarowych.

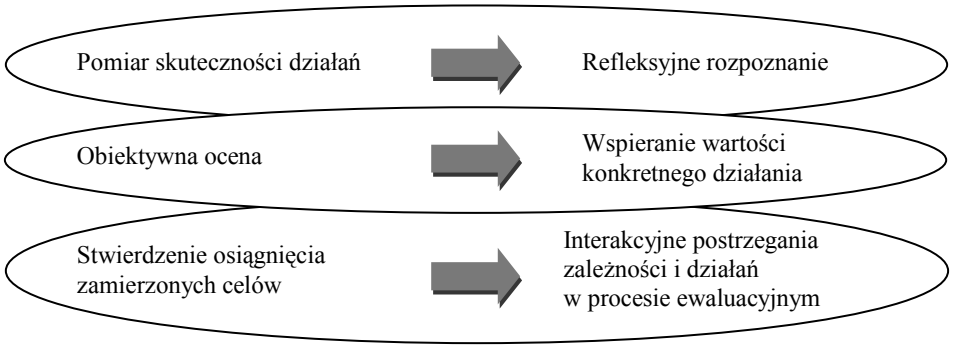
Tabela 1. Proces zmiany rozumienia ewaluacji

Generacje ewaluacji	Rozumienie ewaluacji
Pierwsza generacja ewaluacji – pomiar (*measurement*)	Główny nacisk kładziono na mierzalność efektów i środków. Ilościowe uchwycenie zjawisk miało pozwolić na ich porównywalność i obiektywizm oceny. Rola ewaluatora w oświacie została ograniczona do funkcji technicznych. Miał on wiedzieć jak wykorzystywać ogromną liczbę „psychometrycznych narzędzi pomiarowych", które były dostępne w latach 40. ubiegłego stulecia, do oceny osiągnięć pracy uczniów.
Druga generacja ewaluacji – opis (*description*)	W obszarze zainteresowań ewaluacji nadal pozostawał pomiar, ale rozszerzyła ona zainteresowanie o opis praktyki interwencji, aby usprawniać programy, a także sięgać po alternatywne środki. Rola ewaluatora została w ten sposób sprowadzona do opisu mocnych i słabych stron programów w odniesieniu do założonych celów.
Trzecia generacja ewaluacji – decyzja/ ocena (*judgement*)	W tej generacji, podobnie jak w poprzedniej, ewaluator jest zewnętrznym, neutralnym obserwatorem, który w oparciu o metody naukowe dostarcza danych umożliwiających stwierdzenie stopnia osiągnięcia celów i na ich podstawie wyprowadza konsekwencje w procesie decyzyjnym programu. Ewaluator zajmuje się nie tylko pomiarem osiągniętych celów programu i opisem sposobu funkcjonowania, lecz także dokonuje oceny, czy same cele zostały prawidłowo określone w perspektywie optymalizacji drogi ich osiągania.
Czwarta generacja ewaluacji – proces (*the Process*)	Idea ewaluacji opracowana przez Egon G. Gubę i Yvonne S. Lincoln wykorzystuje metodologię paradygmatu konstruktywizmu. Autorzy postulują oparcie ewaluacji na głębokiej partycypacji wszystkich stron zaangażowanych w program. Ewaluator występuje w roli moderatora w negocjacjach dotyczących ustalania potrzeb informacyjnych, kryteriów i warunków przeprowadzenia ewaluacji, a także w interpretacji danych i wykorzystaniu wyników. Obiektywizm pomiaru i obserwacji schodzi na drugi plan wobec prymatu wiarygodności celów i metodologii badania. W ewaluacji czwartej generacji dominuje podejście jakościowe i nacisk na identyfikowanie złożonych aspektów rzeczywistości, które mają wpływ na całokształt procesu wartościowania.

Źródło: Opracowanie własne na podstawie Guba, Lincoln 1989.

Współczesny sens ewaluacji, jak również jej funkcje i cele, którym ma służyć, przeszły transformację. Główne elementy tej modyfikacji związane są ze

zmianą pewnych obszarów istotności, przechodząc od zawężonego pomiaru skuteczności działań w kierunku szerokiego refleksyjnego rozpoznawania; od obiektywnej oceny, oznaczającej zakończenie działania do procesu wspierania wartości konkretnego działania; wreszcie, od subiektywnego lub intersubiektywnego stwierdzenia osiągnięcia zamierzonych celów do interakcyjnego postrzegania zależności i działań w procesie ewaluacyjnym.



Rysunek 1. Ewolucja ewaluacji
Źródło: opracowanie własne.

Ewaluacja a nauka

Ewaluacja jako dotychczas stosowane badanie nie aspirowała do miana dyscypliny naukowej. Ze względu na swój ocenny charakter, nie wpisywała się w klasyczne rozumienie dyscyplin trzymających się kryterium „nauki wolnej od oceny". Ewaluacja, której istotą jest wartościowanie, w samej swojej naturze nie jest pozbawiona ocen działań, postaw i argumentów.

Współcześnie rozpoczęła się dyskusja nad zmianą statusu ewaluacji – w kontekście coraz częściej dostarczanych dowodów na niesłuszność doktryny nauki wolnej od wartości. Reorientacja ta argumentowana jest wykorzystaniem w szerokim zakresie przez nauki społeczne wartościowania działań i postaw. Wskazuje się też inne dyscypliny będące w ogromnej mierze ewaluacyjnymi, a uprawiane są od tysięcy lat, tj. medycyna, logika, etyka czy matematyka. Pierwsza z wymienionych wykorzystuje element wartościowania przy diagnozie chorób, działaniach prozdrowotnych czy profilaktycznych. Logika w znacznej mierze skupia się na ocenie argumentów, zaś etyka koncentruje się na ewaluacji działań i postaw.

Jednym z orędowników ewaluacji jako dyscypliny naukowej jest Michael Scriven – brytyjski ekspert teorii i praktyk ewaluacyjnych. Prezes Amerykańskiego Stowarzyszenia Badań Edukacyjnych oraz Amerykańskiego Towarzystwa Ewaluacyjnego, będący redaktorem i współzałożycielem „Journal of Multidisciplinary Evaluation" w swoim artykule o znaczącym tytule *Ewaluacja jako rewolucyjna dyscyplina* wskazuje argumenty potwierdzające naukowy charakter ewaluacji. Odcina się od pozytywistycznej lub neopozytywistycznej filozofii nauki, według której nauka nie może obejmować twierdzeń oceniających, ze względu na ich subiektywny charakter, nieprecyzyjność i/lub niepoddawanie się testom (Scriven 2013: 7–16).

Scriven podważa przekonanie o nienaukowym charakterze ewaluacji, argumentując nieprawidłowość jej rozumienia powierzchowną analizą języka ewaluacyjnego. Potwierdza, że może ona poruszać kwestie nierozstrzyganych sporów o gusta, ale większość podejmowanych zagadnień jest sprawdzalna i często wg autora „obiektywnie prawdziwa". Przy tym pełnią one w nauce rolę polegającą na wyrażaniu klasyfikacji dobrych i złych hipotez, teorii i instrumentów.

Rozważania Scrivena wpisują się w problem demarkacji nauki, który stały się jednym z podstawowych zagadnień powstałej w XX wieku filozofii nauki. Przechodząc ewolucję, w której znaczące miejsce zajmowało rygorystyczne stanowisko logicznego pozytywizmu, uznające, że aby zdania miały charakter sensowny, muszą się opierać bądź na obserwacjach empirycznych, bądź wynikać z logiki. Poprzez jedno z najbardziej wpływowych kryteriów demarkacji zaproponowane przez Karla Poppera, który odrzucał indukcjonizm jako drogę dochodzenia do pewnej wiedzy i zaproponował procedurę falsyfikacji (polegająca na stawianiu ambitnych hipotez i dążeniu do ich obalania). W odmienny sposób na problem demarkacji nauki, i czegoś co jedynie do nauki aspiruje, patrzyli historycznie zorientowani badacze, tacy jak Pierre Duhem czy Thomas Kuhn. Analizując rozwój nauki, pokazali, że nie istnieje jedno kryterium naukowości, lecz że są one historycznie zmienne. Poszczególne twierdzenia naukowe znajdują oparcie w systemie wiedzy charakterystycznym dla swojego czasu i operującym odmiennymi kryteriami demarkacji (niewspółmierność interteoretyczna[2])

[2] Ang. *intertheoretical/theoretical incommensurability* – jest to relacja jaka zachodzi między dwiema bardzo ogólnymi teoriami naukowymi, polegająca na tym, że niektóre, rywalizujące ze sobą lub po sobie następujące teorie naukowe są ze sobą niezgodne, bo np. dotyczą częściowo tylko tego samego zakresu zjawisk, lub posługują się odmiennymi metodami badawczymi, czy innymi standardami oceny tego, co jest w danym momencie uznawane za naukę.

(Nickles 2006: 193). Trudności wokół wyłonienia kryterium demarkacji, doprowadziły wielu teoretyków do stwierdzenia, że nie istnieje jedno takie kryterium. Odmienne kryteria demarkacji powinno się stosować do nauk formalnych (logiki i matematyki), nauk ścisłych, przyrodniczych, społecznych i humanistycznych.

Argumenty użyte przez Scrivena znajdują uzasadnienie szczególnie wśród badaczy, którzy stoją na stanowisku, iż niemożliwe jest ustalenie abstrakcyjnego kryterium demarkacji, a odróżnienie nauki od nie-nauki jest ściśle zależne od konkretnego kontekstu, w którym rozróżnienie się przeprowadza. Sama ewaluacja odpowiada założeniom utylitarystów amerykańskich twierdzącym, że o tym, czy dana teoria jest naukowa czy nie, decyduje między innymi jej użyteczność. Oczywiście samo kryterium użyteczności nie wyznacza kryterium naukowości, dlatego, żeby o ewaluacji mówić jako dyscyplinie naukowej musiałaby ona spełnić szereg innych kryteriów.

Problem naukowości ewaluacji nie jest tak prosty i oczywisty, i zapewne potrzeba jeszcze czasu, na ugruntowanie się jej wśród dyscyplin naukowych.

Wykorzystanie wyników ewaluacji

Ewaluacja powinna być skierowana na praktyczność zastosowania jej wyników, co wydaje się być współcześnie kluczowym elementem jej rozumienia i stosowania, bowiem wymóg rozwojowy płynie z praktyki silniej niż z samej teorii. Dlatego ewaluację uznajemy za w pełni wykonaną, jeśli wykorzystane zostaną jej rezultaty. Zastosowanie efektów ewaluacji jest na tyle istotne w całym jej procesie, że już w pierwszym jego elemencie tj. analizie potrzeb rozpoczyna się planowanie możliwych wariantów aplikacji rezultatów badania. Uwzględniając warunek użyteczności badań ewaluacyjnych, jeżeli niemożliwe jest wykorzystanie ich wyników, należy zastanowić się nad potrzebą wykonywania takich badań. W tej perspektywie istotne jest również zwrócenie uwagi na to, jak wszystkie zadania realizowane w toku procesu ewaluacyjnego mogą wpływać na wykorzystanie ewaluacji.

Rolę ewaluacji skoncentrowanej na wykorzystaniu, stale podkreśla w swoich publikacjach, począwszy od lat siedemdziesiątych Michael Quinn Patton (Patton 2008, 2010, 2012). Kilkudziesięcioletnia historia tego podejścia, z jednej strony, wciąż zwiększa potencjalną użyteczność i faktyczne wykorzystanie

badań ewaluacyjnych, a z drugiej, ukazuje przemianę czynników wpływających na zakres, możliwości i charakter owej stosowalności. Rozumienie wspomnianych czynników, uzależnionych od kontekstu społecznego, wspomagają badania nad wykorzystaniem ewaluacji[3], które ukazały jej procesualny charakter, oparty na myśleniu systemowym (Patton 2012). Wykorzystanie wyników ewaluacji, będąc istotnym i jednocześnie złożonym procesem, wymaga starannego zaplanowania. Powinno ono uwzględnić różnych aktorów pełniących wielorakie funkcje w ewaluacji, ale równocześnie, ze względu na wartościującą istotę badań, powinno określić czyje wartości są najważniejsze w kontekście wykorzystania wyników i ewentualnego wdrażania rekomendacji. Tym samym skoncentrowanie na wykorzystaniu wymaga pokonania drogi między ogólnymi założeniami dotyczącymi potencjalnych odbiorców i sposobów aplikacji wyników a szczegółowym określeniem faktycznych tzw. użytkowników docelowych i ich możliwości konkretnego wykorzystania rezultatów badania.

Określenie osób, które mogą i chcą wykorzystać wyniki ewaluacji, staje się kluczowym zadaniem już w pierwszym etapie procesu ewaluacyjnego. Zaangażowanie się w ewaluację, a tym samym wzięcie odpowiedzialności na wszystkich jej etapach, poczynając od zainicjowania badania, określenia celu, poprzez projektowanie, aż po raport końcowy wraz z etapem upowszechnienia wyników, zwiększa w znaczącym stopniu prawdopodobieństwo jej wykorzystania. Jeżeli nie uda się zainteresować badaniem w początkowych jego etapach różnych aktorów sceny ewaluacyjnej, zapewne nie będą oni również chętni a jednocześnie przygotowani do wykorzystania wyników po jego zakończeniu.

Istotnym elementem wykorzystania wyników ewaluacji jest zatem – co często w swoich artykułach podkreśla M.Q. Patton – psychologia wykorzystywania ewaluacji. Uczestnicy procesu ewaluacyjnego tym chętniej wykorzystują badania ewaluacyjne, im lepiej rozumieją ich istotę, im aktywniej uczestniczą w ich projektowaniu i im bardziej czują się za nią nieodpowiedzialni. Tę zależność pokazuje zarówno wieloletnia praktyka ewaluacyjna, jak również systematyczne badania nad wykorzystaniem ewaluacji (Patton 2008).

Michael Q. Patton wprowadzając do terminologii ewaluacyjnej pojęcie „ewaluacja skoncentrowana na wykorzystaniu", z jednej strony, podkreślił istotność

[3] Początkowe badania nad wykorzystaniem badań ewaluacyjnych obejmowały duże ewaluacje prowadzone w sektorze zdrowia publicznego; M.Q. Patton, *Utilization-Focused Evaluation*, Sage, Beverly Hills 1978.

i potrzebę wykorzystania rezultatów uzyskanych w toku przeprowadzonych badań, ale z drugiej, zaproponowanie perspektywy – „skoncentrowania na wykorzystaniu" sugeruje, że istnieją takie ewaluacje, których głównym zadaniem nie jest bezpośrednia i wymierna aplikacja wyników. Tym samym oddzielił on ewaluację skoncentrowaną na wykorzystaniu od ewaluacji ogólnej. Różnica między tymi dwoma rodzajami badań ewaluacyjnych tkwi nie tylko w samym procesie wykorzystywania wyników ewaluacji, ale sięga do rozbieżności w charakterze, celach i motywach ich stosowania. Ewaluacja ogólna zmierza do zrozumienia najistotniejszych wartości badanych działań, podczas gdy ewaluacja skoncentrowana na wykorzystaniu aspiruje poza samo poznanie, dochodząc do ustalenia możliwych dróg ich doskonalenia. Oczywiście dzieje się to poprzez decyzje powzięte na podstawie uczynionych ustaleń. W przypadku tzw. ewaluacji ogólnej trudno też mówić o jedynym tylko jej modelu. Ich odmian może być wiele mieszcząc się w dwubiegunowych granicach, z których jeden kraniec może charakteryzować potrzeba koncentracji na „rozumieniu" problemu, a więc na poszukiwaniu odpowiedzi na pytania „jak jest?" i „dlaczego tak jest?". W drugim przypadku badania ewaluacyjne mogą wyjść poza stwierdzenia o stanie badanego przedmiotu i skoncentrować się na pytaniach: „co należy zrobić, aby poprawić istniejący stan?", „jak to zrobić?", „kogo zaangażować w proces planowania zmian?". Pierwsze podejście jest typowe dla ujęcia socjologicznego, które zmierza do znalezienia ogólnych prawidłowości zjawisk, szczególnie w kontekście społecznym. Z kolei drugie, można określić jako pedagogiczne. Wychodzi ono poza diagnozę sytuacji, próbując określić możliwe sposoby doskonalenia, korekty bądź rozwiązania zbadanego problemu. W pierwszym ujęciu nie ma wymogu przeniesienia wyników badań na działania operacyjne, które towarzyszą ewaluacji skoncentrowanej na wykorzystaniu. Postulat taki pojawia się w ujęciu pedagogicznym, a w najbardziej zaawansowanym stopniu realizuje się w trzecim z proponowanych podejść, które określić możemy jako socjotechniczne. W ostatnim przypadku w procesie badawczym formułowane są wyniki w sposób umożliwiający bezpośrednie wdrażanie poprzez konkretnie podejmowane decyzje. Ujęcie socjotechniczne ewaluacji odnosi się także do planowania i skutecznego kreowania zmian społecznych, poprzez implementowane dyrektywy mające podwaliny w konkretnych ocenach i wartościach będących rezultatem zaprojektowanych i przeprowadzonych badań.

Tabela 2. Ujęcie socjologiczne, pedagogiczne i socjotechniczne ewaluacji

	1	2	3	4	5	6	7
Elementy procesu ewaluacyjnego	Analiza potrzeb	Plan	Projekt	Generowanie danych	Analiza danych	Upowszechnienie wyników	Wykorzystanie wyników
Ujęcie socjologiczne	Poszukuje odpowiedzi na pytania „jak jest?", „dlaczego tak jest?", koncentruje się na „rozumieniu" problemu.						
Ujęcie pedagogiczne	Wychodzi poza stwierdzenie o stanie badanego przedmiotu, koncentruje się na poszukiwaniu odpowiedzi na pytania: „co należy zrobić, aby poprawić istniejący stan?", „jak to zrobić?", „kogo zaangażować w proces planowania zmian?"						
Ujęcie socjotechniczne	Wyniki są formułowane w sposób umożliwiający bezpośrednie wdrażanie poprzez konkretnie podejmowane decyzje.						

Źródło: opracowanie własne.

Michael Q. Patton, istotnie odróżnia pojęcia ewaluacji skoncentrowanej na wykorzystaniu i ogólnie pojętej ewaluacji. Wskazując różnice między nimi podkreśla, iż ewaluacja skoncentrowana na wykorzystaniu:

1. dokonywana jest na rzecz określonych głównych użytkowników docelowych,
2. wykonywana jest wspólnie z nimi,
3. odnosi się do określonego zamierzonego wykorzystania,
4. zawiera konieczność określenia zamierzonego wykorzystania przez użytkowników docelowych (Patton 2013: 19).

Ewaluacja skoncentrowana na wykorzystaniu może badać dowolny przedmiot, pełnić dowolną funkcję, korzystać z dowolnego rodzaju, modelu czy danych. Taka ewaluacja stanowi proces, którego celem jest podjęcie decyzji we współpracy z określoną grupą głównych użytkowników, poprzez skupienie się na ich zamierzonym wykorzystaniu ewaluacji. Znaczenie czynnika ludzkiego w przewidywaniu możliwości zastosowania ewaluacji uwrażliwia na pracę z użytkownikami ukierunkowaną na określenie zamierzonego wykorzystania.

Z tych względów szczególnie istotne staje się określenie grupy osób rozumiejących ewaluację, doceniających jej możliwości i potencjał, jak również zaangażowanych w jej przebieg oraz będących zainteresowanymi jej wykorzystaniem. Spojrzenie na ewaluację z tej perspektywy podkreśla potrzebę skupienia się na użytkowniku. Ponieważ żadne badanie ewaluacyjne nie może służyć interesom wszystkich zainteresowanych jej przebiegiem i wykorzystaniem w równym stopniu, w przypadku ewaluacji skoncentrowanej na wykorzystaniu, ważnym staje się określenie grupy głównych odbiorców wyników, którzy mają faktyczny wpływ na jej aplikowalność.

* * *

O wartości ewaluacji stanowi jej użyteczność, ale nie należy popadać w uproszczone jej definicje sprowadzając do parametrów o charakterze ekonometrycznym lub technokratycznym. Szereg działań podjętych w ramach projektowania badań ewaluacyjnych nabiera swój sens, gdy znajduje praktyczne wykorzystanie w konkretnym środowisku edukacyjnym. Z wyżej wymienionych względów warto odróżnić co najmniej trzy prakseologiczne koncepcje procesu ewaluacyjnego: socjologiczną, pedagogiczną i socjotechniczną. Wszystkie one mają swoje specyficzne cechy i funkcje, ale w obecnych realiach systemu edukacyjnego, ostatnia z nich wydaję się być najbardziej potrzebna i jednocześnie zaniedbana. Fakt ten w szczególny sposób powinni wziąć pod uwagę decydenci obecnego i przyszłego systemu edukacji, aby nie pogłębiać istniejących dysfunkcji w szkolnych realiach społecznej obecności ewaluacji.

Bibliografia

Guba E.G., Lincoln Y.S. 1989. *Fourth generation evaluation*, Sage, London.

Jaskuła S. 2010. *O potrzebie re-ewaluacji w społeczeństwie nadmiaru informacji*, [w:] *Ewaluacja programów operacyjnych – konteksty, dylematy, praktyki*, red. K. Kasprzyk, Urząd Marszałkowski Województwa Kujawsko-Pomorskiego w Toruniu, Toruń.

Jaskuła S. 2011. *Ewaluacja versus akredytacja – dwa światy w jednej przestrzeni*, [w:] *Ewaluacja w edukacji: koncepcje, metody, perspektywy*, red. B. Niemierko, M.K. Szmigiel, Grupa Tomami, Kraków.

Jaskuła S. 2012a. *Ewaluacja rozwojowa jako wyzwanie współczesnego systemu edukacji*, Zarządzanie Publiczne. Zeszyty Naukowe Instytutu Spraw Publicznych Uniwersytetu Jagiellońskiego 3 (19).

Jaskuła S. 2012b. *Mutualizm ewaluacji i diagnozy edukacyjnej*, [w:] *Regionalne i lokalne diagnozy edukacyjne*, red. B. Niemierko, M.K. Szmigiel, Grupa Tomami, Kraków.

Jaskuła S. 2013. *Pedagogika ewaluacji*, „Teraźniejszość – Człowiek – Edukacja". Kwartalnik Myśli Społeczno-Pedagogicznej, Wydawnictwo Naukowe Dolnośląskiej Szkoły Wyższej, nr 4 (64).

Jaskuła S. 2014a. *Edukacja do ewaluacji*, Animacja Życia Publicznego, Zeszyty Centrum Badań Społeczności Polityk Lokalnych, 4 (17).

Jaskuła S., Korporowicz L. 2014b. *Odrębność i współgranie. Polskie relacje diagnostyki i ewaluacji edukacyjnej*, „Ruch Pedagogiczny" nr 2.

Korporowicz L. 2000. *Społeczne wyzwania ewaluacji*, „Remedium", nr 10/92.

Korporowicz L. 2001. *Ewaluacja i społeczeństwo*, [w:] *Rozumienie zmian społecznych*, red. E. Hałas, TN KUL, Lublin.

Korporowicz L. 2010a. *Interakcyjna misja ewaluacji*, [w:] *Ewaluacja w nadzorze pedagogicznym. Konteksty*, red. G. Mazurkiewicz, Wydawnictwo Uniwersytetu Jagiellońskiego, Kraków.

Korporowicz L. 2010b. *Współczesne koncepcje ewaluacji w kontekście wyzwań rozwoju regionalnego*, [w:] *Ewaluacja programów operacyjnych na poziomie regionalnym – teoria i praktyka*, Urząd Marszałkowski Województwa Opolskiego, Opole.

Nickles T. 2006. *Problem of Demarcation*, [w:] *The Philosophy of Science. An Encyclopedia*, eds. S. Sarkar, J. Pfeifer, Routledge, New York–London.

Patton M.Q. 2008. *Utilization-Focused Evaluation*, (4th ed.), Sage, London.

Patton M.Q. 2010. *Developmental Evaluation: Applying Complexity Concepts to Enhance Innovation and Use*, Guilford Press, New York.

Patton M.Q. 2012. *Essentials of Utilization-Focused Evaluation*, Sage Publications, Thousand Oaks.

Patton M.Q. 2013. *Ewaluacja skoncentrowana na wykorzystaniu*, [w:] *(R)ewaluacja. Poszukiwanie nowych metod oceny efektów*, red. A. Haber, Z. Popis, PARP, Warszawa.

Scriven M. 2013. *Ewaluacja jako rewolucyjna dyscyplina*, [w:] *(R)ewaluacja. Poszukiwanie nowych metod oceny efektów*, red. A. Haber, Z. Popis, PARP, Warszawa.

Simons H. 1987. *Getting to Know School in a Democracy. The Politics and Process of Evaluation*, The Flamer Press, London.

KWARTALNIK PEDAGOGICZNY
2015 NUMER 4(238)

Grażyna Szyling
Uniwersytet Gdański

Ewaluacja szkolnych systemów oceniania uczniów. Poszukiwanie znaczeń

Summary

EVALUATION OF SCHOOL STUDENT ASSESSMENT SYSTEMS. SEARCHING FOR MEANINGS

In the text a critical reflection is undertaken on the evaluation of school assessment systems treated as a social category, "pulsating" due to its relation to its changeable context. Analyses have revealed an extensive realm of reduction of meanings and values carried out in the course of assessment systems creation and design, and throughout performance of their evaluation. One of the effects of those actions is dissonance between records in documents and their practical dimension. In turn, the picture of school reality, simplified and subjected to external evaluation procedures, does not so much block a change it does not make it possible to recognise its need should it not fall within the need to meet uniform requirements of the state regarding education. The same logic results in internal evaluation in the school assessment system being granted the status of undisputed yardstick for the educational reality, which, should such a need arise, must be changed for them to better match the procedures formulated in the document.

The text concludes with a question on the need of assessment and values connected with it, the reading of which in the context of specified horizon of meanings may prevent assessment systems from their axiological dimension being reduced, or solidify the current form confined to a set of procedures, whose deeper sense is not reflected on by anyone at school. The latter option does not imply elimination of problems with assessment but those problems remaining unspoken of and their further increase.

Key words: school assessment system, evaluation, reduction of meanings, blocking of change.

red. Paulina Marchlik

Rzeczywistość polskiej edukacji jest zmienna i zależna od często doraźnych koncepcji polityki oświatowej (zob. Niemierko 2013), które w dłuższej perspektywie albo zostają złożone do lamusa i zapomniane, albo stają się – nieco zwietrzałą, acz trwałą – częścią codzienności szkolnej. O tej szczególnej trwałości świadczy ich obecność w potocznych i administracyjnych dyskursach oświatowych, obecność podkreślana skostniałymi formami językowymi i sztywnymi szkolnymi procedurami, których głębszego sensu nikt nie docieka. Jednym z takich funkcjonujących do dziś rozwiązań, które wdrożone zostały wraz z reformą oświaty w 1999 roku (Rozp. MEN 1999), jest wewnątrzszkolny system oceniania uczniów – dalej nazywany WSO – w niniejszym artykule rozpatrywany w kontekście jego ewaluacji. Przyjmuję przy tym, że oczywistość i bezdyskusyjność tego dokumentu oraz jego funkcji regulacyjnych blokuje możliwe zmiany w ocenianiu szkolnym, o których potrzebie zdają się świadczyć między innymi niedawne dyskusje o ocenianiu toczące się wokół zmian w Ustawie o systemie oświaty (2015) czy na forach III Kongresu Polskiej Edukacji.

W dalszych rozważaniach traktować będę szkolny system oceniania jak kategorię społeczną, dynamiczną i nieostrą, „pulsującą" dzięki relacji ze zmiennym kontekstem (Rutkowiak 1995: 25), który wyznaczany jest przez oświatowe regulacje, praktyczne działania, potrzeby społeczeństwa i uczestników procesu edukacyjnego oraz teorię pedagogiczną. Podejście takie wpisuje się w Brunerowskie myślenie o kulturze szkoły, rozumianej „jako sposób życia, który konstruujemy, negocjujemy, poddajemy instytucjonalizacji [...], by nazwać to rzeczywistością" (Bruner 2006: 126). Uzasadnia to podjęcie w niniejszym tekście wieloaspektowej analizy zagadnienia, prowadzącej do jego sproblematyzowania, co w określonych warunkach może pogłębić rozumienie codzienności edukacyjnej i wytworzyć potencjał zmiany (zob. Szyling 2013a).

Założenie to współbrzmi z kluczowym dla tych rozważań rozumieniem ewaluacji jako „zaproszenia do rozwoju", otwartej na przewartościowania i uczenie się (Korporowicz 2010). Jej zasadniczym celem – pisze Leszek Korporowicz (2013: 37) – „jest uczenie się i zmiana o charakterze rozwojowym", której towarzyszy świadomie podejmowana refleksja i poczucie odpowiedzialności za kształtowane przez jej efekty postawy i działania. W takim ujęciu meta-wartością stają się wiedza uczestników ewaluacji oraz ich skupienie „wokół wspólnego dobra wytworzonego podjętym działaniem" (tamże: 42).

Potrzeba ewaluacji? Potrzeba zmiany?

Definicja ewaluacji ujmowanej w kategoriach „zaproszenia do rozwoju", wyklucza wątpliwości, którym dałam wyraz w powyższym śródtytule. Jednak L. Korporowicz zwraca uwagę, że w polskich realiach procesy ewaluacyjne są obciążone nadmiarem zadań kontrolnych i administracyjnych, zawężane do prostego pomiaru, co „prowadzi do postępującej ich alienacji i uprzedmiotowienia" (tamże). Tak sformułowana diagnoza zachęca do poszukiwania impulsu stojącego za powstaniem i ewaluacją WSO, ponieważ on może wskazywać na uspołecznienie tych działań lub ich reaktywny charakter.

Impuls taki z reguły wiąże się z potrzebą, czyli z „czymś, co jest konieczne do przetrwania", ale bywa przeciwstawiane chęciom lub pragnieniom (Marshall 2005: 249), a to już samo w sobie generuje pola napięć. Ponadto badacze nie są zgodni co do wartości potrzeb i sposobu szacowania ich wielkości, ale najczęściej wiążą je z koniecznością, czyli funkcjonalnie pojmowanym warunkiem przetrwania społeczeństwa (tamże). To z kolei podkreśla, że potrzeby powinny być zaspokajane ze względów społecznych. Nie może zatem dziwić, że pojęcie to znalazło swe stałe miejsce w retoryce politycznej – także w polityce oświatowej – jako rodzaj bezdyskusyjnego argumentu uzasadniającego żądanie podjęcia określonych działań.

Zarysowane pole definicyjne pozwala rozpatrywać wdrożenie w praktykę szkoły WSO oraz formalnego wymogu ich ewaluacji jako działania zmierzającego do lepszego zaspokojenia społecznego zapotrzebowania na „dobrą jakościowo edukację". Przyjmuję przy tym dalej, że ową jakość edukacji – silniej niż intencje – konstruują idące za nimi działania oraz ich konsekwencje. Stąd celowe wydaje się dokonanie choć pobieżnego oglądu historii tworzenia analizowanego dokumentu i źródeł wprowadzanych w nim zmian.

Obowiązek stworzenia WSO, uznanego za integralną część statutu placówki, został nałożony na szkoły przez ustawę o systemie oświaty i doprecyzowany po raz pierwszy w 1999 roku, w przywoływanym już rozporządzeniu MEN w sprawie zasad oceniania (...). Poczynione regulacje odczytywane były wówczas – między innymi przez Bolesława Niemierkę (2000) – jako działania sprzyjające większemu zindywidualizowaniu wymagań programowych dzięki uwzględnieniu w ich ustalaniu możliwości uczniów oraz kontekstu pracy szkoły.

Jednak z badań prowadzonych przez Elżbietę Putkiewicz (2000) tuż przed podjęciem przez szkoły prac nad WSO wynikało, że nauczyciele uznawali wytworzenie tego dokumentu za najbardziej niejasne spośród wszystkich postawio-

nych przed nimi zadań związanych z planowaną reformą. Niechętnie opowiadali też o codziennych problemach oceniania szkolnego, „kurczowo trzymając się kilku pojęć sformalizowanego języka pomiaru dydaktycznego, w rodzaju „obiektywności" i „rzetelności" (tamże: 297). Elżbieta Putkiewicz dostrzegała w tym zjawisku podstawową przeszkodę w zbudowaniu przemyślanego systemu szkolnego oceniania. Podobne wątpliwości pojawiały się w nieco późniejszych zewnętrznych badaniach ewaluacyjnych, które prowadził nadzór pedagogiczny i ośrodki doskonalenia nauczycieli (zob. Szyling 2011). Co znamienne, ze zgromadzonych danych wynikało, że w WSO uwzględniano przede wszystkim ministerialne regulacje dotyczące oceniania osiągnięć uczniów, dzięki czemu uzyskiwano efekt ujednolicenia w szkole procedur oceniania. Z drugiej jednak strony, badania te wskazywały, że w dużej części placówek dokumenty związane z ocenianiem zostały potraktowane wyłącznie jako formalne zaspokojenie wymagań systemu oświaty. Już wówczas dostrzegano jednak występowanie tendencji do nadmiernego uszczegółowienia i rygorystycznego przestrzegania zasad określonych w WSO.

Dotychczasowe analizy pozwalają ulokować **zasadniczy impuls prowadzący do powstania szkolnych systemów oceniania w działaniach reformatorskich MEN, niekoniecznie rozumianych i wewnętrznie akceptowanych przez nauczycieli, ale wdrażanych w szkolną codzienność**, co ma duże znaczenie dla kierunku rozważań nie tyle nad potrzebą ewaluacji, ile nad jej charakterem i funkcjonowaniem w realiach edukacyjnych. Leszek Korporowicz (2013) wiąże to zjawisko z bardziej ogólną prawidłowością, czyli nadaniem fundamentalnego znaczenia pewnej koncepcji społecznej, która na użytek jej wykonawców i odbiorców **zostaje stosownie uproszczona, oddalając się od zróżnicowanej prawdy o świecie zewnętrznym**. W proces merytorycznego i formalnego ograniczania kształtu WSO w początkowej fazie ich tworzenia wpisał się także pomiar dydaktyczny w swej najbardziej ortodoksyjnej wersji, która za pożądane i w pewnym sensie oczywiste uznawała blokowanie przez procedury pomiarowe tych działań nauczycieli, które mogłyby zaburzać „obiektywizm" oceny wyrażanej stopniem szkolnym (zob. Szyling 2014).

„Wszystkie te przesunięcia i przewartościowania – pisze L. Korporowicz (2013) – mają bardzo istotny wpływ na pojmowanie, konstrukcję, a następnie oczekiwania, jakie wystosowane zostaną pod adresem badań ewaluacyjnych". Dodam, że zdecydowana większość modyfikacji wprowadzanych w owym czasie w WSO niewiele miała wspólnego z ewaluacją w jakiejkolwiek z jej postaci (zob. Korporowicz, 2011). Sam **przedmiot potencjalnych badań ewaluacyjnych**

znajdował się bowiem w przestrzeni niedomówień i niejasności z jednej strony, a nakazów prawa oświatowego – z drugiej, więc a jedynym sposobem na uniknięcie porażki przez placówkę najczęściej okazywało się naśladowanie zewnętrznych wzorców lub schematów.

Jaki system? Jaka zmiana?

Jak wynika za dotychczasowych rozważań, WSO w początkach swojego funkcjonowania łączył w sobie co najmniej dwa byty: obowiązujące powszechnie wymogi ministerialne oraz właściwości pomiaru dydaktycznego, powiązane z urzędowymi rozwiązaniami, ale zniekształcone przez potoczność. W strukturze tej zwraca uwagę **zmarginalizowanie nauczycielskich systemów oceniania**, które nie są tożsame z przedmiotowymi systemami oceniania (PSO), ale realnie oddziałują na codzienność szkolną. Można przypuszczać, że jest to konsekwencja skażenia nauczycielskich praktyk „grzechem pierworodnym", czyli tak zwaną zmiennością ocen, często postrzeganą w kategoriach dewiacji lub „szarej strefy", którą należy radykalnie ograniczyć, bo jej wyeliminowanie jest niemożliwe ze względów społeczno-wychowawczych (Niemierko 2001). Takie podejście podkreśla **funkcjonalistyczny charakter regulacji wprowadzonych w szkolne ocenianie, sugerując odczytywanie odstępstwa od nich jako dysfunkcji, które w pewnych granicach da się przewidzieć i uwzględnić w procedurach**.

Nadanie takiego znaczenia WSO zostało – jak twierdzi Maria Groenwald (2012) – utrwalone przez stopniowe umacnianie się ważnego „centrum decyzyjnego", czyli systemu egzaminów zewnętrznych. Dostarczał on szkołom przede wszystkim coraz bardziej precyzyjnych procedur i wzorów zadań, a także przykładów formalnego dostosowania jednolitych wymagań do specjalnych potrzeb edukacyjnych uczniów, stwarzając przy tym wrażenie spójnego i niezawodnego systemu. Kilka przykładów praktycznego przenoszenia do codziennego oceniania rozwiązań służących egzaminowaniu zewnętrznemu znaleźć można w wynikach badań Urszuli Opłockiej (2008). Do najbardziej powszechnych należą: 1) procentowe obliczanie wartości zadań, traktowane w WSO jak norma określająca wysokość stopnia; 2) uznanie 30% progu zaliczenia na egzaminie maturalnym za wartość wyznaczającą minimum wymagań; 3) stosowanie procedur egzaminacyjnych (modele odpowiedzi, schematy, klucze) do oceny zadań rozwiązywanych w szkole.

Zarazem jednak nauczyciele przyznają się do manipulowania punktami i kryteriami, by odpowiadały one intuicyjnej ocenie sprawdzanej przez nich pracy, zdarza się im też traktować kryteria oceny jak ochronny parasol, a nie trwały element oceniania (tamże). Oznacza to, że **przychodzące z zewnątrz impulsy do zmiany podlegają redukcji w nauczycielskich systemach oceniania, ale ma ona wymiar nieformalny i jednostkowy**. Potwierdzają to empiryczne ustalenia poczynione przez Grażynę Szyling (2011) ujawniające, że jednym ze sposobów, w jaki nauczyciele radzą sobie ze sprzecznymi oczekiwaniami stron zaangażowanych w proces oceniania, jest okazjonalne posługiwanie się strategiami przetrwania.

Co ważne, wyniki badań nad nauczycielskimi praktykami oceniania, niezależnie od przyjętej orientacji badawczej, pokazują bardzo zbliżony ich obraz, który w przypadku niniejszych rozważań odnieść można do kategorii systemu[1]. Piotr W. Parczewski (2012) odczytuje zgromadzone przez siebie dane za pomocą teorii systemu autopojetycznego, czyli takiego, który rekonstruuje się w reakcji na bodźce zmieniającego się środowiska (Luhmann 2007). Autor stawia tezę, że „ze względu na wysokie skomplikowanie relacji pomiędzy systemem a jego środowiskiem oraz wysoką złożoność relacji wewnątrzsystemowych, próby zewnętrznego sterowania systemem są z góry skazane na niepowodzenie" (Parczewski 2012: 262). Spostrzeżenie to zachowuje zbieżność z interpretacjami U. Opłockiej (2008, 2010), która podporządkowała ogląd praktyk nauczycielskiego oceniania kategorii emergencji, ujawniając, że tworzą one samoorganizującą się złożoną strukturę, nieredukowalną do prostych elementów składowych. Co więcej, autorka twierdzi, że nie można przewidzieć, jak owa indywidualna struktura zareaguje na zmiany przychodzące z zewnątrz.

Przeprowadzone analizy uzasadniają, że mniej lub bardziej świadome ograniczanie w WSO roli, jaką w realiach edukacyjnych odgrywają indywidualne teorie pedagogiczne nauczycieli osadzone w odmiennych często przekonaniach, doprowadza w najlepszym przypadku do **rozdźwięku między formalnymi zapisami a ich wymiarem praktycznym, czyli nauczycielskim działaniem**. Jeśli przyjmiemy dalej, że to działania, a nie myśli i intencje, są głównym przedmiotem regulacji społecznej (Sztompka 2002: 258), możemy do odczytania specyfiki szkolnego oceniania zastosować klasyczną kategorię systemu aksjo-normatywnego. Piotr Sztompka (2002: 278) definiuje go jako „powiązany zespół reguł

[1] Odwołanie do kategorii systemu społecznego ma w moich rozważaniach wyłącznie charakter interpretacyjny: pozwala ukazać niejednorodność i złożoność systemów oceniania jako przedmiotu ewaluacji.

– norm i wartości – [...] charakterystyczny dla danej kultury", przy czym „wartości wskazują, do czego ludzie powinni dążyć, a normy – jak powinni do tego dążyć".

W tak zarysowanym kontekście formalne zapisy WSO współtworzą kulturę oceniania szkolnego. Regulując i ujednolicając sposób postępowania nauczycieli i uczniów oraz ich rodziców w określonych sytuacjach związanych ze szkolnym ocenianiem, stają się normami, za którymi stoją wartości, niekoniecznie zdefiniowane wprost. W potocznym odbiorze sprowadzają się one najczęściej do dwóch dyskursywnych kategorii: sprawiedliwości i obiektywizmu (zob. Szyling 2013b). Tendencja do upraszczania złożonego charakteru obu tych wartości przejawia się w WSO nadmiernym eksponowaniem obiektywizmu, wypierającego sprawiedliwość, która nie poddaje się zabiegom operacjonalizacji. Z kolei sam obiektywizm ograniczony zostaje do aspektów technicznych oceniania oraz procedur zapewniania jego bezstronności, stwarzając wrażenie neutralnego aksjologicznie zabezpieczenia praw nauczycieli, uczniów i ich rodziców. Działania te nie są jednak w stanie wyrugować wartości z oceniania szkolnego a także nie uwalniają nauczycieli od konieczności podejmowania pod wpływem kontekstu kształcenia trudnych i niejednoznacznych decyzji (tamże), co prowadzi do przeciwstawienia sobie dwóch potrzeb: globalnych regulacji oraz indywidualizacji procesu kształcenia i jego wyników. Wytwarzające się w tym obszarze napięcia można powiązać z subiektywnym stosunkiem członków danej społeczności do reguł kulturowych, który bywa niezależny od obiektywnego ich przestrzegania (Sztompka 2002: 283–284). W efekcie u podmiotów zaangażowanych w ocenianie pojawiać się mogą odmienne postawy, lokujące się w przestrzeni wyznaczanej konformizmem, legalizmem lub oportunizmem z jednej strony, a różnymi odmianami buntu – z drugiej.

Zatem **skupienie się na szczegółowym regulowaniu przez WSO działań związanych z ocenianiem, przy równoczesnym pomijaniu stających za nimi wartości, jest kolejnym przejawem instrumentalnej redukcji systemów oceniania**. Dokonany z tak zawężonej perspektywy ogląd szkolnej rzeczywistości oceniania pozostaje zamknięty na jej dyskursywny charakter, wytwarzany przez wartości ukryte za jednostkowymi i grupowymi działaniami. Wartości te mogą wchodzić w konflikt z rozwiązaniami przyjmowanymi z zewnątrz (tamże: 278), nawet jeśli te drugie wydają się neutralne aksjologicznie. **Skutkuje to inercją szkolnej kultury oceniania lub spowalnia zmianę**, zwłaszcza gdy modyfikacji lub rekonstrukcji systemu aksjo-normatywnego nie towarzyszą podejmowane w danej społeczności negocjacje i uzgodnienia dotyczące jego poszczególnych elementów.

Takiego dyskursu zabrakło w procesie wdrażania WSO i brakuje nadal. Dokonany dotychczas przegląd różnych aspektów zmian zachodzących w tym obszarze wskazuje bowiem, że zasadnicze impulsy do modyfikacji szkolnych systemów oceniania są ulokowanie na zewnątrz placówki i docierają do niej jako nakazy lub obowiązki sformułowane wprost przez organy władzy oświatowej różnego szczebla, bądź pośrednio – przez odniesienie do jakiegoś wzoru funkcjonującego w odmiennych realiach, ale stwarzającego złudzenie administracyjnego bezpieczeństwa. Bardziej innowacyjne, ale zarazem trudne do przewidzenia i zewnętrznego rozpoznania są **stosowane przez konkretnych nauczycieli strategie radzenia sobie z włączaniem cudzych propozycji w zakres własnych działań. Mogą one jednak nie tylko prowadzić do zmiany, ale ją hamować, dlatego ich rozpoznanie jest znaczące dla myślenia o ewaluacji WSO**.

Jaka ewaluacja? Jaka „miara jakości"?

Podjęta w poprzedniej części tekstu analiza właściwości szkolnego systemu oceniania, traktowanego jak element kultury szkoły zmieniającej się pod wpływem bodźców zewnętrznych i wewnętrznych działań, ujawnia kilka aspektów ważnych dla myślenia o ewaluacji WSO i wiązanych z nią znaczeniach. Przede wszystkim warto uświadomić sobie, że zmiana zachodzi – czasami nawet szybciej i bardziej efektywnie – niekoniecznie w wyniku ewaluacji działania czy pogłębionej refleksji nad praktyką. Jej źródłem może być nie zawsze przemyślana, a często pozorna reakcja na wymagania zewnętrzne lub napięcia występujące między oficjalnymi regulacjami a realnym kontekstem kształcenia. Taka modyfikacja ma z reguły wymiar nieformalny i wymyka się uogólniającym badaniom zewnętrznym. Niemniej jej bliskie i odległe w czasie, korzystne i szkodliwe konsekwencje są odczuwalne dla społeczności objętej danym działaniem i możliwe do empirycznego ustalenia po spełnieniu określonych wymogów metodologicznych.

Tymczasem analiza raportów ewaluacyjnych w zakresie dotyczącym WSO skłania do sformułowania wniosku, że **badania te najlepiej radzą sobie z redukowaniem nadmiaru, zwłaszcza w obszarze wartości**. Na poparcie tej tezy przytoczę kilka przykładów ilustrujących uproszczenia ewaluacyjne charakterystyczne dla zagadnień oceniania szkolnego.

W ewaluacji zewnętrznej, prowadzonej w ramach sprawowania nadzoru pedagogicznego (Rozp. MEN 2013), wątki związane z ocenianiem są rozproszone

w różnych obszarach wymagań. Koncepcja taka znajduje swoje uzasadnienie w powszechnym przekonaniu, że ocenianie szkolne jest integralną częścią procesu kształcenia, więc nie ma szczególnych przesłanek do wydzielania go jako odrębnego zakresu wymagań państwa wobec szkół. Z drugiej jednak strony przywoływany dokument w wymaganiu nr 11 podkreśla znaczenie, jakie w organizowaniu „procesów edukacyjnych" odgrywa uwzględnianie przez szkołę wniosków z analizy wyników egzaminów zewnętrznych (oraz innych badań, w tym ewaluacji wewnętrznej), co sugeruje obowiązującą hierarchię ważności między ocenianiem szkolnym i egzaminacyjnym, po raz kolejny potwierdzając wnioski z poprzednich rozważań. W kontekście tym zasadne staje się postawienia pytania: **jaki rodzaj wiedzy o swoim systemie oceniania może otrzymać szkoła i nauczyciele dzięki wynikom zewnętrznego badania ewaluacyjnego?**

Przyjmuję dalej – za Howardem Schumanem (2013) – że dane uzyskane w badaniach są silnie zależne od metody i jakości samych narzędzi. Z tej perspektywy zanalizuję dwa celowo dobrane[2] pytania dotyczące wymagania nr 2: „Procesy edukacyjne są zorganizowane w sposób sprzyjający uczeniu się" (http://www.npseo.pl/). Oba mają charakter zamknięty, czyli zakładają, że opcje odpowiedzi wyczerpują całe pole znaczeniowe badanej zmiennej i są na tyle precyzyjne, że ich wyniki dadzą się jednoznacznie odczytać dzięki osadzeniu w zoperacjonalizowanej definicji pojęcia lub przez odniesienie do wartości zawartej w kryterium ewaluacji.

Pierwsze z wybranych pytań jest powiązane z obserwacją zajęć. Ewaluator po jej zakończeniu udziela między innymi odpowiedzi: „Czy uczniowie mają możliwość wykorzystania popełnionych błędów do uczenia się?", korzystając przy tym z następujących opcji do wyboru: „w każdej sytuacji; w większości sytuacji; w niektórych sytuacjach; błędy uczniów nie były wykorzystane w procesie uczenia; nie wystąpiły sytuacje, w których uczniowie popełniali błędy".

Oszacowanie przydatności tego pytania do wnioskowania o szkolnym ocenianiu wymaga przede wszystkim sproblematyzowania znaczenia wykorzystanych w nim pojęć. Błąd może bowiem być – zależnie od bliskiej nauczycielowi koncepcji nauczania – albo czymś nagannym, co należy natychmiast poprawić zgodnie z uznanym wzorem, albo pożądaną informacją o procesie uczenia się ucznia, która wymaga wspólnego namysłu nad kolejnymi krokami, albo jakimś działaniem lokującym się „pomiędzy" tymi sytuacjami. We wszystkich tych przypadkach uczeń może mieć „możliwość wykorzystania popełnionego błędu do

[2] Nie zamierzam dokonywać w tym miejscu szczegółowej merytorycznej i metodologicznej analizy koncepcji ewaluacji zewnętrznej. Interesują mnie tylko wybrane znaczenia, jakie – niekoniecznie świadomie – nadaje ona ocenianiu szkolnemu.

uczenia się", w każdym z nich jednak przebieg tego działania i jego efekty są odmienne, a wartości za nimi stojące rozciągają się od budowania poczucia sprawstwa u uczniów po wytwarzanie u nich bezradności intelektualnej.

Zasygnalizowana redukcja znaczeń wpisuje się w dosyć powszechne przekonanie o ateoretyczności badań praktycznych, czyli także ewaluacji. Problemem jest jednak, że owa **skłonność do potoczności pogłębia i tak nieprzejrzystą codzienność szkolnego oceniania, ponieważ sprowadza ją do prostego schematu** względnie do niej przystającego. Krzysztof Konarzewski (2000) zwraca w tym kontekście uwagę, że brak jasnego teoretycznego osadzenia wartości, do których odwołują się badania praktyczne, oznacza podporządkowanie ich interesom określonej grupy i wpisanie w relacje władzy.

Ten brak zaplecza pedagogicznego i psychologicznego jest wyraźnie widoczny także w sformułowaniu skierowanego do uczniów pytania ankiety ewaluacyjnej: „Jak się zazwyczaj czujesz, kiedy jesteś oceniana/y? Wybierz opisy, które najlepiej pasują do Ciebie: „mam ochotę się uczyć"; „postanawiam, że się poprawię"; „nie chce mi się uczyć dalej"; „nie wiem, co dalej robić"; „wiem, co mam poprawić".

Autorzy pytania przyjmują, że każdy uczeń styka się nie tylko z określoną, ale i powtarzalną sytuacją oceniania, na którą zazwyczaj reaguje zgodnie z jakimś przyjętym wzorem, nigdy jednak nie: radością lub uczuciem ulgi ani złością, żalem lub agresją. Natomiast badania psychologów udowadniają, że wieloaspektowe i złożone relacje między informacją zwrotną o wyniku zadania a dalszym działaniem podmiotu wymykają się modelowym ujęciom i jednoznacznym rozstrzygnięciom (zob. Łukaszewski 2002). Ponadto analizowane pytanie tylko pozornie dotyczy faktów, ponieważ w gruncie rzeczy zawiera wyraźny komponent opinii (por. Schuman 2013). Ta zaś może ulegać zmianie zależnie nie tylko od sytuacji oceniania i wysokości otrzymanego stopnia, ale też od konkretnego przedmiotu i działań nauczyciela, oraz zakresu zbieżności indywidualnych celów ucznia z celami szkoły (zob. np. Brophy 2002).

Zasygnalizowana powyżej **wrażliwość wyników badań ilościowych na indywidualne wartości i osobistą sytuację badanych jest pomijana w ewaluacji zewnętrznej, choć to one w znacznej mierze decydują o społecznym kontekście, w jakim przebiega kształcenie** w klasie i w szkole, czyli wpływają na szeroko rozumianą jakość edukacji. Tego typu redukcji nie zastąpi metodologiczna triangulacja oraz procedury prostego losowego doboru próby (zob. *Model*... 2013), dyskusyjnego przy tak małych i wewnętrznie zróżnicowanych populacjach jak szkoła (por. Patton 2013). Zasadność tego wniosku w odniesieniu do ewaluacji

obszaru szkolnego oceniania jest wyraźnie widoczna w przypadku relatywizacji reguł bezstronności i obiektywizmu, czego przykładem są rozwiązania dotyczące szkolnego funkcjonowania uczniów o specjalnych potrzebach edukacyjnych. Badania wskazują (Brookhart 2004), że zakres akceptacji dla procedur – obowiązujących lub praktycznie stosowanych w ocenianiu tej grupy – dzieli zarówno uczniów jak i nauczycieli, przy czym znaczna część zróżnicowania opinii tej pierwszej grupy wykazuje zależność od statusu wyznaczanego wysokością osiągnięć szkolnych.

Dokonane analizy pozwalają na sformułowanie wniosku, że przydatność wyników ewaluacji zewnętrznej do refleksyjnego namysłu nad szkolnym ocenianiem zdaje się minimalna. **Do rangi poważnego mankamentu poznawczego urasta bowiem schematyzm szkolnej rzeczywistości prezentowanej w raportach**, czego nie można wytłumaczyć wyłącznie systemową rolą ewaluacji zewnętrznej (por. Korporowicz 2011, 2013). Ten uproszczony i podporządkowany zewnętrznym procedurom obraz **nie tyle blokuje zmianę, ile nie pozwala dostrzec jej potrzeby, o ile ta nie wpisuje się w konieczność zaspokojenia zewnętrznych „standardowych" oczekiwań**.

Zjawisko to obecne jest także w wewnętrznej ewaluacji WSO i to niezależnie od stwarzanej przez bliską perspektywę badawczą możliwości upodmiotowienia samego procesu ewaluacyjnego, jak i łatwiejszego identyfikowania uwarunkowań, napięć i wartości występujących w codziennych praktykach oceniania danej placówki. Decydujące znaczenie zdaje się mieć w tym przypadku – na co zwraca uwagę Michael Q. Patton (2013) – doświadczenie nauczycieli-ewaluatorów, kształtowane przede wszystkim przez procedury ewaluacji zewnętrznej, nastawionej na kontrolę pracy szkoły i pozostawiającej poza dyskusją sprawę kryteriów.

Argumentów potwierdzających to spostrzeżenie dostarczają wnioski z analizy 10 losowo wybranych raportów z ewaluacji wewnętrznej WSO, dostępnych w sieci. Dokumenty te przypominają strukturą zminiaturyzowaną kopię raportów, jakie otrzymują szkoły z ewaluacji zewnętrznej, ale znacznie więcej w nich różnych niedostatków metodologicznych i analitycznych. Usterki te można byłoby uznać za drugorzędne, gdyby wyniki ewaluacji były punktem wyjścia do dialogu, a nie podstawą formułowania rekomendacji.

Problematyczność owych zaleceń do wdrożenia ma kilka przyczyn, z których najpowszechniejszą jest brak jasno zdefiniowanych kryteriów ewaluacji albo zbyt ogólnikowe sformułowania np. skuteczność, efektywność, użyteczność (zob. np. Ciężka 2010). Mogłoby to oznaczać, że wystarczającym uprawomocnieniem wartości, do których szkoła odwołuje się w procesie ewaluacji, jest ich zbieżność z kryteriami zewnętrznymi. Jednak niezrozumienie tych wartości

albo ich potoczne odczytanie sprawia, że w **wielu wewnętrznych ewaluacjach WSO kryterium jakości staje się sam przedmiot badań**. W efekcie **lokalne regulacje**, o których mankamentach wspominałam uprzednio, **urastają do rangi niepodważalnego punktu odniesienia dla szkolnych realiów, które – jak zajdzie taka potrzeba – należy zmienić, by przystawały do procedur sformułowanych w dokumencie**. I tak odbiorcy ewaluacji WSO najczęściej dowiadują się, że: „powinni wzmocnić częstotliwość odwoływania się w praktyce do WSO"; „przypominać uczniom i rodzicom w ciągu roku szkolnego zasady WSO"; „jednomyślnie interpretować zawarte w WSO ustalenia"; „wypracować formę sprawdzenia, czy informacja o systemie oceniania została przez wszystkich uczniów zrozumiana i przyjęta do wiadomości"; „bezwzględnie przestrzegać systematyczności oceniania oraz zapisywania ocen w dzienniku"; „oceniać za aktywność na zajęciach edukacyjnych"[3].

Swoista „przezroczystość" aksjologiczna przywołanych rekomendacji oznacza, że ocenianie szkolne jest postrzegane jak zbiór procedur, których głębszy sens nikogo podczas ewaluacji WSO w szkole nie zastanawia. Po raz kolejny potwierdzona zostaje postawiona wcześniej teza o **niedostrzeganiu potrzeby zmian w edukacji albo jej blokowaniu przez rzeczywiste lub tylko oportunistyczne przywiązanie szkoły do zewnętrznych wobec niej, często wyłącznie biurokratycznych regulacji lub schematów wyrosłych z potoczności pedagogicznej**.

Jakie impulsy? Jakie wartości?

W świetle podjętych w tym tekście analiz zdaje się nie budzić wątpliwości, że dla ewaluacji WSO **kluczowy staje się namysł**, co w dokumencie tym jest niezbędnym zapisem, a co tylko naśladownictwem lub modą, co poprawia przejrzystość codziennego szkolnego oceniania, a co wpycha je w gorset zbędnych procedur, co ułatwia komunikację między oceniającymi i ocenianymi, a co stawia ich po dwóch stronach edukacyjnej barykady. **Rozstrzygnięcie tych wątpliwości nie jest możliwe bez ustalenia wspólnych wartości**, które w danej szkole kryją się za nauczycielskimi praktykami oceniania, uczniowskimi strategiami radzenia sobie z ocenianiem i oczekiwaniami rodziców formułowanymi wobec własnych

[3] Dosłowne przytoczenia rekomendacji (zaleceń) znajdujących się w szkolnych raportach z ewaluacji WSO dostępnych w sieci. Ich typowość sprawia, że wskazywanie źródła raportów niosłoby ze sobą niepotrzebne naznaczanie.

dzieci i ich nauczycieli. Zadanie nie jest łatwe, ponieważ miejsce dotychczasowej pewności zastąpią pytania, które mogą sproblematyzować postrzeganie szkolnego oceniania, jeśli postawią je sobie wszystkie podmioty zaangażowane w ten proces. Wspólnym ich trzonem powinna stać się refleksja nad elementarnym zagadnieniem: Po co i komu potrzebne jest ocenianie w szkole?, **refleksja wykraczająca poza skostniałe w potocznym odbiorze i zinstrumentalizowane przez edukację kategorie sprawiedliwości i obiektywizmu**.

Jedną z wartości, która pozwala zobaczyć WSO z innej perspektywy, jest **zaufanie**. Anthony Giddens (2007: 404–405) podkreśla, że systemy małego zaufania (*low-trust systems*) – a takimi są WSO – charakterystyczne dla organizacji i miejsc pracy, w których jednostkom pozostawia się mały zakres kompetencji i kontroli nad tym, co robią, nie przystają do warunków ciągle zmieniającej się rzeczywistości. Ta bowiem wymaga budowania systemów dużego zaufania (*high-trust systems*), zapewniających jednostkom dużo autonomii i samokontroli.

Autonomia, **samokontrola** i **zmienność** – wiązane przez A. Giddensa z „dużym" zaufaniem – stanowią też klucz do odczytania metafory, jaką w swoich rozważaniach o edukacji posłużył się Zygmunt Bauman (2012). Ze znanej tezy, że współczesny świat wymaga ciągłego zbierania i przetwarzania danych, by sprostać nieustannym przewartościowaniom celów, które człowiek chciałby osiągnąć, wysnuł wniosek, że uczniowie przygotowani przez edukację do zmierzania zaprogramowaną drogą do jasno wytyczonego celu, przypominają pociski balistyczne, wypierane na nowoczesnym polu walki przez „pociski uczące się w locie" (tamże: 24). Ta druga grupa uczniów nie tylko potrafi się szybko uczyć, ale też posiada zdolność rozpoznawania zdezaktualizowanej wiedzy, nieprzydatnej do nowych celów, o której po prostu trzeba zapomnieć. Kierują się przy tym „własną oceną: co maksymalnie mogą osiągnąć przy swoich technicznych możliwościach [...]" (tamże: 25), a cele wytyczają podczas działania.

Opozycje występujące w obu koncepcjach unaoczniają, że **każda z opcji do wyboru ma swój horyzont znaczeń, w którego obszarze inaczej odczytywane będą takie wartości związane z ocenianiem**, jak: systematyczność, pilność, posłuszeństwo, autonomia poznawcza, umiejętność radzenia sobie z trudnościami i niepowodzeniami, samokontrola, poczucie sprawstwa, wyrównywanie szans. Pamiętanie o tym **może uchronić ewaluację systemów oceniania przed redukowaniem świata wartości do pozornie neutralnych aksjologicznie procedur**. Pozostaje jednak pytanie, który z horyzontów wybrać? I to odpowiedź na nie decydować będzie o kierunku ewaluacji „zapraszającej do rozwoju" szkolnego oceniania.

Bibliografia

Bauman Z. 2012. *O edukacji. Rozmowy z Riccardo Mazzeo*, tłum. P. Poniatowska, Wydawnictwo Naukowe Dolnośląskiej Szkoły Wyższej, Wrocław.

Brookhart S. 2004. *Grading*, Pearson Education, New Jersey.

Brophy J. 2002. *Motywowanie uczniów do nauki*, tłum. K. Kruszewski, Wydawnictwo Naukowe PWN, Warszawa.

Bruner J. 2006. *Kultura edukacji*, tłum. T. Brzostowska-Tereszkiewicz, Universitas, Kraków.

Ciężka B. 2010. *Planowanie ewaluacji wewnętrznej w szkole*, [w:] *Ewaluacja w nadzorze pedagogicznym. Autonomia*, red. G. Mazurkiewicz, Wydawnictwo Uniwersytetu Jagiellońskiego, Kraków, s. 83–108.

Giddens A. 2007. *Socjologia*, tłum. A. Sulżycka, Wydawnictwo Naukowe PWN, Warszawa.

Groenwald M. 2012. *Co na płaszczyźnie diagnostyki edukacyjnej łączy Henryka Szaleńca z Bruno Latourem i co na to Bolesław Niemierko?*, [w:] *Regionalne i lokalne diagnozy edukacyjne*, red. B. Niemierko, M.K. Szmigel, PTDE, Kraków.

Konarzewski K. 2002. *Jak uprawiać badania oświatowe? Metodologia praktyczna*, WSiP, Warszawa.

Korporowicz L. 2010. *Interakcyjna misja ewaluacji*, [w:] *Ewaluacja w nadzorze pedagogicznym. Konteksty*, red. G. Mazurkiewicz, Wydawnictwo Uniwersytetu Jagiellońskiego, Kraków, s. 29–39.

Korporowicz L. 2011. *Ewaluacja jako animacja. W kierunku ewaluacji piątej generacji*, [w:] *Ewaluacja w nadzorze pedagogicznym. Refleksje*, red. G. Mazurkiewicz, Wydawnictwo Uniwersytetu Jagiellońskiego, Kraków, s. 35–53.

Korporowicz L. 2013. *Społeczna etyka ewaluacji*, „Teraźniejszość – Człowiek – Edukacja", nr 4 (64), s. 35–43.

Luhmann N. 2007. *Systemy społeczne*, „Nomos", Kraków.

Łukaszewski W. 2002. *Zwrotne informacje o wyniku czynności*, [w:] *Psychologia czynności. Nowe perspektywy*, red, I. Kurcz, D. Kądzielawa, Wydawnictwo Naukowe SCHOLAR, Warszawa.

Marshall G. red. 2005. *Słownik socjologii i nauk społecznych*, red. nauk. pol. wyd. M. Tabin, Wydawnictwo Naukowe PWN, Warszawa.

Model i zalecany przebieg ewaluacji zewnętrznej szkół oraz placówek obowiązujący od dnia 1 września 2013 r. ze zmianami z dnia 31 października 2014 r., http://www.npseo.pl/action/download/2/1 (30.04.2015).

Niemierko B. 2000. *Kształcenie według wymagań*, [w:] *Pedagogika w pokoju nauczycielskim*, red. K. Kruszewski, WSiP, Warszawa, s. 74–91.

Niemierko B. 2001. *Chłodne oblicze egzaminu zewnętrznego*, [w:] *Teoria i praktyka oceniania zewnętrznego*, red. B. Niemierko, M.K. Szmigel, Wydawnictwo Pandit, Kraków.

Niemierko B. 2013. *Czy diagnostyka edukacyjna i ewaluacja w edukacji zachowają odrębność?*, „Teraźniejszość – Człowiek – Edukacja", nr 4 (64), s. 7–19.

Opłocka U. 2008. *Emergencja nauczycielskiej pedagogii oceniania – doniesienie badawcze*, [w:] *Uczenie się i egzamin w oczach nauczyciela*, red. B. Niemierko, M.K. Szmigel, PTDE, Kraków, s. 154–161.

Opłocka U. 2010. *Belfer zniewolony*, [w:] *Teraźniejszość i przyszłość oceniania szkolnego*, red. B. Niemierko, M.K. Szmigel, PTDE, Kraków, s. 509–517.

Parczewski P.W. 2012. *Nauczycielskie systemy oceniania. Trafność oceniania na lekcjach języka angielskiego*, Wydawnictwo Adam Marszałek, Toruń.

Patton M.Q. 2013. *Przyszłe trendy w ewaluacji*, [w:] *(R)ewaluacja. Poszukiwanie nowych metod oceny efektów*, red. A. Haber, Z. Popis, tłum. GTC AMG, Polska Agencja Rozwoju Przedsiębiorczości, Warszawa, s. 35–45.

Putkiewicz E. 2000. *O ocenianiu szkolnym i ocenianiu nauczania*, [w:] *Pedagogika w pokoju nauczycielskim*, red. K. Kruszewski, WSiP, Warszawa, s. 294–306.

Rozporządzenie Ministra Edukacji Narodowej z dnia 19 kwietnia 1999 roku w sprawie zasad oceniania, klasyfikowania i promowania uczniów i słuchaczy oraz przeprowadzania egzaminów i sprawdzianów w szkołach publicznych (Dz.U. z 1999 r. Nr 41, poz. 413).

Rozporządzenie Ministra Edukacji Narodowej z dnia 10 maja 2013 roku zmieniające rozporządzenie w sprawie nadzoru pedagogicznego (Dz.U. z 14 maja 2013 r., poz. 560).

Rutkowiak J. 1995. *Pulsujące kategorie jako wyznaczniki mapy odmian myślenia o edukacji*, [w:] *Odmiany myślenia o edukacji*, red. J. Rutkowiak, Oficyna Wydawnicza Impuls, Kraków 1995.

Schuman H. 2013. *Metoda i znaczenie w badaniach sondażowych*, tłum. A. Grabowska, Oficyna Naukowa, Warszawa.

Sztompka P. 2002. *Socjologia. Analiza społeczeństwa*, Wydawnictwo Znak, Kraków.

Szyling G. 2011. *Nauczycielskie praktyki oceniania poza standardami*, Oficyna Wydawnicza Impuls, Kraków.

Szyling G. 2013a. *Diagnoza i ewaluacja a teoretyczne horyzonty refleksji nad działaniem*, „Teraźniejszość – Człowiek – Edukacja", nr 4 (64), s. 55–70.

Szyling G. 2013b. *Obiektywizm i sprawiedliwość – (nie)wygodne kategorie szkolnego oceniania*, [w:] *Sprawcy i/lub ofiary działań pozornych w edukacji szkolnej*, red. M. Dudzikowa, K. Knasiecka-Falbierska, Oficyna Wydawnicza Impuls, Kraków, s. 255–272.

Szyling G. 2014. *Teoria i praktyka oceniania szkolnego: w poszukiwaniu pedagogicznych przesłanek hybrydyczności oceny szkolnej*, „Ruch Pedagogiczny", nr 2, s. 67–82.

Ustawa z dnia 20 lutego 2015 roku o zmianie ustawy o systemie oświaty oraz niektórych innych ustaw (Dz.U. z dnia 16 marca 2015 r., poz. 357).

KWARTALNIK PEDAGOGICZNY
2015 NUMER 4(238)

Piotr Maciej Skorupiński
III Liceum Ogólnokształcące im. gen. Józefa Sowińskiego w Warszawie

Standardy trafności pomiaru

Summary

STANDARDS OF MEASUREMENT VALIDITY

According to the latest edition of the *Standards for Educational and Psychological Testing* (AERA et al. 2014), validity is still the most fundamental construct in educational measurement. The new edition, like the previous versions of the *Testing Standards*, contains standards of professional practice in validation. These standards of validity will be presented in the fourth section of this article. The first three sections explore validity standards included in the *Testing Standards* published before the 2014 edition, in the period of the shift from the trinitarian doctrine validity to the unitarian view of validity. The closing remarks consider the issue of the role of consequences in the most recent validity recommendations.

Key words: validity; validation; AERA, APA, NCME *Standards*.

red. Paulina Marchlik

Idea niezależnego i profesjonalnego monitorowania użytkowania testów ma już ponad 100-letnią historię. Jedną z form jej konkretyzacji stanowią próby ustanowienia standardów w dziedzinie stosowanego pomiaru edukacyjnego (Madaus 2001: 2).

Zarówno w pionierskich *Technicznych rekomendacjach dla testów osiągnięć*[1] (AERA i in. 1955), jak i w kolejnych pięciu edycjach standardów z lat 1966–2014, przygotowanych i popularyzowanych przez Amerykańskie Towarzystwo Badań Edukacyjnych (*American Educational Research Association* – AERA), Narodową Radę do spraw Pomiarów Edukacyjnych (*National Council on Measurement in Education*

[1] Cytaty z publikacji obcojęzycznych w tłumaczeniu autora.

– NCME) oraz Amerykańskie Towarzystwo Psychologiczne (*American Psychological Association* – APA), przed standardami rzetelności zarekomendowane zostały standardy trafności pomiaru. Rekomendacjom tym towarzyszyły wykładnie pojęcia trafności, w których wykorzystano dwa jej modele.

Techniczne rekomendacje współotwierały epokę dominacji trynitarnego modelu trafności, obejmującego trzy zasadnicze typy trafności: kryterialną, treściową oraz teoretyczną. Ostatnie trzy wydania *Standardów edukacyjnego i psychologicznego testowania* upowszechniają natomiast model unitarny, w którym typologię trynitarną zastąpiono zunifikowanym i poszerzonym pojęciem trafności, definiowanym jako sąd oceniający, w jakim stopniu dane empiryczne i konstrukcje teoretyczne upoważniają do inferencji i działań opartych na wynikach pomiaru edukacyjnego (Messick 1989).

W kolejnych częściach niniejszego artykułu te dwa modele trafności zostaną szerzej scharakteryzowane. Szczegółowo przedstawione będą także standardy trafności, rozumiane jako wzorce dobrej praktyki walidacyjnej, rekomendowane przez wspomniane wyżej towarzystwa w ciągu ostatnich sześciu dziesięcioleci. W finalnej części zasygnalizowana jest obiekcja odnosząca się do zasadniczej zmiany w pojęciu trafności, wprowadzonej w najnowszej edycji *Standardów* z 2014 roku.

Pierwsze rekomendacje

Techniczne rekomendacje dla testów osiągnięć zostały przygotowane na podstawie *Technicznych rekomendacji dotyczących testów psychologicznych i technik diagnostycznych* (APA i in. 1954). Adresowane były zarówno do konstruktorów oraz wydawców, jak i do użytkowników testów osiągnięć – jako „standardy profesjonalizmu w praktyce" i „standardy aprobowanej praktyki" (AERA i in. 1955: 3, 5) – z myślą, iż poważna refleksja nad użytecznością danego testu oraz znaczeniem jego wyników wymaga informacji, które powinny zostać najpierw zgromadzone, a następnie udostępnione w materiałach towarzyszących testowi, czyli w tzw. podręczniku jego użytkownika.

Trzydziestosześciostronicowe *Techniczne rekomendacje dla testów osiągnięć* prezentują wspomniane standardy w sześciu działach poświęconych rozpowszechnianiu informacji o teście, interpretacji jego wyników, trafności, rzetelności, przeprowadzaniu testu, a także skalom i normom. Rozdział o trafności zawiera 42 standardy trafności oraz eksplikację trynitarnego jej modelu. W jej świetle,

celem zastosowania testu jest wyrobienie osądu, zaś stopień, w jakim test umożliwia osiągnięcie tego celu, jest przedmiotem informacji o trafności. Informacja ta – w formie dowodów walidacyjnych – może dotyczyć trafności treściowej, diagnostycznej, prognostycznej lub teoretycznej, w zależności od typu osądu, do którego zmierza testowanie. Jak to zostało ujęte: „Dla każdego typu osądu wymagany jest nieco inny typ dowodu walidacyjnego" (AERA i in. 1955: 15).

Standardy trafności również pogrupowano w kilku sekcjach poświęconych poszczególnym typom trafności oraz zagadnieniom ogólnym. W sekcji ogólnej zapisano, iż podręcznik, który należy dołączyć do każdego przeznaczonego do użytkowania testu, „powinien informować o trafności każdego typu inferencji, dla której dany test jest rekomendowany" (standard C 1), „powinien klarownie określać typ trafności, o którym mowa" (C 2) i „wskazywać, które [...] spośród interpretacji wyników testu [...] nie zostały poparte dowodami" walidacyjnymi (C 3).

W sekcji *Trafność treściowa* zawarto rekomendację, iż „jeśli rozwiązywanie testu będzie interpretowane jako próba czynności z pewnego uniwersum sytuacji, podręcznik powinien wyraźnie wskazywać, jakie uniwersum jest reprezentowane oraz jak adekwatny jest dobór próby" (C 4). W związku z powyższym podkreślono konieczność precyzyjnego zdefiniowania wspomnianego uniwersum treści (C 4.1) oraz opisu metody doboru zadań (C 4.2), a także wskazano, iż „jeśli eksperci proszeni są o osąd, czy zadania są odpowiednią próbą z danego uniwersum albo czy są prawidłowo punktowane, należy przedstawić liczbę ekspertów, fakty dotyczące ich odpowiedniego doświadczenia zawodowego i kwalifikacji, szczegóły procesu oceniania oraz dane dotyczące niezależnej zgodności między nimi w ocenie" (C 4.4).

W sekcjach poświęconych trafności diagnostycznej i prognostycznej zalecono zamieszczenie w podręczniku użytkownika testu stosownego opisu wykorzystanych miar kryterialnych, zawierającego także ocenę adekwatności użytej miary, wskazującej te aspekty mierzonej czynności, których nie odzwierciedla dana miara, oraz czynniki irrelewantne, które znalazły odbicie w wynikach (C 6). Opis taki powinien również informować o rzetelności miar kryterialnych (C 7) oraz wyczerpująco charakteryzować próby walidacyjne, pod wskazanymi względami porównywalne do grup poddawanych testowaniu (C 12). W tym zakresie rekomendowane jest przedstawienie „odpowiednich miar tendencji centralnej i wariancji wyników testu w próbie walidacyjnej" (C 12.1), oraz korelacji między wynikami testu i miar kryterialnych (C 6.1).

Rozdział *Trafność teoretyczna* zawiera 5 standardów poprzedzonych uwagami dotyczącymi roli tego typu trafności w pomiarze edukacyjnym. W świetle tych

uwag, „trafność teoretyczna jest szczególnie ważna w przypadkach, które zazwyczaj dominują w dziedzinie testów osiągnięć edukacyjnych, kiedy nieosiągalne jest odpowiednie niezależne kryterium. Walidacja testów osiągnięć poprzez zademonstrowanie ich mocy różnicowania pomiędzy wybranymi kryterialnymi grupami odniesienia jest w efekcie trafnością teoretyczną [...]. Dane wskazujące na moc różnicującą w koniunkcji z trafnością treści są wystarczające dla większości testów osiągnięć, zorganizowanych wokół materii przedmiotowej. Pozostałe rekomendacje [...] są znaczące dla nowszego typu testów osiągnięć, które usiłują mierzyć takie konstrukty, jak rozumienia, postawy, zainteresowania, zamiłowania i nawyki lub umiejętności uczenia się. Te [...] nie reprezentują identyfikowalnego uniwersum treści [...]" (AERA i in. 1955: 26). Podręcznik użytkownika testu mierzącego tego typu konstrukty „powinien przedstawiać wszelkie dostępne informacje, które pomogą użytkownikowi w określeniu, jakie właściwości [...] znajdują odbicie w wariancji wyników testu" (C 16), oraz „przedstawić zarys teorii, na której oparty jest test, i porządkować wszelkie dostępne dane walidacyjne, aby pokazać, w jaki sposób wspierają one teorię" (C 17).

Trynitarne standardy trafności

W okresie dominacji trynitarnego modelu trafności opublikowane zostały – oprócz *Technicznych rekomendacji* – dwa zbiory standardów pomiaru, w których model ten propagowano.

Pierwszy z nich: *Standardy dla Testów Edukacyjnych i Psychologicznych i Podręczników Testowania* (APA i in. 1966), powielał strukturę *Technicznych rekomendacji*, przy czym w rozdziale *Trafność* w miejsce sekcji *Trafność prognostyczna* i *Trafność diagnostyczna* wprowadzono gruntownie przeredagowaną *Trafność kryterialną*. W *Trafności teoretycznej* powtórzono 4 rekomendacje z *Technicznych rekomendacji* APA, wskazujące konieczność przedstawienia korelacji między testem i innymi testami, „których interpretacja jest relatywnie klarowna", a także innymi akceptowanymi miarami tego samego konstruktu (standardy C 7.12 i C 7.13), konieczność zaprezentowania danych dotyczących efektu prędkości w wypadku testu przeprowadzanego w określonych granicach czasowych (C 7.25) oraz informowania o wynikach analizy czynnikowej, jeżeli w takiej wykorzystano dany test (C 7.22). Rekomendacje te stanowiły rozwinięcie jednego z dwóch podstawowych zaleceń APA z 1954 roku: o konieczności „przedstawienia wszelkich dostępnych informacji, które towarzyszyły będą użytkownikowi w rozpoznaniu tych

właściwości psychologicznych, które odbijają się w wariancji wyników" (C 18). W wydaniu z 1966 roku zalecenie to zmieniono, nadając mu postać: „Jeżeli autor proponuje interpretować test jako miarę pewnej zmiennej teoretycznej – zdolności, cechy czy postawy – proponowana interpretacja powinna zostać w pełni wyrażona", a także „odróżniona od interpretacji wynikających z innych teorii" (C 7). Ponadto uszczegółowiono je w dwóch nowych stwierdzeniach: o konieczności podsumowania badań związanych z hipotezami wywiedzionymi z teorii oraz przedstawienia danych dotyczących stopnia, w jakim inne konstrukty znajdują odbicie w wariancji wyników.

Wspomniane wyżej rekomendacje oraz 49 pozostałych standardów trafności powiązano z trynitarną koncepcją trafności. Świadczy o tym także wstęp wprowadzający pojęcie trafności, w którym powtórzono – z drobnymi korektami – sformułowania z *Technicznych rekomendacji* APA, na przykład stwierdzenie, iż „ocena trafności teoretycznej dokonywana jest w trakcie poszukiwania tych jakości, które mierzone są za pomocą testu, to znaczy ustalenia stopnia, w jakim określone pojęcia eksplikatywne albo konstrukty wyjaśniają wykonanie danego testu" (APA i in. 1966: 13). Jedyną znaczącą zmianą w tym wstępie jest termin „trafność kryterialna".

Trynitarna koncepcja trafności przyjęta została także w kolejnych *Standardach dla Testów Edukacyjnych i Psychologicznych* (APA i in. 1974). W rozdziale *Standardy dla raportów z badań rzetelności i trafności* zamieszczono 82 standardy trafności. W sekcji *Trafność treściowa* powtórzono 6 standardów z edycji poprzedniej, dotyczących doboru próby treści z uniwersum programowego, oraz dodano zalecenie sprawdzania treści testu pod kątem ewentualnej jej stronniczości (standard E 12.1.2) oraz sugestię dotyczącą dowodów trafności treściowej testów zawodowych (E 12.4). W *Trafności teoretycznej* powtórzono 10 z 12 rekomendacji z 1966 roku, utrzymując strukturę ich zapisu: zalecenie główne (pełnego wyrażenia proponowanej interpretacji wyników testu jako miary konstruktu), dwa zalecenia podrzędne (wskazania stopnia, w jakim proponowana interpretacja została zweryfikowana, i przedstawienia danych dotyczących stopnia, w jakim inne konstrukty znajdują odbicie w wariancji wyników) oraz kilka zaleceń szczegółowych. Dodano przy tym tylko jedną nową rekomendację wskazującą, iż „jeżeli różnice w strategiach rozwiązywania testu, które mogą wpływać na interpretację wyników, są związane z identyfikowalnymi właściwościami grup społecznych, ta informacja powinna zostać klarownie przedstawiona [...]" (E 13.2.25).

W omawianym rozdziale gruntownie zmieniono natomiast wstęp wprowadzający w problematykę trafności. Podjęto w nim próbę zdefiniowania terminu

„konstrukt" – jako „idei rozwiniętej lub «skonstruowanej» w wyniku pracy wyobraźni naukowej i poinformowanej [...], mającej na celu wyjaśnienie i uporządkowanie niektórych aspektów istniejącej wiedzy" (APA i in. 1974: 29) – oraz nowego opisu koncepcji trafności.

Trafność – w tym ujęciu – odnosi się do odpowiedniości wniosków wyprowadzonych z wyników testu lub innych form oceny. Wyróżnione typy trafności – kryterialna, treściowa i teoretyczna – dotyczą różnych typów wniosków, które sformułowane być mogą na podstawie wyników testowania. Te aspekty trafności są ze sobą powiązane tak, iż „wszelka wiedza dotycząca trafności jest relewantna w trakcie ewaluacji trafności teoretycznej" (APA i in. 1974: 26). Ocena testu w świetle określonego konstruktu wymaga zatem akumulacji wyników wszelkich dostępnych badań.

We wstępie zarysowano także przykładowe koncepcje badania trafności. Walidacja – rozumiana jako proces gromadzenia i oceny danych związanych z interpretacją wyników – rozpoczynać się zatem może od sformułowania hipotez dotyczących właściwości osób, które uzyskały wysokie i niskie wyniki. W jej trakcie możliwe jest chociażby wykorzystanie wyniku testu jako zmiennej zależnej bądź niezależnej w badaniu eksperymentalnym. Można również przeprowadzić weryfikację prognoz zachowania się osób o odmiennych wynikach w różnych sytuacjach czy analizę zadań w powiązaniu z zewnętrznym kryterium. Jak jednak podkreślono: „Dowody trafności teoretycznej nie są owocem pojedynczego badania; sądy dotyczące trafności teoretycznej opierają się raczej na zakumulowanych wynikach wielu badań" (APA i in. 1974: 30).

Unitarne standardy trafności

Kolejna edycja *Standardów edukacyjnego i psychologicznego testowania* (AERA i in. 1985) różniła się zasadniczo od wersji z 1974 roku. Po pierwsze, była obszerniejsza. Po drugie, ustępy na temat trafności zostały ulokowane na początku części pierwszej: *Technicznych standardów konstrukcji i ewaluacji testu*. Po trzecie, zostały one zasadniczo przeredagowane.

Zmiany w podrozdziale *Trafność* wiązały się z fundamentalną zmianą w koncepcji trafności, która została wyrażona już w pierwszym jego akapicie. Ponieważ trafność – definiowana jako odpowiedniość, sensowność i użyteczność określonej inferencji wyprowadzonej z wyników testowania, a więc jako stopień, w jakim zgromadzone dowody empiryczne i teoretyczne uzasadniają daną infe-

rencję – jest „unitarną koncepcją" (AERA i in. 1985: 9), w dokumencie nie stosuje się już tradycyjnych terminów określających typy czy aspekty trafności. W ich miejsce wprowadza się nazwy trzech kategorii dowodów trafności, dotyczących konstruktu, treści testu i miar kryterialnych. Odchodzi się także od wyodrębniania sekcji zawierających standardy związane z typami trafności.

W *Trafności* zamieszczono 25 standardów trafności. Trzy spośród nich odnoszą się do problemów łączonych wcześniej z trafnością teoretyczną, a w tej wersji – z dowodami dotyczącymi konstruktu. Standard 1.8, czyli przeredagowany standard E 13 z poprzedniej edycji, stanowi, iż „jeżeli test proponowany jest jako miara konstruktu, konstrukt ów powinien być wyodrębniony spośród innych konstruktów; proponowana interpretacja wyników testu powinna być jednoznacznie określona; a związane z tym konstruktem dowody trafności powinny być przedstawione w celu uzasadnienia takich inferencji. W szczególności należy przedstawić dowody, które ukazują, że test nie zależy silnie od innych konstruktów". Dwie kolejne rekomendacje, nieobecne w *Standardach* z 1974 roku, precyzują, iż „powinno się przedstawić dowody wskazujące, że wynik znajduje się w bliższej relacji z wynikami pomiaru tego samego konstruktu innymi metodami niż z miarami innych konstruktów" (standard 1.9), oraz „demonstrujące, że wyniki testu są w bliższej relacji ze zmiennymi stanowiącymi przedmiot zainteresowania niż ze zmiennymi niewłączonymi do przyjętej teorii" (1.10).

Te trzy standardy poprzedzone są powtórzonymi za *Standardami* z 1974 roku ogólnymi zaleceniami. Zgodnie z nimi, „należy przedstawiać dowody trafności głównych typów inferencji, dla których rekomendowane jest użycie danego testu" (1.1), zaś w przypadku „jeżeli trafność danej interpretacji nie była przedmiotem badania, ten fakt powinno się wyraźnie wskazać" (1.2). Zaleca się także, aby stwierdzenia na temat trafności odnosiły się do określonych interpretacji czy decyzji (1.2).

Przywołanym powyżej standardom towarzyszą rekomendacje związane z pozostałymi dwoma typami dowodów trafności: treściowymi i kryterialnymi, a także charakterystyka źródeł tych dowodów, takich jak korelacje między zadaniami testu, korelacje między wynikami testu a wynikami zastosowania różnorodnych metod pomiaru danego i innych konstruktów, analiza procesów aktywnych w czasie pracy z zadaniem, a także sąd ekspercki, obejmujący relacje pomiędzy zadaniami a treściową dziedziną testowania. W tym ostatnim przypadku rekomenduje się – podobnie jak w poprzednich edycjach – udostępnienie użytkownikowi testu informacji na temat doświadczenia, kwalifikacji i przygotowania ekspertów oraz procedur wykorzystanych w trakcie uzgadniania sądów (1.7).

Unitarna zmiana w koncepcji trafności została pogłębiona w kolejnej edycji *Standardów edukacyjnego i psychologicznego testowania* (AERA i in. 1999). W tym wydaniu, podobnie jak w edycji poprzedniej, podkreśla się, że trafność jest unitarną koncepcją, i odchodzi od tradycyjnej terminologii, nazywającej typy trafności, w stronę terminologii określającej typy dowodów trafności.

Wśród wspomnianych dowodów znalazły się dowody oparte na badaniach treści i wewnętrznej struktury testu, procesów odpowiedzi na zadania, relacji pomiędzy wynikami zastosowania testu oraz miar tego samego i innych konstruktów, technicznej jakości systemu testowania (na przykład procedur konstrukcji i przeprowadzania testu), a także obserwowanych i przewidywanych konsekwencji testowania i ich źródeł. Do tych ostatnich zaliczono niedoreprezentowanie konstruktu (*construct underrepresetation*) w teście, definiowane jako stopień, w jakim test nie obejmuje ważnych aspektów konstruktu, oraz komponenty irrelewantne względem konstruktu (*construct-irrelevant components*), rozumiane jako te elementy składowe testu, które stymulują procesy niezwiązane z konstruktem.

Zasygnalizowane typy dowodów trafności składają się na argumentację uzasadniającą zamierzoną interpretację wyników, zaś stopień, w jakim przygotowane dowody wspierają tę interpretację, stanowi przedmiot osądu. Dziedziną walidacji jest zatem interpretacja wyników testu, nie zaś sam test.

Dwadzieścia cztery standardy trafności, dotyczące tak rozumianej walidacji, która rozpoczyna się od sformułowania proponowanej interpretacji wyników, obejmują zalecenia uzasadniania – za pomocą dowodów empirycznych i teoretycznych – każdej zamierzonej interpretacji wyników (standard 1.1), jednoznacznego sformułowania planowanej interpretacji wyników i opisu konstruktu stanowiącego przedmiot pomiaru (1.2), ostrzeżenia użytkowników testu w sytuacji braku dowodów uzasadniających interpretację wyników (1.3), a także zgromadzenia przez użytkownika testu danych niezbędnych do uzasadnienia takiego wykorzystania testu, które nie zostało poddane walidacji przez jego konstruktora (1.4). Oprócz tego włączono do zbioru standardów wymóg opisu właściwości próby walidacyjnej (1.5), procedur i kryteriów wyboru treści testu (1.6) oraz ekspertów – w przypadku walidacji z wykorzystaniem opinii eksperckich (1.7). Zalecono także przedstawienie odpowiednich dowodów w wypadku uzasadnienia odwołującego się do procesów psychologicznych i operacji kognitywnych, zachodzących u egzaminowanych (1.8), oraz opis warunków, w jakich zgromadzono dane wykorzystane w analizach statystycznych, jeżeli wyniki tego rodzaju analiz zawarte są w dowodach trafności (1.13). Ponadto zarekomendowano standard uzasadniania wyboru zmiennych (wraz z opisem konstruktów,

które one reprezentują) wykorzystanych w analizach porównawczych (1.14) oraz praktykę prezentowania danych o wewnętrznej strukturze testu w wypadku interpretacji opartej na wnioskowaniu z przesłanek dotyczących relacji między częściami testu (1.11).

W zbiorze standardów trafności znalazły się również rozstrzygnięcia w zakresie problematyki skutków zastosowania testu. Przyjęto, iż jeżeli oczekuje się, że zastosowanie lub interpretacja wyników testu przyniosą określone następstwa, podstawy antycypacji rezultatów użycia testu powinny być przedstawione (1.22 i 1.23). Zalecono także podjęcie próby badania niezamierzonych konsekwencji użycia testu w celu rozstrzygnięcia, czy nie wynikają one z niedoreprezentowania konstruktu w teście lub obecności komponentów irrelewantnych względem konstruktu, lub czułości testu na inne niż przedmiot pomiaru właściwości (1.24).

W komentarzu poprzedzającym standardy trafności podkreślono, że za walidację – rozumianą jako rozwijanie naukowych argumentów dotyczących trafności zamierzonej interpretacji wyników – odpowiedzialni są wspólnie konstruktor i użytkownik testu. Ten pierwszy odpowiada za zebranie dowodów trafności proponowanego odczytania wyników pomiaru, ten drugi – za ocenę tych dowodów z uwzględnieniem tych okoliczności, w których projektuje on zastosowanie testu, a także za zgromadzenie dodatkowych dowodów, jeżeli użytkowanie testu różni się od określonego przez jego konstruktora.

Translokacja trafności konsekwencyjnej

Unitarny model trafności został zastosowany także w najnowszych – niemalże 230-stronicowych – *Standardach edukacyjnego i psychologicznego testowania* (AERA i in. 2014). Edycja ta stanowi zrewidowaną wersję *Standardów* z 1999 roku.

W *Trafności*, w pierwszym jej rodziale, do tekstu z poprzedniej edycji wprowadzono bądź poprawki, bądź dodatkowe ustępy. Zmieniono także układ sekcji *Standardy trafności*, zawierającej tym razem 26 standardów trafności, pogrupowanych w trzech tzw. klastrach. Klastry te poprzedzono rekomendacją nadrzędną, wyrażającą fundamentalną zasadę konstruktorską – zasadę konstruowania i udostępniania językowej interpretacji wyników testu oraz jej uzasadnienia – w formule: „Klarowna artykulacja każdej zamierzonej interpretacji wyniku testu dla określonego zastosowania powinna być przedstawiona i odpowiednie dowody trafności na poparcie każdej zamierzonej interpretacji powinny być udostępnione" (standard 1.0).

W klastrze *Ustalanie zamierzonych zastosowań i interpretacji* umieszczono siedem standardów trafności, przeniesionych – ze zmianami – z edycji poprzedniej (1.1, 1.2, 1.3, 1.4, 1.9, 1.22 i 1.23). Obejmują one między innymi standard „klarownego przedstawiania, jak wyniki testu mają być interpretowane i w konsekwencji stosowane", a w tym także przejrzystego opisu mierzonych konstruktów (1.1), standard prezentowania uzasadnienia każdej zamierzonej interpretacji wyników, obejmującego dowody empiryczne i teoretyczne (1.2), oraz standard informowania użytkowników testu o ewentualnym braku walidacji określonej interpretacji jego wyników lub o jej niezgodności z dostępnymi dowodami (1.3). Wspomniane wyżej zmiany leksykalne uściślają, iż przedmiotem walidacji jest interpretacja wyników testu w konkretnej sytuacji jego użytkowania.

W klastrze *Kwestie dotyczące prób i metod wykorzystanych w walidacji* powtórzono trzy standardy z edycji wcześniejszej (1.4, 1.7 i 1. 13) właściwie bez zmian. Rekomendacje te dotyczą sposobu opisu próby walidacyjnej (1.8), procedur związanych z udziałem sędziów kompetentnych w badaniach trafności (1.9) oraz metod gromadzenia danych wykorzystanych w analizach statystycznych (1.10).

Klaster *Specyficzne formy dowodów trafności* został podzielony na 6 części – zawierających 15 standardów, w tym 14 przeniesionych z wydania poprzedniego – zatytułowanych: *Dowody ukierunkowane na treść* (standard 1.11, czyli 1.6 z 1999 roku), *Dowody ukierunkowane na procesy kognitywne* (1.12/1.8), *Dowody ukierunkowane na wewnętrzną strukturę* (1.13/1.11, 1.14/1.12 i 1.15/1.10), *Dowody ukierunkowane na relacje z konceptualnie powiązanymi konstruktami* (1.16/1.14), *Dowody ukierunkowane na relacje z kryteriami* (1.17/1.16, 1.18/1.15, 1.19/1.17, 1.20, 1.21/1.18, 1.22/1.20, 1.23/1.21 i 1.24/1.19) oraz *Dowody oparte na konsekwencjach testów* (1.25/1.24). Jedyny nowy standard w tym klastrze stanowi zalecenie raportowania stopnia niepewności związanej z miarami wielkości efektu – w sytuacji wykorzystywania ich we wnioskowaniu (1.20). Nieliczne zmiany w pozostałych standardach to albo korekty leksykalne, albo kilkuwyrazowe uzupełnienia.

Najobszerniejsze zmiany w komentarzach dołączonych do każdego standardu wprowadzone zostały w ustępie poświęconym standardowi 1.25, który określa działania badawcze, jakie należy podjąć w sytuacji wystąpienia niezamierzonych konsekwencji zastosowania testu, w słowach: „Gdy niezamierzone konsekwencje stanowią rezultat zastosowania testu, należy podjąć próbę zbadania, czy konsekwencje te wynikają z czułości testu na cechy inne niż te, które zamierzano ocenić, czy też z defektu testu w zakresie pełnego reprezentowania zamierzonego konstruktu" (1.25). W ustępie dodanym do trzyzdaniowego komentarza z 1999 roku, poświęconego kwestii relacji między konsekwencjami użycia testu a zachodzącym w nim niedoreprezentowaniem konstruktu i funkcjonującymi w jego

ramach komponentami irrelewantnymi względem konstruktu, stwierdzono, że „odkrycie niezamierzonych konsekwencji może prowadzić do namysłu nad odpowiedniością zastosowanego konstruktu". Dopowiedziano także, iż „odpowiedzialność za upewnienie się, że niezamierzone konsekwencje są przedmiotem ewaluacji, leży po stronie tych, którzy podejmują decyzję w sprawie użycia danego testu" (AERA i in. 2014: 31).

Również w sekcji poprzedzającej *Standardy trafności*, w której zreferowano problematykę źródeł dowodów trafności interpretacji wyników testu w danym jego zastosowaniu, najobszerniejsze zmiany wprowadzono w dziale dotyczącym konsekwencji testowania, który w edycji z 1999 roku nosił tytuł *Dowody oparte na konsekwencjach testowania*, zaś w wydaniu ostatnim – *Dowody trafności i konsekwencje testowania*. W dziale tym – obok ustępów z poprzedniej edycji, w których podjęto próbę modyfikacji unitarnego modelu Messicka w jego wymiarze konsekwencyjnym – uwzględniono bardziej zdecydowane rozgraniczenie między problematyką trafności interpretacji wyników testu w danym jego zastosowaniu a problematyką konsekwencji praktyk pomiarowych. I o ile we wcześniejszej edycji zwracano już uwagę na „rozróżnienie pomiędzy dowodami, które są bezpośrednio relewantne względem trafności, i dowodami, które mogą wspierać decyzje dotyczące polityki społecznej, ale które znajdują się poza sferą trafności", i przestrzegano, iż „chociaż informacje o konsekwencjach testowania mogą mieć wpływ na decyzje dotyczące stosowania testów, to takie konsekwencje nie umniejszają, same w sobie, trafności zamierzonych interpretacji wyników testu" (AERA i in. 1999: 16, AERA i in. 2014: 20), to dopiero w wydaniu ostatnim jednoznacznie stwierdzono, że konstruktor testu odpowiedzialny jest jedynie za przeprowadzenie walidacji rozumianej jako zebranie dowodów trafności proponowanej przez niego interpretacji wyników otrzymanych w zamierzonym zastosowaniu testu i że odpowiedzialność za zgromadzenie dowodów dotyczących twierdzeń wskazujących na pozytywne lub negatywne konsekwencje stosowania testu spoczywa na stronie, która takie twierdzenia formułuje.

W dziale *Dowody trafności i konsekwencje testowania* przywołano także przypadek ścisłego związku pomiędzy trafnością i konsekwencjami. Podobnie jak w edycji z 1999 roku, stwierdzono bowiem, iż „dowody dotyczące konsekwencji są relewantne względem trafności wtedy, kiedy mogą być powiązane ze źródłem braku trafności, takim jak niedoreprezentowanie konstruktu czy komponenty irrelewantne względem konstruktu" (AERA i in. 2014: 21). Przypadek ten ujęty został także w ramach standardu 25, stanowiącego, iż w sytuacji wystąpienia niezamierzonych konsekwencji użycia testu należy podjąć próbę zbadania, czy nie wypływają one z przywołanych wyżej źródeł braku trafności.

W innych działach sekcji poprzedzającej *Standardy trafności*, zawierających tekst z poprzedniej edycji wraz z drobnymi korektami oraz obszernymi uzupełnieniami, zasygnalizowaną powyżej kwestię relacji między trafnością i konsekwencjami poruszono w ustępach dodanych, na przykład w drugim akapicie tej sekcji, a więc już na pierwszej stronie rozdziału poświęconego trafności. Stwierdzono w nim, iż – po pierwsze – warunkiem *sine qua non* uzasadnionego zastosowania testu jest posiadanie dowodu trafności interpretacji jego wyników uzyskanych w podobnych okolicznościach jego zastosowania oraz że – po drugie – dysponowanie takim dowodem stanowi niewystarczający warunek jego użycia. Jak to ujęto w przywołanym ustępie: „W przypadku, gdy istnieją wystarczające dowody trafności, decyzja, czy przeprowadzić właśnie dany test, wymaga zazwyczaj dodatkowego namysłu. Obejmuje on kwestie kosztów i korzyści, ujęte w różnych subdyscyplinach jako analiza użyteczności lub refleksja nad negatywnymi konsekwencjami zastosowania testu, oraz wszelkie negatywne konsekwencje skonfrontowane z pozytywnymi konsekwencjami zastosowania testu" (AERA i in. 2014: 11). Takie ujęcie problemu konsekwencji użycia testu, iż oto należy je rozważyć, dysponując już dostarczonymi przez konstruktora testu dowodami trafności, stanowi poważną próbę zawężenia unitarnego pojęcia trafności i wyłączenia z jego zakresu tzw. trafności konsekwencyjnej (Niemierko 2009: 148).

Pozostałe ustępy dodane do tekstu zamieszczonego we wspomnianych wyżej działach nie zawierają tak zasadniczych zamian pojęciowych. Znajdują się w nich dodatkowe przykłady ilustrujące kontaminację konstruktu (wariancję irrelewantną względem konstruktu) oraz uwagi dotyczące dostosowania treści testu do programu kształcenia czy eksperckiego osądu jako metody oceny, czy jakość i ilość dowodów trafności są wystarczające. Podkreśla się tam także, iż w celu określonego zastosowania testu nie zachodzi potrzeba gromadzenia wszystkich typów dowodów trafności. W dziale końcowym, poświęconym integrowaniu dowodów trafności, zawarta jest konstatacja, iż mimo że „wnioskowanie walidacyjne podobne jest do wnioskowania naukowego", gdyż „proces walidacji nigdy się nie kończy", to „w pewnym punkcie dowody walidacyjne pozwalają na sumatywny osąd zamierzonej interpretacji, że jest dobrze poparta i możliwa do obronienia", przy czym „ilość i charakter dowodów wymaganych do wsparcia prowizorycznego osądu trafności często różnią się pomiędzy dziedzinami" (AERA i in. 2014: 22).

Pomimo zasygnalizowanych powyżej zmian w koncepcji trafności samą jej definicję powtórzono – z drobną korektą leksykalną – za *Standardami* z 1999 roku. Do definicji tej – w brzmieniu: „trafność odnosi się do stopnia, w jakim

dowody empiryczne i teoria wspierają interpretacje wyników testu dla zaproponowanych zastosowań testu" – dodano jednakże komentarz uściślający, iż „twierdzenia o trafności powinny odnosić się do konkretnych interpretacji związanych ze ściśle określonymi zastosowaniami". W tym kontekście zdecydowanie podkreślono także, że „używanie frazy «trafność testu» [...] jest niepoprawne" (AERA i in. 2014: 11).

Obiekcja

Standardom trafności, sześciokrotnie uzgadnianym i publikowanym przez NCME, AERA i APA[2], towarzyszyła konceptualizacja pojęcia trafności – od modelu trynitarnego, popularyzowanego w trzech pierwszych wydaniach standardów pomiarowych, do modelu unitarnego, upowszechnianego w ostatnich trzech edycjach. W wersji najnowszej model unitarny poddany został istotnej modyfikacji, która polega na translokacji odpowiedzialności za badanie trafności konsekwencyjnej i której skutkiem może okazać się zwrot pragmatyczny w teoriotwórczej refleksji nad trafnością.

Podstawą unitarnego modelu trafności jest bowiem pojęcie trafności konstruktu reprezentującego – jako to ujął Samuel Messick – „nasze najlepsze, chociaż niedoskonałe i omylne, wysiłki uchwycenia esencji cech, które posiadają walor realności w zachowaniach niezależnych od naszych prób ich scharakteryzowania" (Messick 1998: 35). Czy zatem wspomniana translokacja stanowi jedynie próbę scedowania odpowiedzialności za niedoskonałą i omylną interpretację wyników z konstruktora na użytkownika testu, czy raczej wyraża zmianę poważniejszą – epistemologiczną i etyczną zarazem? Czy przejawem tej właśnie zmiany nie jest także argumentacyjny model trafności Michaela T. Kane'a (Skorupiński 2013), będący wyrazem – jak to określił sam Kane (2013: 64) – „bardziej pragmatycznego podejścia", w ramach którego trafność stanowi funkcję nie tylko aktywności badawczej, lecz także perswazji?

Bibliografia

AERA, APA, NCME. 1985. *Standards for educational and psychological testing*, AERA, Washington.
AERA, APA, NCME. 1999. *Standards for educational and psychological testing*, AERA, Washington.

[2] Dwie edycje standardów zostały w całości przełożone na język polski i opublikowane pod redakcją Jerzego Brzezińskiego (Polskie Towarzystwo Psychologiczne 1985; AERA i in. 2007).

AERA, APA, NCME. 2007. *Standardy dla testów stosowanych w psychologii i pedagogice*, tłum. E. Hornowska, GWP, Gdańsk [pierwodruk: AERA i in. 1999].

AERA, APA, NCME. 2014. *Standards for educational and psychological testing*, AERA, Washington.

AERA, NCMUE. 1955. *Technical recommendations for achievement tests*, NEA, Washington.

APA, AERA, NCMUE. 1954. *Technical recommendations for psychological tests and diagnostic techniques*, „Psychological Bulletin", t. 51, nr 2 (suplement), s. 1–38 (201–238).

APA, AERA, NCME. 1966. *Standards for educational and psychological tests and manuals*, APA, Washington.

APA, AERA, NCME. 1974. *Standards for educational and psychological tests*, APA, Washington.

Kane M.T. 2013. *Validating the Interpretations and Uses of Test Scores*, „Journal of Educational Measurement", t. 50, nr 1, s. 1–73.

Madaus G.F. 2001. *A Brief History of Attempts To Monitor Testing*, „NBETPP Statements", t. 2, nr 2, Chestnut Hill, MA, dostępny na: http://files.eric.ed.gov/fulltext/ED456145.pdf [otwarty 14 marca 2015].

Messick S. 1989. *Validity*, [w:] *Educational measurement*, red. R.L. Linn, New York, s. 13–103.

Messick S. 1998. *Test validity: a matter of consequence*, „Social Indicators Research", t. 45, nr 1–3, s. 35–44.

Niemierko B. 2009. *Diagnostyka edukacyjna. Podręcznik akademicki*, Wydawnictwo Naukowe PWN, Warszawa.

Polskie Towarzystwo Psychologiczne. 1985. *Standardy dla testów stosowanych w psychologii i pedagogice*, tłum. E. Hornowska, PTP, Warszawa [pierwodruk: APA i in. 1974].

Skorupiński P.M. 2013. *Modele trafności pomiaru*, [w:] *Ścieżki rozwoju edukacyjnego młodzieży – szkoły pogimnazjalne. Trafność wskaźników edukacyjnej wartości dodanej dla szkół maturalnych*, red. M. Karwowski, IFiS PAN, Warszawa, s. 13–25.

KWARTALNIK PEDAGOGICZNY
2015 NUMER 4(238)

Aleksandra Jasińska-Maciążek
Grzegorz Humenny
Instytut Badań Edukacyjnych w Warszawie

Wykorzystanie egzaminów o niskiej doniosłości do oceny efektywności nauczania – możliwości i ograniczenia

Summary

LOW-STAKES TESTS AS A BASIS FOR THE ASSESSMENT OF TEACHING EFFECTIVENESS

The article discusses the possibility of using low-stakes tests to assess teaching effectiveness measures (educational value-added, EVA) for primary schools in Poland. It was shown that EVA, calculated using the results of the OBUT test for diagnosis of skills of third grade pupils, which is voluntary for schools, and the *Sprawdzian* – external examination conducted in the sixth grade, can be a valuable tool for evaluation, provided that the third-grade test is conducted and assessed in accordance with the procedures. The authors identified problems in this field. Disturbing characteristics of the OBUT test may suggest that these procedures have not been complied with in every school. For this reason, the authors recommend changes in the implementation of such diagnoses of school achievements.

Key words: low-stakes tests, assessment of teaching effectiveness, educational value-added, primary school.

red. Paulina Marchlik

Testy osiągnięć szkolnych możemy podzielić, ze względu na konsekwencje, jakie wiążą się z wykorzystaniem ich wyników, na testy o doniosłym znaczeniu (wysokiej stawki – *high stakes*) i testy o niskiej doniosłości (powszednie, niskiej stawki – *low stakes*) (Niemierko 2009). W podziale tym bierze się pod uwagę konsekwencje dla uczniów. Egzaminy doniosłe decydują o ich losach na kolejnych etapach kształcenia. Testy o niskiej doniosłości nie wiążą się z poważnymi konsekwencjami (są to np.: szkolne klasówki, testy diagnostyczne). Jednak test

o niskiej doniosłości dla ucznia, może być testem doniosłym z punktu widzenia szkoły lub nauczycieli, jeśli jest wykorzystywany do oceny ich pracy.

Obecnie coraz większe znaczenie przypisuje się testom osiągnięć szkolnych jako narzędziu ewaluacji placówek edukacyjnych oraz monitorowania systemu oświaty, służącemu poprawie efektywności nauczania (EACEA 2010; OECD 2013). W niektórych krajach wykorzystuje się do tego celu egzaminy doniosłe (np. Szkocja, Irlandia, Łotwa), w innych testy o niskiej doniosłości (np. Australia, Węgry, Anglia). Poniżej przybliżamy założenia systemów opartych na testach niskiej stawki zaznaczając, że są to przykłady, a nie wyczerpująca lista takich rozwiązań.

Programy ewaluacyjne w Australii, Anglii i na Węgrzech

Australijski system ogólnokrajowych egzaminów *National Assessment Program – Literacy and Numeracy* (NAPLAN) istnieje od 2008 roku. Do testów przystępują wszyscy uczniowie kończący naukę w klasach 3, 5, 7 i 9 (Freeman 2009). Główną ich funkcją jest monitorowanie postępów uczniów, w celu zapewnienia im możliwości jak najlepszego wykorzystania potencjału. Na Węgrzech od 2001 roku istnieje ogólnokrajowa ocena podstawowych kompetencji (*National Assessment of Basic Competencies* – Sinka 2010). Od 2008 roku pomiar obejmuje wszystkich uczniów kończących 6, 8 i 10 klasę (wcześniej badania były realizowane na próbach), a uczniowie dodatkowo wypełniają ankietę dotyczącą ich pochodzenia społecznego, co pozwala na kontekstową ocenę wyników szkół. Ponadto w czwartej klasie (10 lat) realizowane jest badanie diagnostyczne (*Diagnostic skills and ability assessment*), które ma na celu ocenę poziomu umiejętności uczniów i dostarczenie nauczycielom informacji potrzebnych do oceny i planowania własnej pracy z uczniami (Hungarian Eurydice Unit, 2009). Program węgierski został wprowadzony w celu rozwoju systemu ewaluacji jednostek edukacyjnych, dostarczenia szkołom danych i narzędzi umożliwiających profesjonalną ocenę własnej pracy oraz porównanie swoich wyników z innymi szkołami w kraju (Balázsi 2006). Podobnie w angielskim systemie egzaminów krajowych (*National Curriculum Assessment*) od 2008 roku testuje się wszystkich uczniów w 2, 6 i 11 (lub 10) roku nauki. Dodatkowo istnieje dobrowolny pomiar w dziewiątym roku nauki. System ten stworzono w przekonaniu, że monitorowanie rozwoju poznawczego uczniów jest dobrą praktyką, a ocena poziomu osiągnięć może wspierać proces nauczania i uczenia się (National Foundation for

Educational Research 2009). Żaden z wymienionych pomiarów umiejętności nie ma charakteru doniosłego dla uczniów.

Testy te, w zależności od systemu i etapu edukacji, przeprowadzane i oceniane są w różny sposób. W Anglii pomiar po drugim roku nauki oceniany i przeprowadzany jest przez nauczycieli badanych klas. Podobnie jak egzamin w czwartej klasie na Węgrzech (za wyjątkiem arkuszy wybranych do ogólnokrajowej próby, ocenianych zewnętrznie). Pozostałe testy w tych krajach, oceniane są przez zewnętrznych egzaminatorów. W systemach australijskim i angielskim wyniki kolejnych pomiarów są przedstawiane na wspólnej skali osiągnięć szkolnych odpowiadającej założonym poziomom realizacji krajowego programu nauczania. Pozwala to na ocenę postępów ucznia oraz sprawdzenie, czy jego wyniki są zgodne z oczekiwanymi (EACEA 2010; ACARA 2014). Na Węgrzech testy sprawdzają nie tyle stopień opanowania programu nauczania, co umiejętności wykorzystania posiadanej wiedzy do rozwiązywania problemów praktycznych (Hungarian Eurydice Unit 2009).

W każdym z tych systemów najważniejszym odbiorcą wyników są szkoły i nauczyciele, a podstawowym celem pomiarów jest poprawa efektywności nauczania. Wyniki szkół z większości pomiarów obowiązkowych są powszechnie dostępne[1]. Są one prezentowane nie tylko na tle średniej w kraju i w mniejszych jednostkach terytorialnych, ale także jako przetworzone wskaźniki uwzględniające kontekst pracy szkoły lub uprzednie osiągnięcia szkolne uczniów. Wyniki te wykorzystywane są nie tylko przez szkoły. Stanowią też źródło informacji dla rodziców zainteresowanych wyborem szkoły dla dziecka. Służą także władzom i innym organom nadzoru do ewaluacji zewnętrznej (Balázsi 2006; EACEA 2010). To sprawia, że stają się one dla szkół testami wysokiej stawki.

Egzaminy w Polsce

W Polsce powszechny pomiar osiągnięć odbywa się od 2002 roku w ramach systemu egzaminów zewnętrznych. Egzamin gimnazjalny oraz matura pełnią m.in. funkcję selekcyjną, ewaluacyjną i monitorującą (Dolata, Szaleniec 2012). Są wykorzystywane przez szkoły do oceny i planowania własnej pracy (Matuszczak, Wasilewska 2015; Wasilewska i in., 2014) oraz przez organy prowadzące

[1] Portal australijski: www.myschool.edu.au i angielski: http://www.education.gov.uk/schools/performance.

i inne podmioty w ewaluacji zewnętrznej (Herczyński i in. 2012; Bąbiak i in. 2013; SEO 2013).

Obowiązkowy egzamin na zakończenie szkoły podstawowej – sprawdzian – miał pełnić funkcję diagnostyczną, jednak jest testem krótkim, mało zróżnicowanym pod względem treści zadań, nie posiadającym norm dla poszczególnych wymagań programowych (Dolata, Szaleniec 2012), co utrudnia wykorzystanie jego wyników do oceny realizacji wymagań programowych. Dla uczniów, co do zasady, jest to test niskiej stawki. W praktyce niektóre gimnazja wykorzystują jednak jego wyniki jako jedno z kryteriów selekcji do szkoły uczniów spoza rejonu[2]. Sprawdzian wypełnia także funkcję ewaluacyjną – pomaga szkołom w ocenie efektów nauczania i znajdowaniu obszarów do poprawy (Matuszczak, Wasilewska 2015).

Egzaminy zewnętrzne nie obejmują wcześniejszych etapów edukacji. Do 2010 roku szkoły mogły korzystać jedynie z narzędzi przygotowanych samodzielnie albo z oferty wydawnictw proponujących własne testy. Wadą tych rozwiązań był brak możliwości zobaczenia rezultatów na tle wyników populacyjnych. Od 2011 do 2014 roku organizowane było Ogólnopolskie Badanie Umiejętności Trzecioklasistów (OBUT)[3] – dobrowolna diagnoza umiejętności z języka polskiego i matematyki. W 2011 roku wzięło w niej udział ponad 75% szkół podstawowych, co dało możliwość analizowania wyników szkół na tle wyników krajowych (Pregler, Wiatrak 2011).

Wykorzystanie wyników testów do oceny efektywności nauczania

Jednym z negatywnych skutków egzaminów zewnętrznych jest traktowanie ich wyników jako głównego lub jedynego wskaźnika oceny pracy szkoły i nauczycieli (Szaleniec 2004). Średnie wyniki wykorzystywane są np. do tworzenia prasowych rankingów szkół. Zestawienia te nie biorą pod uwagę faktu, że szkoły nie mają takich samych szans na uzyskanie porównywalnych rezultatów. Wyniki zależą od szeregu czynników, na które szkoła nie ma wpływu, takich jak inteligencja, status społeczno-ekonomiczny rodziny, wykształcenie i aspiracje rodziców (Konarzewski 2012; Dolata i in. 2014; Jasińska-Maciążek, Modzelewski 2014).

[2] Np. Gimnazjum nr 46 w Bydgoszczy: http://www.2lo.bydgoszcz.pl

[3] W latach 2006–2011 badanie było realizowane na reprezentatywnych próbach ogólnopolskich pod kierownictwem M. Dąbrowskiego i M. Żytko.

Szkoły, wykorzystując wyniki testów do oceny własnej pracy, przenoszą uwagę z ucznia na proces nauczania. Może przełożyć się to na skuteczniejsze nauczanie. Aby szkoła mogła podejmować trafne decyzje, potrzebuje jednak dobrych wskaźników procesów w niej zachodzących (MacBeath i in. 2005). Wskaźniki te powinny być niezależne od tego, na co szkoła nie ma wpływu. Wyniki testowe szkół powinny być odnoszone do wyników innych placówek pracujących z podobnymi uczniami. Z takiego myślenia wywodzi się idea edukacyjnej wartości dodanej.

Edukacyjna wartość dodana (EWD) jako miara efektywności nauczania

Wskaźniki EWD są miarą względnego przyrostu osiągnięć szkolnych uczniów na danym etapie kształcenia, a do ich obliczenia potrzebne są co najmniej dwa pomiary osiągnięć: na początku i na zakończenie tego etapu. Pokazują one, czy uczniowie w danej szkole osiągnęli średnio wyniki wyższe, niższe czy porównywalne do wyników uczniów podobnych pod względem uprzednich osiągnięć. Z tego względu uważane są za lepszą miarę jakości nauczania niż np. średnie wyniki z testu (OECD 2008).

Edukacyjna wartość dodana traktowana jest jako miara wkładu szkoły lub nauczycieli w wyniki egzaminacyjne – wskaźnik efektywności nauczania (McCaffrey i in. 2003; OECD 2008). Najprostsze modele uwzględniają jedynie informację o poziomie uprzednich osiągnięć uczniów, bowiem nie tylko bezpośrednio determinuje on wyniki na kolejnych etapach nauczania, ale jest także nośnikiem informacji o czynnikach niezależnych od szkoły (Dolata i in. 2013). W modelach EWD można również uwzględniać charakterystyki uczniów (np. inteligencja, status społeczny rodziny) i otoczenia szkoły (np. wielkość miejscowości, zamożność mieszkańców, poziom bezrobocia), by w ten sposób oczyścić wskaźniki z ich wpływu. To, jakie zmienne zostaną uwzględnione powinno zależeć od celów, jakim mają służyć obliczone miary (Raudenbush, Willms 1995). Niezależnie od liczby zmiennych kontekstowych nie należy jednak traktować wskaźników EWD jako bezwzględnego dowodu na efektywność działań podejmowanych przez szkoły i nauczycieli (McCaffrey i in. 2003; Raudenbush 2004). Dobre wykorzystanie miary EWD w ewaluacji wymaga sięgania po dodatkowe informacje dostępne nauczycielom pracującym w szkole. Analiza wskaźników

EWD razem z kontekstem pracy szkoły może pomóc w ocenie skuteczności podejmowanych działań i bardziej świadomym planowaniu pracy dydaktycznej (Stożek 2012).

W Polsce metoda EWD dopiero się rozwija. Gimnazja mogą analizować wskaźniki EWD od 2006 roku, a licea i technika od 2010 roku. Dla szkół podstawowych do 2014 roku nie było możliwości obliczania wskaźników EWD, ponieważ jedynym powszechnym pomiarem osiągnięć szkolnych był sprawdzian w klasie szóstej. Badanie OBUT umożliwiło myślenie o modelach EWD dla drugiego etapu edukacyjnego (klasy IV–VI). Uczniowie biorący w nim udział w 2011 roku trzy lata później przystąpili do sprawdzianu. Był to pierwszy rocznik uczniów, dla którego można było obliczyć wskaźniki EWD dla szkół podstawowych.

Niebezpieczeństwa wynikające z wykorzystania egzaminów o niskiej doniosłości do oceny nauczania

Wyniki testów o niskiej doniosłości, jako że nie wiążą się z poważnymi konsekwencjami dla ucznia, mogą być szczególnie podatne na zniekształcenia związane z wpływem motywacji testowej. Determinuje ona podjęcie przez ucznia próby rozwiązania zadań, wysiłek poznawczy i wytrwałość w ich rozwiązaniu (Dolata, Pokropek 2010; Skórska i in. 2014). Wyniki takich testów prawdopodobnie zaniżają poziom umiejętności uczniów (Wolf, Smith 1995; Wise, DeMars 2005), przekładając się na niedoszacowanie efektów kształcenia.

Fakt, że test nie ma znaczących konsekwencji dla uczniów nie oznacza, że jest on testem niskiej stawki dla nauczycieli czy szkół (Corcoran i in. 2011). Jeśli wyniki wykorzystywane są do oceniania pracy nauczycieli albo podejmowania istotnych dla nich decyzji, mogą oni próbować wpłynąć na poziom motywacji testowej uczniów. W efekcie uzyskane wyniki mogą być inne niż w sytuacji, gdy pomiar jest testem o niskiej doniosłości dla wszystkich zainteresowanych (Papay 2011; Jager i in. 2012). W kontekście porównywania wyników szkół bardziej problematyczny jest zróżnicowany stosunek placówek do testów. W przypadku badania OBUT jest prawdopodobne, że taka sytuacja miała miejsce. Autorom artykułu znane są przypadki, w których wyniki testu OBUT niosły za sobą ważne konsekwencje dla nauczycieli, np. były komunikowane rodzicom,

którzy mieli wybrać nauczyciela dla swojego dziecka. Brakuje jednak danych o skali tego zjawiska.

Nie tylko konsekwencje formułowane wprost mogą sprawić, że nauczyciele będą postrzegać dany test jako doniosły. Ważna może być obawa przed negatywną oceną koleżeńską lub dyrektorską. Im wyższa postrzegana lub rzeczywista stawka, o której decyduje wynik, tym większe prawdopodobieństwo pojawienia się niezamierzonych, negatywnych konsekwencji. W literaturze można znaleźć różne przykłady oszustw i nadużyć związanych z sytuacją testową. W Holandii, niektóre szkoły podstawowe, nie obejmowały pomiarem słabszych uczniów, próbując w ten sposób podnieść średni wynik (EACEA 2010). W USA wskazuje się na takie praktyki jak zawężanie programów nauczania do treści egzaminu, przekazywanie nauczycielom testów z wyprzedzeniem, by przygotowali uczniów do egzaminu, podpowiadane uczniom podczas pisania testu, podawanie im wcześniej prawidłowych odpowiedzi, „naciąganie" wyników podczas sprawdzania prac, czy zmiana uczniowskich odpowiedzi na kartach egzaminacyjnych (Jacob, Levitt 2003; Nichols, Berliner 2007).

Procedury przeprowadzania i oceniania testów o niskiej doniosłości często są mniej rygorystyczne, niż egzaminów doniosłych. Badanie OBUT było na przykład prowadzone przez pracowników szkoły (zwykle wychowawców klas), bez zewnętrznych obserwatorów – jak w przypadku egzaminów zewnętrznych. Dodatkowo, testy z badania OBUT były sprawdzane przez nauczycieli badanych klas lub innych pracowników szkoły według reguł opracowanych przez zespół badawczy (CKE 2011). Jest to zrozumiałe z punktu widzenia kosztów testowania, jednak sytuacja taka może nasilać problemy związane z oszustwami egzaminacyjnymi, a także skutkować zniekształceniem wyników wynikającym z różnej (niewystandaryzowanej) interpretacji procedur testowania czy reguł oceniania prac.

Warunki obliczenia wartościowych wskaźników efektywności nauczania dla drugiego etapu edukacyjnego

W literaturze poświęconej metodzie EWD zwraca się uwagę na kilka kluczowych dla obliczenia wskaźników aspektów (OECD 2008; Chudowsky i in. 2010). Przede wszystkim potrzebne są dwa dobre pomiary osiągnięć szkolnych: na początku i na zakończenie etapu nauczania, który chcemy oceniać. Dobry

w tym kontekście test, to taki, który trafnie mierzy poziom realizacji celów kształcenia. Pomiary muszą być rzetelne oraz nieobciążone działaniem czynników niezwiązanych z poziomem osiągnięć (np. motywacją testową, czy niedotrzymaniem procedur realizacji lub oceniania testów). Dodatkowo skala, na której prezentowane są wyniki, powinna być interwałowa (Ballou 2009; Reardon, Raudenbush 2009) i mieć wystarczająco dużo wartości, by mogła być traktowana w analizach jako zmienna ciągła.

Do obliczenia wskaźników EWD można wykorzystać dane populacyjne, jak i dane z próby szkół. Musi to być jednak próba reprezentatywna. Niezbędne jest także, aby dla każdego ucznia wyniki obu pomiarów były poprawnie połączone.

Kolejna kwestia wiąże się z właściwym statystycznym wymodelowaniem względnych postępów uczniów. Modele EWD opierają się na zdroworozsądkowym założeniu, że funkcja opisująca zależność wyników egzaminu na zakończenie danego etapu edukacji od wyników egzaminu na jego początku powinna być niemalejąca (Żółtak 2013). Musi być ona dobrze dopasowana do danych, co w przypadku związku nieliniowego wymaga zastosowania odpowiednich metod.

Istotne jest również to, na ile skutecznie w modelu EWD udaje się kontrolować znaczenie czynników wpływających na postępy uczniów, które są niezależne od szkoły. Dostępne dane pozwalają jedynie na sprawdzenie, czy obliczone wskaźniki są niezależne od średniego w szkole poziomu osiągnięć „na wejściu" (czyli po III klasie). Brak takiej zależności na poziomie uczniów wynika bezpośrednio z modelu. Związek na poziomie szkół wskazywałby na problemy w tym zakresie.

Zdajemy sobie sprawę z tego, że przytoczona powyżej lista warunków nie jest wyczerpująca i dodatkowe badania są potrzebne, by z pełnym zaufaniem wykorzystywać wskaźniki EWD. Nie uda nam się też wystarczająco wnikliwie przeanalizować każdego z przytoczonych problemów. Celem artykułu nie jest jednak sformułowanie jednoznacznej i wielowymiarowej oceny jakości wskaźników dla drugiego etapu kształcenia. Został on zarysowany skromniej, jako zbadanie możliwości wykorzystania wyników pomiaru takiego, jak badanie OBUT i wyników sprawdzianu do szacowania wskaźników EWD oraz naświetlenie możliwych problemów, jakie możemy napotkać wykorzystując do tego testy o niskiej doniosłości. Jest to więc dopiero początek drogi do zbudowania wartościowych miar efektywności nauczania dla szkół podstawowych.

Metoda

Wykorzystane dane

W analizach wykorzystano wyniki sprawdzianu w szóstej klasie – obowiązkowego dla uczniów egzaminu zewnętrznego realizowanego 1.04.2014 (dane z arkuszy standardowych) oraz dane z badania OBUT realizowanego 17.05.2011.

Skalowanie wyników przeprowadzono na danych populacyjnych. Z uwagi na fakt, że wyniki piszących test OBUT zostały zanonimizowane, niemożliwe było połączenie ich z wynikami sprawdzianu na poziomie centralnym. Dlatego modele EWD obliczono na danych z losowej ogólnopolskiej próby szkół, dla których pozyskano połączone dla uczniów wyniki testów. Szkoły przekazały sumy punktów z testów OBUT i sprawdzianu, informację o płci ucznia oraz posiadaniu przez niego zaświadczenia o potrzebie dostosowania warunków na sprawdzianie (tzw. dysleksji rozwojowej).

Dobór próby szkół odbył się przez systematyczne losowanie placówek, które przystąpiły do testu OBUT w 2011 roku, w warstwach wydzielonych ze względu na województwa oraz średni wynik testu OBUT w szkołach. W warstwach szkoły uporządkowane zostały ze względu na liczbę uczniów przystępujących do tego pomiaru. Liczba szkół losowana w każdej warstwie była proporcjonalna do liczby uczniów. Próba założona liczyła 200 szkół. W ustalonym terminie udało się pozyskać dane od 181 szkół.

Charakterystyka narzędzi i skalowanie wyników testów

Test OBUT składał się z dwóch części (języka polskiego i matematyki). Łącznie test zawierał 20 zadań, za które można było uzyskać 41 punktów. Test był relatywnie łatwy (lewoskośny rozkład wyników, współczynnik skośności = –0,70). Sprawdzian zawierał 26 zadań, za które można było uzyskać 40 punktów. Był to również test łatwy (współczynnik skośności = –0,47).

Wyniki obu pomiarów zostały wyskalowane z zastosowaniem modelu Rascha dla zadań ocenianych binarnie oraz modelu *partial credit* dla zadań punktowanych na dłuższych skalach (modele funkcjonujące w ramach teorii IRT – *item response theory*) (De Ayala 2009; Szaleniec 2009). Zastosowano metodę estymacji *marginal maximum likelihood*. Za wskaźnik poziomu osiągnięć przyjęto oszacowania *expected a posteriori* (EAP). Wyniki badania OBUT przedstawiono na jednej

jednowymiarowej skali opisującej osiągnięcia szkolne po III klasie (w modelu uwzględniono łącznie zadania z testu z języka polskiego i matematyki). Wyniki sprawdzianu również przedstawiono na jednowymiarowej skali.

Model Rascha wybrano z dwóch powodów. Po pierwsze, uważa się, że tylko modele jednoparametryczne (a takim jest model Rascha, w którym jedynym szacowanym parametrem jest trudność zadania) mają teoretyczne podstawy co do tego, by skale utworzone z ich zastosowaniem można było traktować jako interwałowe (Ballou 2009). Czyli, by można było przyjąć, że wzrost o jedną jednostkę na skali oznacza równoważny przyrost osiągnięć szkolnych w każdym jej zakresie. Przy czym można tak stwierdzić tylko wtedy, gdy model jest dobrze dopasowany. Dopasowanie to było przedmiotem analiz i zostało omówione w części wynikowej. Założenie to nie jest natomiast spełnione dla skali nieprzetworzonej sumy punktów.

Ponadto spośród różnych modeli IRT tylko w modelu Rascha suma punktów zdobyta w teście jest statystyką dostateczną dla oszacowania poziomu osiągnięć (wyniku wyskalowanego) (De Ayala 2009). Miało to bardzo duże znaczenie praktyczne. Wyniki testu OBUT są przechowywane jedynie w szkołach, zwykle w postaci sumy punktów uzyskanych w teście. Aby z wyników analiz mogło korzystać jak najwięcej szkół, należało wybrać taki model, który wymaga znajomości jedynie sumy punktów. Z tego względu nie można też było podczas skalowania usunąć słabszych pod względem psychometrycznym zadań, ani skracać skal dla zadań punktowanych na dłuższych skalach.

Zastosowanie modeli IRT wymaga spełnienia założenia o lokalnej niezależność pozycji. W celu wykrycia grup zadań, które mogą łamać to założenie wykonano analizę skupień zmiennych wokół cechy latentnej (CLV – Vigneau, Qannari 2003). W efekcie po kilka pozycji zarówno w teście OBUT, jak i ze sprawdzianu zostało połączonych, co pozwoliło spełnić założenia modelu.

Wyskalowanie wyniki testów przedstawiono na skali o średniej 100 i odchyleniu standardowym 15 w populacji uczniów.

Schemat analizy

Ocena jakości skal osiągnięć szkolnych

Jakość dopasowania danych do modelu Rascha została zweryfikowana za pomocą miar *outfit* i *infit*. Miary te liczone są dla zadań. *Outfit* jest równie wrażliwa na odstępstwa od modelu dla całego zakresu skali, natomiast *infit* dla tych

przypadków, które znajdują się w okolicy parametru trudności zadania. Jeśli miary te przyjmują wartość 1, uznaje się, że zadanie jest dobrze dopasowane do modelu. Dopuszcza się jednak pewne rozchwianie ich wartości (De Ayala 2009: 55–57), przyjmując przedział wartości od 0,8 do 1,2 jako świadczący o dobrym dopasowaniu.

Jako miary rzetelności wykorzystano alfę Cronbacha oraz współczynnik rzetelności EAP/PV (Adams 2005; Jasińska, Modzelewski 2014). EAP/PV jest jedną z miar rzetelności wykorzystywaną w modelach IRT. Jest to stosunek wariancji oszacowania punktowego (EAP) do całkowitej wariancji zmiennej ukrytej (obliczanej jako wariancja PV – *plausible values*). Obie miary mogą przyjmować wartości od 0 do 1.

W celu przeprowadzenia analizy ewentualnego obciążenia wyników testowych błędem nielosowym związanym z oddziaływaniem różnych czynników na poziomie szkoły (np. różnej motywacji testowej, niedotrzymywaniem reguł realizacji badania lub kodowania prac) zestawiono międzyszkolną wariancję wyników testu OBUT z danymi pokazującymi zróżnicowanie międzyszkolne po III klasie w innych badaniach. Dekompozycji wariancji dokonano wykorzystując dwupoziomowe modele mieszanych efektów, a wariancję międzyszkolną obliczono jako stosunek wariancji efektów losowych związanych z podziałem na szkoły do wariancji całkowitej (Domański, Pokropek 2011). Współczynnik ten mówi nam o tym, jaką część zmienności wyników możemy przypisać podziałowi na szkoły. Większe niż oczekiwane międzyszkolne zróżnicowanie wyników testu OBUT mogłoby świadczyć o istnieniu w szkołach procesów zniekształcających wyniki testowe.

Analiza reprezentatywności próby

Analizę reprezentatywności próby przeprowadzono porównując średnią arytmetyczną oraz odchylenie standardowe wyników uczniów z próby z wynikami populacyjnymi. Dla wyników OBUT wykorzystano dane ze wszystkich szkół, które przystąpiły do badania, a dla sprawdzianu wyniki populacyjne.

Modelowanie EWD

Poziom uprzednich osiągnięć szkolnych jest najważniejszą zmienną niezależną w modelach EWD. Dlatego kluczowe jest dobre dopasowanie funkcji

łączącej wynik „na wyjściu" z wynikiem „na wejściu". W praktyce funkcja ta przeważnie jest krzywoliniowa i z tego powodu modeluje się ją wielomianem odpowiedniego stopnia (Żółtak 2013). W prezentowanych analizach przyjęto analogiczne podejście. Poza wynikiem z testu OBUT w modelu uwzględniono informację o płci i dysleksji. Ponadto przyjęto, że model EWD powinien spełniać dwa kryteria. (1) Funkcja łącząca wyniki sprawdzianu z wynikami testu OBUT powinna być niemalejąca. (2) Wszystkie współczynniki regresji (w tym kolejne potęgi wyników testu OBUT) powinny być istotne statystycznie; przy czym im niższy stopień wielomianu, tym lepiej. Modelowanie przeprowadzone zostało na połączonych danych z próby szkół.

Procedura wyboru najlepiej dopasowanego modelu polegała na przeprowadzeniu szeregu regresji estymowanych metodą najmniejszych kwadratów (MNK), do których wprowadzane były: informacja o płci, dysleksji, wyniku testu OBUT oraz krokowo kolejne jego potęgi. Wyniki każdego modelu były weryfikowane pod względem spełniania przyjętych kryteriów. Regresję MNK wybrano, gdyż daje ona możliwość obliczania wskaźników EWD dla dowolnie zdefiniowanych grup uczniów (np. uczniów jakiejś szkoły) jako średniej z reszt. Dzięki temu użytkownicy metody mogą prowadzić analizy w podziale na różne podgrupy (np. na klasy lub płeć).

Ocena właściwości wskaźników EWD

Weryfikację, czy wskaźniki EWD są niezależne od tego, z jakimi uczniami (o jakich cechach) szkoła pracuje, przeprowadzono licząc współczynnik korelacji między średnim wynikiem testu OBUT w szkole a wartością wskaźnika EWD dla szkoły.

Wyniki

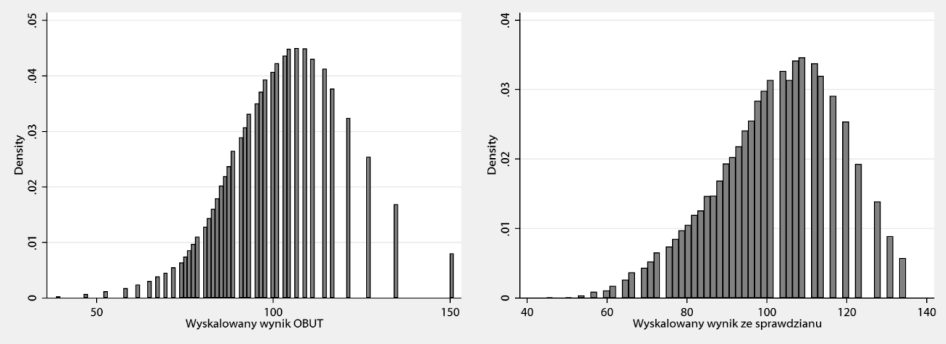
Właściwości wykorzystanych testów

Analiza miar *infit* i *outfit* dla zadań z testu OBUT i sprawdzianu pokazała zadowalające dopasowanie danych do modelu Rascha (statystyki opisowe zamieszczono w Tabeli 1). W przypadku testu OBUT tylko dla dwóch zadań wartości

miar dopasowania wykroczyły poza przyjęty zakres (0,8–1,2), w tym tylko *outfit* dla jednego zadania przyjął wartość odbiegającą od tego zakresu w stopniu znaczącym (1,77). W przypadku sprawdzianu miary dla dwóch zadań w sposób nieznaczny odbiegały od przyjętego zakresu świadczącego o dobrym dopasowaniu. Wyniki te pokazały, że ze względu na kryterium dopasowania skale zbudowane z wykorzystaniem modelu Rascha można traktować jak interwałowe.

Tabela 1. Statystyki opisowe miar dopasowania zadań do modelu

Statystyki	OBUT		Sprawdzian	
	outfit	*infit*	*outfit*	*infit*
Minimum	0,80	0,87	0,77	0,88
1 kwartyl	0,88	0,93	0,93	0,96
3 kwartyl	1,09	1,04	1,04	1,04
Maksimum	1,77	1,39	1,05	1,04



Rysunek 1. Rozkłady wyskalowanych wyników testu OBUT i sprawdzianu

Bardziej problematyczne okazało się założenie o ciągłości skal z uwagi na małą „gęstość" wyników w zakresie wysokiego poziomu osiągnięć. Rysunek 1 przedstawia rozkłady wyskalowanych wyników testu OBUT i sprawdzianu. Duża łatwość testu OBUT spowodowała, że nie różnicuje on dobrze poziomu osiągnięć uczniów zdolnych. Ich wyniki zawierają się w kilku wartościach punktowych, znacznie od siebie oddalonych (zbliżenie kształtu rozkładu wyników do rozkładu normalnego jest pożądaną konsekwencją zastosowanego modelu skalowania). W zakresie wyników wysokich skala ma więc charakter skokowy. W przypadku sprawdzianu analogiczny problem jest widoczny, choć nie jest tak silny. Wyniki

te pokazują, że trzeba zachować ostrożność w interpretowaniu wskaźników obliczonych na podstawie tych danych (zarówno średnich wyników, jak i miar EWD) w szczególności dla szkół pracujących z uczniami o wyższych osiągnięciach szkolnych.

Rzetelność obu miar wykorzystywanych w modelach EWD jest na akceptowalnym poziomie dla potrzeb budowania wskaźników zagregowanych. Alfa Cronbacha dla testu OBUT wyniosła 0,72, a dla sprawdzianu 0,86. Współczynnik rzetelności EAP/PV dla testu OBUT wyniósł 0,85, a dla sprawdzianu 0,86.

Analiza zróżnicowania międzyszkolnego wyników testu OBUT doprowadziła do niepokojących wniosków. Wariancja międzyszkolna tego pomiaru jest znacząco wyższa niż obserwowana w innych badaniach dla pomiarów po III klasie szkoły podstawowej. W przypadku wyników wszystkich uczniów piszących test OBUT, podział na szkoły wyjaśnia prawie 18% zróżnicowania wyników. Na danych z próby wykorzystywanej do modelowania EWD, współczynnik ten wyniósł niemal 20%. Wartości te są znacznie wyższe niż uzyskane w innych badaniach, w których wariancja międzyszkolna osiągnięć na koniec I etapu edukacyjnego wahała się od 8,6% do 13,7% (Dolata i in. 2014; Jasińska, Modzelewski 2014 – por. Tabela 2). Może to przemawiać za hipotezą, że przynajmniej w części szkół uzyskane wyniki są obciążone nielosowym błędem związanym z różną motywacją testową, niedotrzymaniem reguł przeprowadzania lub oceniania testu OBUT w szkołach.

Tabela 2. Wariancja międzyszkolna osiągnięć uczniów na koniec I etapu edukacyjnego

Dane		Odsetek wariancji międzyszkolnej (w %)
OBUT 2011	populacja	17,7
	próba	19,8
Badanie podłużne EWD (2012)	czytanie	8,6[a]
	świadomość językowa	11,8[a]
	matematyka	10,5[a]
PIRLS 2011	czytanie	12,2[b]
TIMSS 2011	matematyka	13,7[b]

Na podstawie: [a] Dolata i in. (2013); [b] Jasińska i Modzelewski (2013).

Reprezentatywność próby

Średnie wyniki uczniów z próby w obu pomiarach okazały się nieco lepsze niż w populacji, a odchylenia standardowe zbliżone (patrz Tabela 3). Można zatem przyjąć, że próba dobrze odzwierciedla wyniki populacyjne.

Tabela 3. Porównanie danych z próby i populacji

Pomiar	Grupa	Średnia	Odchylenie st.
OBUT 2011	populacja	100,0	15,0
	próba	101,7	15,6
Sprawdzian 2014	populacja	100,0	15,0
	próba	101,4	13,7

Model EWD dla II etapu edukacyjnego

Na tak zgromadzonych danych testowano różne modele EWD dla II etapu edukacyjnego. Najlepszym okazał się model, w którym wyniki testu OBUT zostały uwzględnione jako wielomian 5-go stopnia.

Tabela 4. Współczynniki regresji w modelu EWD

Model	Współczynniki niestandaryzowane		Współczynniki standaryzowane	*t*	Istotność
	B	Bł. st.	Beta		
(Stała)	100,069	0,194		516,681	$p < 0,001$
Płeć	1,973	0,223	0,072	8,859	$p < 0,001$
Dysleksja	3,388	0,339	0,082	9,996	$p < 0,001$
OBUT	11,585	0,235	0,876	49,381	$p < 0,001$
$OBUT^2$	–1,130	0,138	–0,151	–8,158	$p < 0,001$
$OBUT^3$	–0,999	0,107	–0,428	–9,308	$p < 0,001$
$OBUT^4$	0,060	0,015	0,081	4,153	$p < 0,001$
$OBUT^5$	0,045	0,008	0,211	5,905	$p < 0,001$

W modelu tym funkcja łącząca wyniki na sprawdzianie z wynikami testu OBUT jest niemalejąca (w przeciwieństwie do modeli z niższą potęgą wielomianu) i wszystkie współczynniki regresji są istotne statystycznie (por. Tabela 4). Dopasowanie modelu do danych jest dobre – R-kwadrat wyniósł 0,45.

Podstawowe właściwości wskaźników EWD

Obliczone z opisanego powyżej modelu wskaźniki EWD skorelowano ze średnim wynikiem testu OBUT dla szkoły. Analiza ta miała pokazać, czy obliczone wskaźniki EWD są niezależne od tego, z jakimi uczniami pracuje szkoła. Niestety stwierdzono istnienie silnego negatywnego związku między tymi zmiennymi – korelacja Pearsona jest równa –0,37. Wynik ten okazał się zaskakujący, bowiem badania właściwości wskaźników EWD dla gimnazjów, liceów i techników wskazują na istnienie odwrotnej relacji (Żółtak 2013). Dla tych szkół obserwuje się słabszy, jednak pozytywny związek średniej egzaminu „na wejściu" ze wskaźnikami EWD (w szczególności w szkołach o wyższych wynikach), a zjawisko to tłumaczy się problemem selektywności szkół. W przypadku testu OBUT zaobserwowany związek może występować w sytuacji, w której w pewnych szkołach wyniki tego pomiaru są zawyżone. W efekcie w placówkach tych trudno o dodatnią EWD, gdyż obliczana jest ona na podstawie nierealistycznie wysokich uprzednich osiągnięć. W powiązaniu z informacją o większej niż oczekiwana wariancji międzyszkolnej wyników tego testu, hipoteza ta wydaje się całkiem prawdopodobna.

Podsumowanie i dyskusja

Głównym celem testów diagnostycznych o niskiej doniosłości jest przekazanie nauczycielom, rodzicom i uczniom profesjonalnej informacji o poziomie osiągnięć szkolnych uczniów. Ich wyniki mogą również być podstawą wartościowych danych o efektach i efektywności nauczania w szkołach, potrzebnych placówkom do świadomego planowania pracy i oceny skuteczności podejmowanych działań. Cele te mogą być jednak zrealizowane tylko wówczas, gdy wykorzystane narzędzia są odpowiedniej jakości a wyniki pomiarów są wiarygodne.

Przedstawione w artykule wyniki pokazały jednak, że wiarygodność danych z testu OBUT może być podważona, przynajmniej w części szkół. Wnioski te

są spójne ze spostrzeżeniami niektórych nauczycieli klas IV–VI szkół podstawowych, którzy: „od początku wyrażali wątpliwości, czy wyniki Ogólnopolskiego Badania Umiejętności Trzecioklasistów są na tyle wiarygodne, że mogą być podstawą do określania wyników przewidywanych na sprawdzianie po szóstej klasie. Małe zaufanie wynikało z przekonania, że w zbyt dużym stopniu na wyniki wpływał czynnik ludzki – przeprowadzanie badania w obecności nauczyciela prowadzącego, sprawdzanie przez nich prac, rywalizacyjna postawa wobec wyników badania, itp." (cytat z raportu z badania jakościowego z jednej ze szkół, za: Pfeiffer 2015: 22). Tak więc pomiar z założenia o niskiej doniosłości okazał się testem wysokiej stawki dla nauczycieli przynajmniej części szkół, co mogło doprowadzić do zniekształcenia danych. Wyniki te pokazały, że przynajmniej niektórzy nauczyciele nie postrzegają diagnozy osiągnięć szkolnych uczniów jako szansy na uzyskanie wartościowych danych o własnej pracy mogących im pomóc w rozwoju zawodowym. Nie czują, że współuczestniczą w procesie zbierania ważnych dla nich informacji. Jest to w ich oczach zewnętrzne, narzucone przez kogoś narzędzie kontroli czy opresji. Postawę taką może nasilać brak pewności własnych kompetencji dydaktycznych lub brak zaufania do przełożonych lub innych odbiorców wyników testowych (rodziców, innych nauczycieli, organu prowadzącego). W takiej kulturze testowania trudno o wiarygodne dane.

Zmiana procedur realizacji tego typu badania i kodowania testów mogłaby poprawić jakość danych. Niektóre szkoły same wyszły z inicjatywą rozwiązań korzystnie przekładających się na trafność pomiaru, zlecając przeprowadzenie badania i sprawdzanie prac innym nauczycielom ze szkoły (zamiast nauczającym dane oddziały) (Pfeiffer 2015). Wiarygodność wskaźników byłaby większa, gdyby zostały one obliczone na podstawie losowej próby szkół, w których zarówno realizacja, jak i kodowanie testów byłyby powierzone osobom z zewnątrz. Rozwiązanie to wprowadzono do organizowanego w 2015 roku przez Instytut Badań Edukacyjnych pomiaru kompetencji trzecioklasistów.

Wydaje się jednak, że dla jakości wyników kluczowe znaczenie ma nastawienie użytkowników testów. Jeśli nauczyciele będą postrzegać je jako narzędzie służące „rozliczalności", a nie jako źródło rzetelnych danych pozwalających na profesjonalne planowanie własnej pracy, nieuczciwości i wypaczenia będą się zdarzały. Warto podkreślić, że na stosunek nauczycieli wpływa w dużej mierze sposób wykorzystania wyników testów przez innych odbiorców. Ważne jest więc zarówno mądre konstruowanie wskaźników, jak i mądry sposób ich wykorzystania.

Wartościowe mogłyby być też badania nad możliwością wykorzystania metod wykrywania oszustw związanych z sytuacją testową (Jacob, Levitt 2003; Karabatsos 2003) do oczyszczenia danych z badania OBUT przed wykorzystaniem

ich do modelowania. Autorzy podjęli takie próby, jednak nie wpłynęły one na zmianę postaci modelu. Pogłębienie tych analiz jest wskazane.

Modele EWD dla drugiego etapu nauczania zostały obliczone z wykorzystaniem prezentowanego w artykule podejścia w ramach prac nad rozwojem metody w Polsce. Szkoły podstawowe mogą korzystać z przygotowanego dla nich narzędzia do analizy wyników[4]. Użytkownicy metody powinni jednak mieć na uwadze to, że daje ona o tyle wartościową informację, o ile pomiar osiągnięć testem OBUT został przeprowadzony w sposób rzetelny i uczciwy.

Bibliografia

ACARA. 2014. *NAPLAN Achievement in Reading, Persuasive Writing, Language Conventions and Numeracy: National Report for 2014*, Australian Curriculum, Assessment and Reporting Authority, Sydney.

Adams R.J. 2005. *Reliability as a measurement design effect*, „Studies in Educational Evaluation", nr 312–313, s. 162–172.

Balázsi I. 2006. *National Assessment of Basic Competencies in Hungary*, Center for Evaluation Studies, dostęp: http://www.iaea.info/documents/paper_1162a1d92d.pdf [08.10.2015].

Ballou D. 2009. *Test Scaling and Value-Added Measurement*, „Education Finance and Policy", nr 4 (4), s. 351–383.

Bąbiak I., Matuszczak K., Zielonka P. 2013. *Raport z ewaluacji wewnętrznej projektu EWD*, Instytut Badań Edukacyjnych, Warszawa.

Chudowsky N., Koenig J.A., Braun H.I. 2010. *Getting value out of value-added report of a workshop*, National Academies Press, Washington.

Centralna Komisja Egzaminacyjna 2011. *Ogólnopolskie badanie umiejętności trzecioklasistów OBUT 2011. Przewodnik przeprowadzającego badanie*, CKE, Warszawa.

Corcoran S.P., Jennings J. i L., Beveridge A.A. 2011. *Teacher Effectiveness on High- and Low-Stakes Tests*, Society for Research on Educational Effectiveness, New York.

De Ayala R.J. 2009. *The theory and practice of item response theory*, Guilford Press, New York.

Dolata R., Hawrot A., Humenny G., Jasińska A., Koniewski M., Majkut P., Żółtak T. 2013. *Trafność metody edukacyjnej wartości dodanej dla gimnazjów*, Instytut Badań Edukacyjnych, New York.

Dolata R., Hawrot A., Humenny G., Jasińska-Maciążek A., Koniewski M., Majkut P. 2014. *Kontekstowy model efektywności nauczania po pierwszym etapie edukacyjnym. Edukacyjna Wartość Dodana*, Instytut Badan Edukacyjnych, Warszawa.

Dolata R., Pokropek A. 2010. *Motywacja a wynik testu z nauk przyrodniczych. Studium na przykładzie PISA 2006*, [w:] *Teraźniejszość i przyszłość oceniania szkolnego: XVI Krajowa Konferencja Diagnostyki Edukacyjnej*, red. B. Niemierko, M.K. Szmigel, Grupa Tomami, Kraków.

[4] http://ewd.edu.pl/pobierz-2/

Dolata R., Szaleniec H. 2012. *Funkcje krajowych egzaminów w systemie edukacji*, „Polityka Społeczna", nr tematyczny 1, s. 37–41.

Domański H., Pokropek A. 2011. *Podziały terytorialne, globalizacja a nierówności społeczne: wprowadzenie do modeli wielopoziomowych*, Wydawnictwo IFiS PAN, Warszawa.

EACEA. 2010. *Ogólnokrajowe egzaminowanie uczniów w Europie: cele, organizacja i wykorzystanie wyników*, tłum. M. Kowalczyk, Fundacja Rozwoju Systemu Edukacji, Warszawa.

Freeman C. 2009. *First national literacy and numeracy tests introduced*, „Research Developments", nr 20 (20), s. 4.

Herczyński J., Borek A., Stożek E. 2012. *Ocena potrzeb szkoły na podstawie osiągnięć uczniów oraz wyników ewaluacji zewnętrznej*, [w:] *Finansowanie oświaty*, red. M. Herbst, t. 3, Biblioteczka Oświaty Samorządowej, Warszawa, s. 171–232.

Hungarian Eurydice Unit. 2009. *National Testing of Pupils in Europe: Objectives, Organisation and Use of Results. Hungary*, Hungarian Eurydice Unit, Bruksela.

Jacob B.A., Levitt S.D. 2003. *Rotten Apples: An Investigation of the Prevalence and Predictors of Teacher Cheating*, „The Quarterly Journal of Economics", nr 118 (3), s. 843–877.

Jager D.J., Merki K.M., Oerke B., Holmeier M. 2012. *Statewide Low-Stakes Tests and a Teaching to the Test Effect? An Analysis of Teacher Survey Data from Two German States*, „Assessment in Education: Principles, Policy Practice", nr 19 (4), s. 451–467.

Jasińska A., Modzelewski M. 2013. *Międzyszkolne zróżnicowanie wyników nauczania po pierwszym etapie kształcenia*, [w:] *Polska edukacja w świetle diagnoz prowadzonych z różnych perspektyw badawczych: XIX Krajowa Konferencja Diagnostyki Edukacyjnej*, red. B. Niemierko, M.K. Szmigel, Grupa Tomami, Kraków, s. 165–178.

Jasińska A., Modzelewski M. 2014. *Testy osiągnięć szkolnych TOS3: przykład narzędzia skonstruowanego z wykorzystaniem modelu Rascha*, „Edukacja", nr 2 (127), s. 85–107.

Jasińska-Maciążek A., Modzelewski M. 2014. *Modele analizy zróżnicowania wyników nauczania*, [w:] R. Dolata, *Czy szkoła ma znaczenie? Analiza zróżnicowania efektywności nauczania na pierwszym etapie edukacyjnym*, Instytut Badań Edukacyjnych, Warszawa.

Karabatsos G. 2003. *Comparing the aberrant response detection performance of thirty-six person-fit statistics*, „Applied Measurement in Education", nr 16 (4), s. 277–298.

Karwowski M. red. 2013. *Ścieżki rozwoju edukacyjnego młodzieży – szkoły pogimnazjalne: trafność wskaźników edukacyjnej wartości dodanej dla szkół maturalnych*, Wydawnictwo Instytutu Filozofii i Socjologii Polskiej Akademii Nauk, Warszawa.

Konarzewski K. 2012. *TIMSS i PIRLS 2011. Osiągnięcia szkolne polskich trzecioklasistów w perspektywie międzynarodowej*, CKE, Warszawa.

MacBeath J., Schratz M., Meuret D., Jakobsen L. 2005. *Czy nasza szkoła jest dobra?*, tłum. K. Kruszewski, Wydawnictwa Szkolne i Pedagogiczne, Warszawa.

Matuszczak K., Wasilewska O. 2015. *Wyniki egzaminów zewnętrznych w pracy szkoły*, [w:] *Egzaminy zewnętrzne w polityce i praktyce edukacyjnej. Raport o stanie edukacji 2014*, red. R. Dolata, M. Sitek, Instytut Badań Edukacyjnych, Warszawa.

McCaffrey D.F., Lockwood J.R., Koretz D.M., Hamilton L.S. 2003. *Evaluating Value-Added Models for Teacher Accountability*, RAND Corporation, Santa Monica, CA.

National Foundation for Educational Research. 2009. *National Testing of Pupils in England, Wales and NorthernIreland*, dostęp: https://www.nfer.ac.uk/shadomx/apps/fms/fmsdownload.cfm?file_uuid=67EAAF91-C29E-AD4D-07F1-A9373EA17105&siteName=nfer.

Nichols S.L., Berliner D. 2007. *The pressure to cheat in a high-stakes testing environment*, [w:] *Psychology of academic cheating*, red. E.M. Anderman, T.B. Murdock, Elsevier Academic Press, Amsterdam–Boston, s. 289–311.

Niemierko B. 2009. *Diagnostyka edukacyjna: podręcznik akademicki*, Wydawnictwo Naukowe PWN, Warszawa.

OECD. 2008. *Measuring improvements in learning outcomes: Best practices to assess the value-added of schools*, Organisation for Economic Co-operation and Development, Paris.

OECD. 2013. *Synergies for better learning: an international perspective on evaluation and assessment*, Organisation for Economic Co-operation and Development, Paris.

Papay J.P. 2011. *Different Tests, Different Answers The Stability of Teacher Value-Added Estimates Across Outcome Measures*, „American Educational Research Journal", nr 481, s. 163–193.

Pfeiffer A. 2015. *Raport zbiorczy z badania uczestniczącego w szkołach podstawowych*, Instytut Badań Edukacyjnych, Warszawa.

Pregler A., Wiatrak E. red. 2011. *Ogólnopolskie Badanie Umiejętności Trzecioklasistów Raport OBUT 2011*, CKE, Warszawa.

Raudenbush S.W. 2004. *What are value-added models estimating and what does this imply for statistical practice?*, „Journal of Educational and Behavioral Statistics", nr 291, s. 121–130.

Raudenbush S.W., Willms J.D. 1995. *The Estimation of School Effects*, „Journal of Educational and Behavioral Statistics", nr 20 (4), s. 307–335.

Reardon S.F., Raudenbush S.W. 2009. *Assumptions of value added models for estimating school effects*, „Education Finance and Policy", nr 4 (4), s. 492–519.

SEO. 2013. *Wymagania państwa wobec szkół*, dostęp: http://www.npseo.pl/action/requirements/wymagania_panstwa_wobec_szkol.

Sinka E. 2010. *OECD Review on Evaluation and Assessment Frameworks for Improving School Outcomes. Hungary. Country Background Report*, Hungary, dostęp: http://www.oecd.org/edu/school/50484774.pdf.

Skórska P., Świst K., Pokropek A. 2014. *Indywidualne i grupowe efekty motywacji testowej uczniów*, [w:] *Diagnozy edukacyjne: dorobek i nowe zadania: XX Krajowa Konferencja Diagnostyki Edukacyjnej*, red. B. Niemierko, M.K. Szmigel, Grupa Tomami, Kraków, s. 161–171.

Stożek E. 2012. *Czy egzaminy mogą pomóc szkole w rozwoju?*, „Polityka Społeczna", nr tematyczny 1, s. 30–33.

Szaleniec H. 2004. *Jak wykorzystywać wyniki egzaminów zewnętrznych*, WSiP, Warszawa.

Szaleniec H. red. 2009. *Teoria wyniku zadania IRT: zastosowania w polskim systemie egzaminów zewnętrznych: praca zbiorowa*, CKE, Warszawa.

Vigneau E., Qannari E.M. 2003. *Clustering of Variables Around Latent Components*, „Communications in Statistics – Simulation and Computation", nr 32 (4), s. 1131–1150.

Wasilewska O., Rybińska A., Muzyk A. 2014. *Wykorzystanie ewaluacji zewnętrznej i wewnętrznej przez szkoły*, Instytut Badań Edukacyjnych, Warszawa.

Wise S.L., DeMars C.E. 2005. *Low Examinee Effort in Low-Stakes Assessment: Problems and Potential Solutions*, „Educational Assessment", nr 10 (1), s. 1–17.

Wolf L.F., Smith J.K. 1995. *The Consequence of Consequence: Motivation, Anxiety, and Test Performance*, „Applied Measurement in Education", nr 8 (3), s. 227–242.

Żółtak T. 2013. *Statystyczne modelowanie wskaźników edukacyjnej wartości dodanej – podsumowanie polskich doświadczeń (Raport tematyczny z badania)*, Instytut Badań Edukacyjnych, Warszawa.

KWARTALNIK PEDAGOGICZNY
2015 NUMER 4(238)

COLLOQUIA

Bolesław Niemierko[1]
SWPS Uniwersytet Humanistycznospołeczny, Wydział w Sopocie

Psychologists as educational diagnosticians. On the track of typology

Streszczenie

PSYCHOLOGOWIE W ROLI DIAGNOSTÓW EDUKACYJNYCH. NA TROPIE TYPOLOGII

Celem tego artykułu jest przedstawienie założeń teoretycznych i prób empirycznych dotyczących postaw, jakie przyjmują psychologowie pełniący funkcje diagnostów edukacyjnych. Zawiera on także próbę wyjaśnienia związku tych postaw ze stanami Ja ukształtowanymi we własnym dzieciństwie i w późniejszych doświadczeniach społecznych diagnosty.

Diagnostyka edukacyjna jest teorią i praktyką rozpoznawania warunków, przebiegu i wyników uczenia się. Obejmuje nieformalne i standaryzowane procedury usuwania zaburzeń w uczeniu się oraz zaspokajania innych potrzeb rozwojowych dzieci, młodzieży i dorosłych. Od psychologów oczekuje się, że pokierują pedagogami w zdobywaniu wiedzy o zdolnościach, motywacji i osiągnięciach uczniów.

Na podstawie typologii umysłów zaproponowanej przez Czesława Nosala zbudowano typologię diagnostów edukacyjnych obejmującą cztery typy: proceduralistę, teoretyka, ryzykanta oraz intuicjonistę. W fenomenologicznych i kwestionariuszowych badaniach studentów psychologii potwierdzono występowanie tych czterech typów diagnosty, a także wykryto ich związki z niektórymi stanami Ja określonymi przez Eryka Berne'a w teorii analizy transakcyjnej.

Słowa kluczowe: diagnostyka edukacyjna, diagnosta edukacyjny, student psychologii, analiza transakcyjna.

[1] I am deeply indebted to two anonymous reviewers of my paper. Their remarks allowed me to remove some errors and omissisons they found in its previous version.

Educational Diagnostics

Educational diagnostics will be conceived here as the theory and practice of sound recognizing of (1) environments, (2) processes, and (3) outcomes of learning, in particular scholastic learning. The three targets of diagnostics are equally important. Research and applications may be focused on all or on any of them: environment, process or product of learning, not only at schools but also outside, under any circumstances and to whoever's benefit.

Broadly comprehended educational diagnostics embraces many elements of (1) *medical diagnostics*, dealing with students' organic development, their health preservation and illness risks, (2) *psychological diagnosis*, covering abilities, emotional characteristics, attitudes and aspirations, (3) *sociological diagnosis*, inquiring into expected and performed students' social roles, and (4) *historic diagnosis*, pertaining to the students' biographies, their individual and generation--bound experiences (Niemierko 2009).

According to Stefan Ziemski (1972), every *full diagnosis* contains five segments:

1. *Typological diagnosis*, categorizing individuals into subgroups of ability, achievement, home environment, culture, study skills, etc.
2. *Genetic diagnosis*, registering events connected to the present situations with alleged causal relationship. For example, a teacher may learn students' personalities going through their biographic documents.
3. *Meaning diagnosis,* which consists in identifying previous events and present attitudes that could influence current learning processes. Unconscious anxiety, resentments and constraints may be harmful to learning new behaviours.
4. *Stage diagnosis*, applying a developmental model to student characteristics and behaviour. The model may be very general, like the Piaget's (1924) model of mental development and the Kohlberg's (1976) model of moral development, or may be quite specific, based on subject-matter didactics and teacher observations of learning.
5. *Prognostic diagnosis*, being in education a prediction of learning results under certain conditions. It often begins: "Assuming your further consisted attempts, you will probably achieve..." It is important not to deform prognoses into effort-exempting prophecies.

The first question to be asked about recognizing school learning is whether we are going (1) to perform it in a participant way, not disturbing and/or changing

the usual course of teaching/learning process, or (2) to arrange quasi-experimental situations in which main context variables are controlled and ready-made measurement tools are applied. Accordingly, we may distinguish (1) the *informal*, teacher-performed educational diagnostics, and (2) the *standardized*, professional educational diagnostics. Both are indispensable to competent managing of contemporary educational systems.

The second question to educational diagnostics is whether it is designed (1) to deal with learning disabilities and disorders or (2) to advance regular learning in a normal situation. To treat illnesses or to establish healthcare? We will refer to the former in (1) *disorder diagnostics*, and to the latter in (2) *developmental diagnostics*. Generally, breakdowns need more specialized consultants than everyday monitoring of student growth but both diagnostic domains require theory and research.

Plotting the two divisions of diagnostic procedures we arrive to the classification presented in Table 1.

Table 1. Four paradigms of educational diagnostics

Aim \ Method	Informal diagnostics	Standardized diagnostics
Recognizing disorders	Informal intervening diagnostics	Standardized intervening diagnostics
Monitoring development	Informal developmental diagnostics	Standardized developmental diagnostics

Informal intervening diagnostics is aimed at undisturbed class management. Its application is almost exclusively based on pre-service training and professional experience of teachers. According to Good and Brophy (1987), every day, a teacher interacts with individual students as many as 1,000 times, and she evaluates students' behaviour at least 90 times. She has to make many managerial decisions not having enough time to gather sufficient information about the cases of disturbance and the troublemakers. Psychologists can assist teachers in some more challenging activities.

Standardized intervening diagnostics is focused on learning prerequisites. It is conducted either by school psychologists or by institutions of educational counselling where psychologists constitute the majority of staff. They use tests, questionnaires and observation techniques to detect a child's innate and acquired cognitive skills and to recommend teaching content and methods suitable for improving the child's school achievement.

Informal developmental diagnostics emphasizes systematic monitoring achievement progress of both the slower learners and the faster learners. The rate of learning and the students' attitudes toward the school subjects in which they attain some achievement criteria are the focus of attention in this paradigm. Entries, processes, and outcomes of learning are equally important in this approach. Looking at, and in consequence seeing variations in student motivation and effort appears to be the key factor of educational success (Good & Brophy 1987).

Standardized developmental diagnostics is intended for learning support by providing students and teachers with precise information on learning outcomes. It needs professional measurement tools but may also be performed at the classroom level, which leads some theorists to the idea of "classometry" as a separate domain of psychometric methods (Benson 2003). Standardized diagnostic methods are based on measurement, mainly of its *norm-referenced* type, where student score is interpreted by means of population score distribution. The *criterion-referenced* tests, in which a student score is interpreted according to educational standards, prove less useful in standardized diagnostics as they are too subjective in standard setting procedure and may tell us more about the standard developers than on the students themselves (Hambleton 1994). However, both types of educational measurement are firmly based on psychometric theory and governed by interdisciplinary standards (*Standards*... 2014).

Educational Diagnosticians

As educational diagnostics comprises both informal and standardized procedures to help out learning disorders as well as general developmental needs of children, youth, and adults, cooperative work of teachers, pedagogues, and psychologists is necessary. Psychologists are supposed to guide collective judgments on the learners' aptitudes, intensions, and gains, particularly when cases of maladjustment are at stakes. It results in demand for psychologists inclined to perform educational service.

A variety of attitudes toward diagnostic methodology and educational advisory styles should be expected among psychologists. Following some concepts of the well-known Myers-Briggs Type Indicator (Myers 1962), Czesław Nosal presented his own general classification of *mental constitutions*. It contains four categories (Nosal 1990, 1999):

1. *Objective-concrete type* – directed to facts, observations, measurement, precise reasoning.
2. *Objective-global type* – directed to theories, models, generalities, scientific laws.
3. *Subjective-concrete type* – directed to needs, experience, experiments, personal opinions.
4. *Subjective-global type* – directed to feelings, half-hidden desires, syntheses, philosophies.

It may be assumed that mental characteristics are parallel to learning styles of diagnosticians. Wincenty Okoń's (1967) general classification of learning ways shaped by Bolesław Niemierko (2009) turned into the following:

Form *alpha*. *Learning by acquiring*, which leads to *know-what* knowledge, reproduction of existing information on the world outside.

Form *gamma*. *Learning by discovering*, which leads to *know-why* knowledge, theory building and verification, looking for general truths.

Form *beta*. *Learning by doing*, which leads to *know-how* knowledge, solving real-world problems and performing serviceable activities.

Form *delta*. *Learning by experiencing*, which leads to *know-who* knowledge, emotional gains, value clarification, personal commitment.

Lower secondary school student preferences in a nation-wide representative study were distributed 30% – 40% – 15% – 15% (Niemierko 2009) but, unfortunately, the proportion of alpha learners grows in the upper secondary school (lyceum) and the proportion of gamma-learners diminishes. Moreover, the tendency to learning by acquiring prevails in the majority of European universities (Okoń 1971).

In his handbook of psychological diagnostics, Władysław Paluchowski (2001) applied Nosal's typology to distinguish four types of psychological diagnosticians:

Type 1: the unemotional logician persistently looking for particular facts, precise observations, and measurements, generalizing them into consistent and completely justified categories.

Type 2: the diagnostician who uses the smallest possible number of assumptions, productive models, and general rules to explain recorded phenomena and to include them into certain determinative categories.

Type 3: the diagnostician sensitive to emotional interactions, disposed to taking immediate hazardous but positive and useful actions in order to gather necessary information.

Type 4: the impressionist who strives for deep understanding of personally appreciated problems and valuing his/her own feelings higher than particular facts.

The above-raised four types of educational diagnosticians will be given here the following working labels:

1. *Proceduralist*
2. *Theoretician*
3. *Gambler*
4. *Intuitionist*

These names will be used in my reports on the empirical try-outs of the typology.

Qualitative study

In the course of educational diagnostics, 194 second and third-grade psychology students were tested with several short stories on real-world educational problems. One of them sounded (in translation into English) as follows:

Creative work # 4

A student of education has brought to a college seminar the poem written by a twenty-year old prisoner during the theatrical-literary classes in a house of correction for juvenile delinquents:

Empty walls tightened
Sorrow that I wasn't right
And there's your view beloved
Though we are out of sight
I wish I felt your lips
And saw your lovely eyes
But I've got loneliness
And sufferings of nights
I will survive it all
And overcome despair
Just stand by me, stand by me
Then I'll manage the way

Describe this prisoner's social adjustment in the house of correction and his emotional maturity. As the psychological protector of the prisoner propose three questions in order to help him in the process of socialization.

The students' positions on the problem appeared to pretty clearly classifiable according to the previously discussed typology:

1. *Proceduralists* (about 15%) complain of being deprived of the prisoner's judicial and psychological documentation. Any supposition on his social adjustment and emotional maturity they count as precluded from consideration before seeing him and making a conversation.
2. *Theoreticians* (about 30%) reflect on the concepts of social adjustment and emotional maturity and find some indexes of the concepts in lines of the poem. However, they warn against general conclusions before the propositions are verified.
3. *Gamblers* (about 15%) pay their attention to the fact of the prisoner's attendance to the theatrical-literary classes. It means for them that his socialization is in progress and his future after leaving the house of correction becomes promising.
4. *Intuitionists* (about 40%) notice that the poem expresses deep and candid feelings. The prisoner lacks support in the correction house and he depends upon a person who stays outside and her connection with him remains uncertain.

In this exercise, the global reasoning manner represented by theoreticians and intuitionists prevail over the concrete ones, procedural and gambling interpretations. It could be owed to academic circumstances of the inquiry. The students tried perhaps to exhibit their general psychological competence rather than just find the course of action to help the prisoner solve his existential problems.

This kind of preliminary, qualitative and highly subjective studies based on exemplary real-world educational problems confirmed our capacity for identifying the four above-distinguished diagnostic attitudes. To analyse their dependence on inborn and environmental variables more structuralized research methods were needed.

Quantitative study

"The Educational Diagnostician" inventory consisting of four ten-item subscales has been constructed according to Paluchowski's outlines. Thirty items were scored with Likert's 1 – 5 scale and ten reversed items were scored 5 – 1.

When reliability coefficients of subscales appeared intolerably low ($\alpha \approx 0.4$) item analysis was performed but it resulted only in slight improvement ($\alpha \approx 0.5$) of reconstructed subscales. Obviously, the content range of indexes for particular scales should get narrowed and if it is not sufficient for a decent reliability, they should be enlarged to at least 15 items.

In spite of the reliability inadequate for individual diagnosis some general regularities of the domain may be taken into consideration. They will aid construct validity of the typology and direct future research on the topic.

The main descriptive statistics of "The Educational Diagnostician" questionnaire study in a sample of 50 psychology five-graders are presented in Table 2.

Table 2. Descriptive statistics of psychology students' scores in "The Educational Diagnostician" questionnaire

	Proceduralist	Theoretician	Gambler	Intuitionist
Mean	34.6	34.2	32.1	32.2
Variance	17.9	14.7	25.5	21.3
Modal choice*	20	16	12	13

* Number of students who selected this type as their favourite or one of two equally preferred.

As we can see in Table 2, average acceptance to objective types of educational diagnosis, procedural and theoretical, is slightly ($p = 0.98$) higher with psychology students than their acceptance to subjective methods, by trial-and-error and intuitive, and first choices of objective types are also more frequent. It may be interpreted as a signal that psychologists are inclined to adopt an objective manner of educational diagnosing.

The four subscales were low positively intercorrelated (0.13–0.22) with exception of gambler – intuitionist (0.32, $p > 0.95$) and proceduralist – intuitionist (–0.18).

Transactional Analysis

Transactional analysis is a personality theory and psychotherapy serving personal development of humans (Berne 1961). Its basic model of *ego states* is the following:

1. *The Parent* who frequently and automatically fills the position of initiator and decision-maker responsible for activities of others being treated as children. This attitude is learned from one's own parents or from persons taking action of parents.
2. *The Adult* who is rational, autonomic, self-governing, who understands the reality of situations and maintains partnership relations. This attitude is neither influenced by parents nor a repetition of childish behaviours and feelings.
3. *The Child* is governed by uncontrolled emotions and yields to those who fill the position of parents. This attitude reproduces the spontaneous behaviours, thoughts, and feelings of childhood, when nature and people were originally got to know.

The basic typology was presently multiplied into plurality of subcategories in which a kind of fractal phenomenon (self-similar pattern that replicates at diminishing scale) plays the crucial part. Accordingly, the Parent in the Parent, the Adult in the Parent, the Child in the Parent, etc. is recognized and characterized (Jagieła 2012). From among the nine subcategories the following four were selected as positively correlated with particular types of educational diagnosticians in preliminary studies (Niemierko 2014):

1. The Adult in the Parent, called also the *Practical Parent*, who is pragmatic, operative, and efficient in actions imitating the ones performed once by parents and then by knowledgeable and significant persons.
2. The Child in the Parent, called also the *Protective Parent*, who is attentive, careful, tutelary, and helpful, inclined to become involved in nurturing processes.
3. The Parent in the Adult, called also the *Ethical Adult* or Ethos, who learned and transformed values and capacities from all the sources recognized as trustworthy.
4. The Adult in the Child, called also the *Creative Child* or the Little Professor, who represents original strategies of exploration the reality and of problem solving.

Coefficients of correlation (Pearson r) between four types of diagnosticians and four ego states in a joined group of second, third and fifth-grade psychology students ($n = 244$) are presented in Table 3.

Table 3. Correlations between types of educational diagnosticians and selected ego states

	Practical Parent	*Protective Parent*	*Ethical Adult*	*Creative Child*
Proceduralist	0.34**	0.06	0.40**	–0.14*
Theoretician	0.24**	–0.05	0.17**	0.22**
Gambler	0.01	0.15*	–0.12	0.15*
Intuitionist	0.07	0.35**	0.12	0.18**

* $p > 0.95$; ** $p > 0.99$.

Ego states of students were measured with the *Egogram* inventory created in the Research Group on Educational Transactional Analysis in Jan Długosz Academy in Częstochowa (Pierzchała & Sarnat-Ciastko 2012). Its application to psychology students brought moderate subscale reliability coefficient of about 0.6, which additionally lowered the indexes of relationship. However, the following remarks seem worth undertaking:

1. The *objective* sets, procedural and theoretical, correlate with the Practical Parent and the Ethical Adult ego states. Respective psychologists have deep inclination to adopt psychometrical techniques learned in professional training.
2. In their research strategies and hardy problem solving, Theoreticians display some characteristics of the Creative Child. It may prolong diagnostic processes but make them more flexible and sometimes more valid.
3. Proceduralists explicitly avoid behaviours of the Creative Child. Their resemblance to the Practical Parent and the Ethical Adult seems to be stronger, more rigid, and more consistent than with Theoreticians.
4. The *subjective* sets, hazardous and intuitive, correlate with the Protective Parent and the Creative Child ego states. Respective psychologists are tender, receptive, and gentle. They probably base their findings on free observations rather than on exact procedures.
5. Intuitionists have probably ($p = 0.95$ for the difference between correlation coefficients) more attributes of the Ethical Adult than Gamblers. They are also more protective ($p = 0.95$) than gamblers who look for the shortest way to educational success.

All the generalizations based on Table 3 are only hypothetical. They should be proved in broader and better structuralized studies.

Conclusion

Advantageous changes in educational systems make psychologists more and more significant participants of educational activities. Their most weighty contributions to the students' development processes are diagnosis and psychotherapy.

In a quality of educational diagnostician a psychologist may take up the position of (1) a technique-oriented, conscientious specialist or (2) a human-oriented, empathic tutor. Both attitudes were spotted among psychology students.

A psychologist may be mainly concerned with (1) the particular case to assist and (2) the global human needs to be satisfied by educational processes. These attitudes were also identified in psychology student questioning.

Crossing the two dichotomies we obtain the typology used in the study reported in this paper. The four types of educational diagnosticians were called: the Proceduralist, the Theoretician, the Gambler, and the Intuitionist.

Low reliabilities of scales measuring the particular types forbid us from making more categorical statements than some suggestions about:

(1) usefulness of the approach for the studies on psychologists who perform educational diagnoses,
(2) perspectives on improving pre-service and in-service training of psychologists and pedagogues to function in educational diagnostics.

Some of the differences are for a certainty anchored in home and environment education in the early childhood of diagnosticians. Further studies should explain their significance for the communicative transactions with children and youth.

References

Benson J. 2003. Editorial. *Educational Measurement: Issues and Practice*. Special Issue: Changing the Way Measurement Theorists Think about Classroom Assessment, *4*.

Berne E. 1961. *Transactional Analysis in Psychotherapy. A Systematic Individual and Social Psychiatry*, Grove Press, New York.

Good T.L., Brophy J.E. 1987. *Looking in Classrooms*, Harper & Row, New York.

Hambleton R.K. 1994. *The Rise and Fall of Criterion-Referenced Measurement?*, "Educational Measurement: Issues and Practice", vol. 13, 4, pp. 21–26.

Jagieła J. 2012. *Słownik analizy transakcyjnej*, Wydawnictwo Akademii im. Jana Długosza, Częstochowa.

Kohlberg L. 1976. *Moral stage and moralization: The cognitive-developmental approach*, [in:] *Moral Development and Behavior: Theory, Research, and Social Issues*, ed. T. Lickona, Holt, New York.

Myers I.B. 1962. *The Myers-Briggs Type Indicator Manual*, Educational Testing Service, Princeton.

Niemierko B. 2009. *Diagnostyka edukacyjna. Podręcznik akademicki*, Wydawnictwo Naukowe PWN, Warszawa.

Niemierko B. 2014. *Cztery typy diagnostów edukacyjnych*, [in:] B. Niemierko & M.K. Szmigel, *Diagnozy edukacyjne. Dorobek i nowe zadania*, PTDE, Kraków.

Nosal C.S. 1990. *Psychologiczne modele umysłu*, PWN, Warszawa.

Nosal C.S. 1999. *Psychologia decyzji kadrowych*, Wydawnictwo Profesjonalnej Szkoły Biznesu, Kraków.

Okoń W. 1967. *Podstawy wykształcenia ogólnego*, Nasza Księgarnia, Warszawa.

Okoń W. 1971. *Elementy dydaktyki szkoły wyższej*, PWN, Warszawa.

Paluchowski W.J. 2001. *Diagnoza psychologiczna. Podejście ilościowe i jakościowe*, Wydawnictwo Naukowe Scholar, Warszawa.

Piaget J. 1953. *The Origins of the Intelligence in Children*, International University Press, New York.

Pierzchała A. & Sarnat-Ciastko A. 2012. *Egogram strukturalny. Kwestionariusz samooceny stanów Ja*, "Edukacyjna Analiza Transakcyjna", 1.

Standards for Educational and Psychological Testing. 2014. American Educational Research Association, Washington.

Ziemski S. 1972. *Problemy dobrej diagnozy*, Wiedza Powszechna, Warszawa.

KWARTALNIK PEDAGOGICZNY
2015 NUMER 4(238)

Michał Daszkiewicz
Uniwersytet Gdański

Learner's utterance as a foundation of *personal-experience*-oriented classroom diagnostics

Streszczenie

WYPOWIEDŹ UCZNIA JAKO PODSTAWA DIAGNOSTYKI NIEFORMALNEJ UKIERUNKOWANEJ NA DOŚWIADCZENIE OSOBISTE

Artykuł jest próbą odpowiedzi na pytanie, w jaki sposób (ustne i pisemne) wypowiedzi uczniów mogą być wykorzystane do rozpoznania tego, jak uczniowie rozumieją treści przedmiotowe i jak osobiście doświadczają edukacji. Tekst wychodzi od przedstawienia *uczenia się* oraz *egzaminu* – tj. dwóch składników konstytutywnych *diagnostyki edukacyjnej* – jako dwóch uzupełniających się składników *osobistego doświadczenia edukacyjnego*, silnie uzależnionego do zawartego w nich (rozumianego i wykorzystywanego przez ucznia) języka. W celu wyjaśnienia znaczenia kwestii językowych w tworzeniu solidnej diagnostyki szkolnej, tekst omawia zbiór „prawd" leżących u podstaw analizy wypowiedzi uczniowskich i tym samym determinujących komunikację między nauczycielem i uczniem. Następnie przedstawione są argumenty przemawiające za tym, iż przyszłość diagnostyki edukacyjnej zależy w znaczącym stopniu od tego, jak pomyślnie uda się wypełnić lukę między diagnostyką a lingwistyką. Pokazane jest to za pomocą pytań, które „stoją" na skrzyżowaniu tychże dwóch dyscyplin i które wymagają znalezienia odpowiedzi. Tekst zamykają: (a) wnioski wyciągnięte z przykładowych wypowiedzi uczniów na temat egzaminów i procesu uczenia się oraz (b) wskazania dotyczące pozyskiwania danych z tego, co autor nazywa tu *językowymi zdarzeniami edukacyjnymi* (stanowiącymi jednostkę analizy komunikacji szkolnej).

Słowa kluczowe: wypowiedź ucznia, diagnostyka edukacyjna, językowe zdarzenie edukacyjne, osobiste zdarzenie edukacyjne.

1.1. HUMAN STRATUM: Personal Educational Experience

It is the *personal experiencing of education* that justifies the very existence of the discipline of educational diagnostics. Meant to describe and explain the learner's situation[1], this consistently developing discipline has essentially been focused on two types of his or her experience: learning and examinations. Whilst the former has long been – as far as the personal dimension of education goes – associated with and perceived through such issues relating to the learner as learning styles, motivation to learn, intelligence, or the context of learning, the interest in the latter has directed educators' attention to such learner-related concepts as test wiseness, anxiety, patterns of achievements, or – consistently analysed – self-assessment.

What, most strikingly, proves to be missing, though, when one studies the scope of issues addressed over the last twenty years in Poland, is the discipline's emphasis on language. Both the constitutive components of the personal educational experience mentioned above rest strongly on (the learner's mastery, i.e. understanding and application of) language. To put it in the simplest way possible, if the language of learning or examinations is familiar to learners, the personal experiencing of them has every chance of becoming beneficial and learner-friendly, whereas if the language departs from the learner's experiencing of reality, the two types of experience are bound to prove detrimental to the learner's development. Hence, if the language itself constitutes such a significant common denominator of learning and examinations, it does merit a more prominent position in educational diagnostics, which this paper is yet one more attempt to signalize[2].

Beyond every utterance produced by the learner there lies not only knowledge and the ability to say things, but also (largely personal) beliefs and convictions that a learner happens to hold. First, s/he may find it not worthwhile articulating educational content, or, second, s/he may view the learning of a particular school subject as not boiling down to speech (the example of which

[1] The scope of discipline has been defined on the website of the Polish Association of Educational Diagnostics: http://www.ptde.org/old/html/potrzeba_PTDE.html (with cognitive and emotional aspects lying at its basis); due to the fact that its scope encompasses various learning contexts, this paper consistently refers to 'learners' (rather than 'pupils' or 'students'). The website opened on 2 April 2015.

[2] This text can be seen as both a digest and a complementation of its author's conference papers, addressing analyses of language used by school students and teachers: Daszkiewicz (2007, 2008). With this path of educational diagnostics being represented by few other texts, the said papers are referred to a number of times.

was presented in one of the author's earlier publications, with a learner viewing civics as a discipline requiring him to memorise concepts and terms related to the functioning of a civic society, which, needs no articulation to be performed (Daszkiewicz 2008: 329)). Such views on whether it is worth or necessary to utter subject matter does not need to be realized to play an important function, and so, quite undeniably, they constitute an underlying component of anything that learners learn. Needless to say, the differences between learners and between their approach to subject matter also pertain to how scientific or colloquial language they use, which implies that e.g. educational measurement tools (intended to be objective) will frequently fail to catch learners' (subjective) verbalization of issues. Although the causes of differences between learners' language lie beyond the scope of this text, it is worth noting here that, to refer to Bruner's rationale, their language is in each case differently 'scaffolded' by parent-child interactions.

What follows from the above is that due to the use of language some moments in the course of learning acquire a special status, meaning those when the learner happens to articulate (and/or write) specific subject matter, either because s/he chooses to do so or is somehow prompted by circumstances or other participants of the learning environment to speak. These moments have been coined by the author as '**language educational events**' and they will be presented below as a unit of analysis that may help educational diagnostics advance on the stratum of its linguistic aspects. A better recognition of these moments (events) brings us closer to diagnosing the (degree of the) learner's creativity, which is a direct consequence of, as Krzysztof Mudyń puts it, the language being a social entity internalized by an individual in the course of ontological development, which sets constraints to the individual's efforts at creativity not only from the outside but also from the inside (Cf. Mudyń 1997: 79). To refer here to Vygotsky's rationale of the social character of language, we can also observe that the language used in an educational event may or may not fall within the learner's zone of proximal understanding rendering the event less or more personal (and, *nota bene*, social at the same time).

1.2. HUMAN STRATUM: Teacher-Learner Communication

Language educational events, as defined above, are crucial to teacher-learner communication: what the learner's skills are determines the teacher's reception

of the learner's mastery and understanding of the subject matter – regardless of how valid that reception might happen to be. This aligns with J.F. Fourny's statement noting that "speech is the only window through which the physiologist can view the cerebral life": without this "window" the teacher's attempts to "see through" the learner will be to no avail. This being the case, the key aspiration of educational diagnostics consisting in recognizing the learner's position (with regard to the subject matter learnt) can only be fulfilled with an ample collection of his or her (spoken or written) utterances. Only with a sufficient load of the learner's speech are we able to observe, for instance, whether the learner herself/himself is capable to applying (and thus understanding) what Piaget coined figurative thought (conveyed through figurative language).

With the learner's utterances never fully reflecting what is in his or her mind, though, there is a need to construe learner talk through the distinction between the latent and the observable. These two categories have long been used in methodology (e.g. in item response theory, otherwise called probabilistic theory) to designate hidden mental constructs, on the one hand, and their observable consequences, on the other[3]. If we apply this distinction to the learner's (latent) knowledge and his/her (audible/visible) speech, we may note – among others – the following consequences:

(a) learners with the same knowledge may articulate it differently, meaning in an either less or more communicative manner,
(b) learners with different knowledge may articulate the same sentences, which implies the need for recognition of how linguistically skilled all learners are[4].

It is rather obvious that the teachers' comments formulated with reference to learners need to take this distinction into account as otherwise their comments prove but over-simplistic. Let's consider then how justifiable a selection of teacher commentaries will prove to be if we apply the latent-observable lenses.

> Commentary 1. *When I'm reading Anne's assignments, I can see she thinks properly. She describes geographical phenomena in a way typical of her: full of technical expressions, but at the same time such that her individual or even emotional approach to the issue is demonstrated. It's good because it shows that Anne is learning lots of new things. Sometimes you can have an impression that the description formulated by Anne has been constructed by a person that deals with geography on a daily basis*.

[3] For example, in: Konarski (2004).

[4] This approach can be seen as one of key communication strategies in the classroom, as presented in Twardowski (2005: 26).

Commentary 2. *What Luke writes and says draws one's attention to his eloquence and high level of intelligence. What is particularly striking is that Luke presents historical knowledge in his own way. Doing so, he uses very few terms and names which occurred during lessons, but his involved form of expression points to correct understanding of the sequence of events. In Luke's assignments it's always possible to find an interesting presentation of historical issues and I never get bored when reading his works*.

Commentary 3. *Ola has big gaps in maths. If she is to explain how she's solved a task, she gets lost. Even when the teacher praises her because the result is fine, and asks her to explain it to other learners, she cannot do that. She thinks and mutters something to herself, but it doesn't look like a clear mathematical explanation. What helps her is participation of other learners in the talk, or situations when she's helping on a task in a team, but then one has the impression that she uses her friends' ideas, not her own. And it is the way of solving the problem that matters. She knows the result, but she always says that she's short of time. She's up to something, but it's not quite clear what the target is*.

Commentary 4. *Sławek's statements don't belong to the desirable ones and they don't merit a high grade. The chaos in which they are formulated signalizes chaos in the information he has acquired. Civics is a subject in which one needs to be able to express oneself properly and to present one's views, and this is exactly what is missing in what Sławek writes or speaks. A conclusion that one draws when reading or listening to Sławek is that he learns hardly anything about civics and he only relies on his memory. It's rather thoughtless of him*.

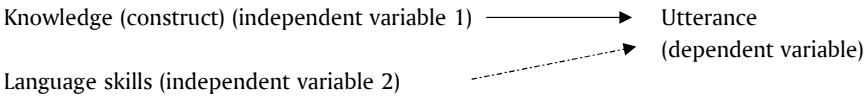
What is clear in all these four situations is that the teacher's idea of the learner's knowledge (which we may view as the *what* of subject matter expressed by learners) is strongly shaped by his or her individual behaviours, personal reasoning and unique position with regard to the use of language (which, on the other hand, we may see as the *how* of that subject matter expressed). Needless to say, whilst the thematic content of utterances will vary across school subjects, the learner's language practices and habits seem to be more fixed and not easily fall subject to modifications, which follows from the simple fact that all learners (or people, in general) have become accustomed to certain structure and lexis prior to any form of school learning.

It may thus be argued that if all the situations referred to by the teachers above were treated by them as language educational events, that is seen as moments when their learners demonstrate individual language-based behaviours (rather than directly unravel their personal knowledge), these situations would be much easier to interpret by teachers, that is make it easier for them to distinguish between the what and the how of what they say and/or write, and thus be better informed about the learner's competence in any particular subject or issue studied.

2.1. THEORETICAL STRATUM: Bridging the gap between linguistics and diagnostics

With the learner's language skills determining the teacher's perception of the learner's knowledge on practically every school subject, the theory of what we have coined here language educational events calls for rapid advancement so that the gap between diagnostics and linguistics would be effectively bridged. By emphasizing the special status of these events we address the following three issues at a time: we (1) acknowledge the fact that not all teachers, to say the least, are "linguistically reflective" enough to notice that they fall subject to language-caused false impressions; (2) give prominence to those moments in which there appears to be more contact with the learner's mind, who, quite undeniably, is also learning at the point when s/he is being quiet and the thoughts remain unuttered, and (3) foster validity of educational diagnostics in that reflection is undertaken on what the utterances made, collected and jointly analysed signify.

The case of the teacher assuming that a given utterance unambiguously reflects the learner's knowledge (as if the two could ever be equated) can be regarded as a common instance of the attribution error (understood as "a tendency to overestimate internal factors, attributable to a disposition, and to underestimate the role of external factors/situations" (Ross 1977: 173)). This simplified view can be represented graphically in the form a relationship between independent variables (evading the teacher's attention) and the dependent one (taking the shape of a learner's utterance):



The attribution error in question can take a different form and be either beneficial or detrimental to the learner, depending on how familiar the learner is with the subject matter and how well s/he demonstrates it by means of language[5]:

Table 1. Four types of knowledge-language relationship and respective attribution errors

Learner	Type represented by the learner	Direction of influence of language skills	Variant of basic attribution error	Impact for the learner	Possible label
L1	knows a lot and speaks/writes nicely	inflating	angel halo effect	beneficial	'walking encyclopedia'
L2	knows little but speaks/writes nicely	inflating	leniency effect	beneficial	'beats about the bush'
L3	knows a lot but speaks/writes poorly	depressing /deflating	first-glance effect	detrimental	'uncertain of oneself'
L4	knows little and speaks/writes poorly	depressing /deflating	satan halo effect	detrimental	'lazybones'

The four variants of the error are represented by the four commentaries cited in section 1.2 of this text. All four interpretations of the language educational events they refer to are at least partially dubious: Commentary 1 (inflating the teacher's evaluation) follows the reasoning "because the learner speaks/writes about an issue vividly and with conviction, s/he feels that issue and knows it well", which may be too prompt a conclusion, not to mention the remark that the teacher "can see Anne thinks properly"; Commentary 2 (inflating, too, but starting from a lower level of the learner's competence) relates to a learner who presents very few subject specifics but is still seen as well-oriented thanks to the involved language used, with the (clement) teacher's interpretation proving that "who seeks shall find" (and so the learner may not need to try

[5] The table below has been derived from a similar one presented in one of the two conference texts ("XIV KDE") discussing language bases of educational diagnostics.

too hard to demonstrate very specific knowledge); Commentary 3 (depressing[6]) appears to fail to recognize the fact that Ola may find it hard to verbalise what she knows (especially in the case of discipline that itself, by definition, is difficult to put into words) and the error seems to consist in the simplistic belief "if she knew, she'd say"; Commentary 4 (depressing, too) is particularly harmful to the learner, with the teacher finding in the learner's plain utterances confirmation of prior hypotheses stipulating that very little knowledge has been acquired; as noted earlier, Sławek's brevity may follow from his understanding of how the subject school be learnt and his limited (for untrained) ability to present what he knows.

The above shows that to conduct classroom diagnostics of the learners' utterances certain rules need to be devised and must be applied with regard to language educational events. Four rules pertaining to the four cases above have been proposed by the author earlier[7], with the two chief proposals having been made by them (two rules per either proposal):

(i) Do not view proper language forms as unambiguously indicative of subject competence.

(ii) Do not view improper language forms as overtly indicative of subject ignorance.

The "improper" language may be – wrongly and detrimentally to learners – seen by teachers as any form of language that departs from that introduced by the school or the classroom; taking into account the fact that the learner's colloquial language allows one to emanate one's own self in contacts with another man (Aleksander 2009: 41).

It follows from the above that disrespect of language issues jeopardizes validity of educational diagnostics as without sufficient attention being given to the language of utterances the construct of diagnostics remains partially vague. Conversely, adequate recognition of the learner's (individual) language used is tantamount to recognition of their personality and their experiencing of education.

---

[6] The adjective needs to be read not in its emotional sense (equivalent to 'upsetting') but rather as denoting a reduction in the value of the language encountered (as in economics, where it is opposite to an increase or inflation. The analogy to economics is by no means accidental here for the learner's utterances can be seen to "sell badly or well").

[7] The previous analysis presented at one of the KDE conferences addressed the issue of how each of the four rules relates to the relationship between the language itself and the personality or the construct (Daszkiewicz 2008: 327–329).

2.2. THEORETICAL STRATUM: Exploiting other disciplines in language-informed diagnostics

The fact that beyond each utterance there is a person implies that each language educational event is a highly personal occurrence. No single utterance produced by two different people is an exactly same event. As a consequence, analyses of language educational events need to exploit all disciplines which help understand the learner's position from different perspectives, be it sociology, psychology or any other social science. Although it is rather obvious that knowledge and language are strongly interwoven and so neither of them can be stripped of the other, there are numerous facets to utterances that render it irrational to view the language produced as if created in a vacuum without a very specific personality. To give an example of the lenses through which learners' utterances need to be construed, we may refer to a few cognitive-psychology--related truths. Namely:

(i) due to a multidimensional character of knowledge and achievements, analyses of learners' utterances cannot be conducted through the criterion of sequentiality (e.g. no single utterance can be viewed to signify any other preceding or following utterance);

(ii) due to the unique character of each learner's system of concepts, the most appropriate benchmark for interpretation of learners' utterances could be the other terms, sentences, words, or arguments presented by the very same learner;

(iii) due to an unlimited number of paraphrases and the arbitrariness of a language sign, no term (word, expression) denoting a given concept can be viewed as better than any other language sign[8];

(iv) due to the language shaping the framework of educational practices, that is the fact that it constitutes the means through which rights and duties of school-life participants are created and questioned (Cf. Wawrzyniak-Beszterda 2002: 27), the learner will generally "filter" his or her utterance through the framework established, which means that the utterance heard or written will frequently be formulated in a language which departs from the learner's (emotionally the closest) personal interlanguage[9].

[8] This property of language is also referred to as 'conventionality' (derived from transformational generative linguistics, also known in semiotics). Cf. Kurcz (2000: 38).

[9] Introduced first by Larry Selinker in 1972 and frequently referred to in applied linguistics, the concept (understood as a learner's own independent grammar system at any given point of

The analysis of language used is made difficult by the fact of each person **experiencing** his or her utterance, meaning undertaking decisions as to the choice of words and expressions, having his or her own convictions as to when it is worth using the language and/or finding certain language labels as more suitable, understandable and worth remembering than others. Hence, the theory of language educational events may rest primarily on answers to the following two key questions, which can only be answered by an interdisciplinary approach to educational diagnostics:

> Question 1. What qualities of an utterance determine the distance between latent and observable data and determine validity of inferences concerning knowledge made?
> Tentative answer: Language educational events become far more informative (i.e. the said distance can be said to be smaller) in the case of utterances marked by redundancy and subjectivity (as opposed to those representing a greater degree of entropy[10] and objectivity). In language terms, the size of distance (and so the degree of certainty of interpretation) seems most linked to, for instance, the use of pronouns, conventional collocations, etc.[11]

> Question 2. What amount of observable data can serve as a basis for assessment of latent qualities (i.e. how much text is needed to justify inferences concerning knowledge)?
> Tentative answer: Vague as it may sound at first here, there seems to be a point in (our reception of) a learner's utterance at which we, as listeners or readers, have gathered enough textual data to be disturbed by a change in lexis, style, grammar or register.

Language educational events are best to be viewed as one type or one facet of personal educational events: when experiencing by themselves the process of learning or the course of examinations[12], the learners may or may not choose to articulate their thoughts and thus provide substance for a language-based educational diagnosis. By recognizing the personal edge of their language choices

learning) merits a special place in educational diagnostics (See e.g. Johnson 1999: 174–6).

[10] The redundancy-entropy distinction follows the reasoning presented by Friske (1999: 34).

[11] More on the issue: Daszkiewicz (2007: 90).

[12] An exemplary set of learners' utterances concerning these two issues is discussed in the author's XIII KDE conference text.

and personal decisions, we prevent normalization of their education (and its assessment), whilst the fact that explaining what particular utterances prove is a most taxing challenge must not serve as an excuse for not undertaking to unravel the mental construct behind what the learner says or writes.

References

Aleksander Z. 2009. *Szansa na międzypokoleniowy dialog w kontekście rozumienia słów z zakresu moralności przez współczesne podmioty edukacyjne*, Wydawnictwo Uniwersytetu Gdańskiego, Gdańsk.

Daszkiewicz M. 2007. *Uczenie się i egzamin „na ustach" uczniów i nauczycieli – kontrastywna analiza wypowiedzi obojga stron*, [in:] *Uczenie się i egzamin w oczach uczniów*, eds. B. Niemierko, M.K. Szmigel, PTDE, Łomża ('XIII KDE'), pp. 78–92.

Daszkiewicz M. 2008. *Uczenie się i egzamin „na ustach" uczniów i nauczycieli – post scriptum do tekstu z XIII KDE na temat błędów atrybucji*, [in:] *Uczenie się i egzamin w oczach nauczyciela*, eds. B. Niemierko, M.K. Szmigel, PTDE, Opole ('XIV KDE'), pp. 321–330.

Friske J. 1999. *Wprowadzenie do badań nad komunikowaniem*, Wydawnictwo Astrum, Wrocław.

Johnson K., Johnson H. eds. 1999. *Encyclopedic Dictionary of Applied Linguistics. A Handbook for Language Teaching*, Blackwell Publishers Ltd., Oxford.

Konarski R. 2004. *Model cechy latentnej w analizie psychometrycznej testów i pozycji testowych*, [in:] *Standardy wymagań i normy testowe w diagnostyce edukacyjnej*, eds. B. Niemierko, H. Szaleniec, PTDE, Kraków.

Kurcz I. 2000. *Psychologia języka i komunikacji*, Wydawnictwo Naukowe „Scholar", Warszawa.

Mudyń K. 1997. *Zdarza się, że myślimy*, Wydawnictwo Profesjonalnej Szkoły Biznesu, Kraków.

Ross L. 1977. *The intuitive psychologist and his shortcomings: Distortions in the attribution process*, [in:] *Advances in experimental social psychology*, ed. L. Berkowitz, FL: Academic Press, Orlando, Vol. 10, pp. 173–220).

Twardowski A. 2005. *Trudności w porozumiewaniu się nauczyciela z uczniami*, [in:] *Porozumiewanie się w szkole na rzecz tworzenia edukacyjnej wspólnoty*, ed. A. Sajdak, Wydawnictwo Uniwersytetu Jagiellońskiego, Kraków, pp. 21–30.

Wawrzyniak-Beszterda R. 2002. *Doświadczenia komunikacyjne uczniów w czasie lekcji. Studium empiryczne*, Oficyna Wydawnicza Impuls, Kraków.

KWARTALNIK PEDAGOGICZNY
2015 NUMER 4(238)

Paulina Skórska
Instytut Badań Edukacyjnych w Warszawie

Gender gap in reading and writing achievements

Streszczenie

RÓŻNICE OSIĄGNIĘĆ EDUKACYJNYCH DZIEWCZĄT I CHŁOPCÓW W ZAKRESIE CZYTANIA I PISANIA

Istniejące badania empiryczne wskazują, że dziewczynki uzyskują wyższe wyniki w testach czytania i pisania w języku ojczystym w porównaniu do chłopców. Różnice te wahają się od 0,15 do 0,6 odchylenia standardowego, a więc od niewielkiego do znaczącego. Wielkość tego efektu różni się także między krajami, etapami edukacyjnymi i latami. Celem tego badania było oszacowanie wielkości efektu różnic płciowych w wynikach polskich standaryzowanych testów o wysokiej doniosłości, w tym wypadku egzaminu gimnazjalnego z języka polskiego. W badaniu wykorzystano modelowanie *Item Response Theory*, a następnie oszacowano wielkość efektu różnic płciowych w wynikach, biorąc pod uwagę wrażliwość efektu na typ wykorzystanej statystyki. Wzięto pod uwagę różnice w wariancji wyników w grupie dziewcząt i chłopców, gdyż dotychczasowe badania często pomijały ten fakt, co mogło prowadzić do błędów interpretacyjnych. Wyniki wskazują na znaczący efekt różnic płciowych w wynikach z języka polskiego, większy niż można by się spodziewać na podstawie istniejących badań empirycznych, zarówno w zakresie średniej wielkości efektu, jak i różnic wśród uczniów o najwyższych i najniższych umiejętnościach (krańce rozkładu umiejętności). Rezultaty badania mogą być użyteczne dla nauczycieli i twórców edukacyjnych polityk publicznych, którzy są zainteresowani zapewnieniem realizacji zasady sprawiedliwości i równości w edukacji.

Słowa kluczowe: różnice osiągnięć chłopców i dziewcząt, umiejętności czytania i pisania, testy doniosłe, osiągnięcia uczniów, teoria odpowiedzi na pozycje testowe (*Item Response Theory*, *IRT*).

Introduction

In modern societies, education is one of the most important foundations of the whole system of social stratification and can be considered as one of the

channels of social mobility (Gromkowska-Melosik 2011). Therefore, fairness is one of the most fundamental issues in educational measurement as a basis for protecting both test takers and test users in all aspects of testing (AERA et al. 2014). The testing process is unfair if it does not represent certain individuals or groups in terms of the construct being measured (Kane 2010). In other words, a test is unfair if its results systematically underestimate or overestimate abilities of certain groups, e.g. boys and girls. Test fairness can be viewed as an important aspect of test validity (Messick 1989): any factors which reduce fairness also reduce measurement validity. Latest edition of Standards for Educational and Psychological Testing (AERA et al. 2014) underlines that fairness concerns the validity of individual scores interpretations for intended uses. In this approach, all members of a group, regardless of whether we are dealing with a girl or a boy, should be treated similarly when making interpretations of test scores for individuals.

Demarcation lines of access to various forms of socialization, as well as education and the labour market run along with gender (Gromkowska-Melosik 2011). Therefore, one of the primary tasks in the evaluation of test fairness according to gender differences is to separate the real differences in ability levels between boys and girls and the bias in the testing process and the test itself (gender as a factor generating construct-irrelevant variance). The realization of this objective requires, however, answering the question of whether gender gap in results of high-stakes tests still exists, as empirical studies indicate inconclusive evidences. This article focuses on gender gap in literacy tests, as many studies show that the gender gap in reading and writing abilities is still visible.

The majority of research concerning the ability to use native language relates to reading comprehension. The Progress in International Reading Literacy Study (PIRLS) indicated that 10-year-old girls tend to perform better in reading comprehension in 35 out of 40 countries (Mullis et al. 2007). In 2010 in all states girls performed better than boys in tests of reading ability (Chudowsky & Chudowsky 2010). Notwithstanding, the National Assessment of Educational Progress (NAEP) showed that gender differences in favour of girls decreased from 13 points (0.3 standard deviation) in 1971 to 7 points currently (0.15 deviation) (Rampey et al. 2009). However, empirical studies suggested different effect sizes of gender gap in reading ability. Sarah Logan and Rhona Johnston (2009) indicated that girls outperform boys in the reading tests of about 2/3 of the standard deviation, however Lietz (2006) argued slightly lower effect of gender gap amounting to 0.19 standard deviation. The Programme for International Student

Assessment (PISA) showed the difference of 0.3 standard deviations (OECD 2009) and PIRLS the gap of 0.2 standard deviation favouring girls (Robinson & Lubienski 2011). In 2010 edition of PISA, Polish students obtained results indicating that girls outperform boys by 0.36 standard deviation (Logan & Johnston 2010). Polish studies (Skórska & Świst 2014) indicated a gender gap of about 0.4 standard deviation favouring girls. At the same time, researchers pointed out a smaller gender gap in writing ability (Logan & Johnston 2009), which may reach about 0.04 of standard deviation and can be treated as negligible (Lietz 2006).

Some studies, however, showed very different results: there is no gender gap in reading ability or boys tend to slightly outperform girls (e.g. Knickerbocker 1989). The effect of gender gap in the reading comprehension can be sensitive to the culture: Johnson (1973–1974) pointed out that while in Canada and the USA girls outperform boys in reading, in England and Nigeria the effect was opposite. Studies showed that in some areas of literacy boys are better than girls, e.g. verbal analogies (Lietz 2006). Although it has been shown that reading gender gap exists in over 100 countries, the differences can be in large part explained by the different methodology of studies included in the meta-analyses and reviews (Lietz 2006). Recently Mikkel Wallentin (2009) presented a critique of a well-known meta-analysis computed by Janet Hyde and Marcia Linn (1988) and showed that if the effect sizes were weighed by group frequencies, the effect of gender gap would be opposite in direction, suggesting slightly better verbal skills of boys. In this meta-analysis, 66% of the studies indicated a negligible effect of gender differences. Wallentin (2009) noted that the biggest effects resulted from smaller studies (smaller sample sizes) – bigger studies suggested no effect at all.

The diversity of the results of the gender gap in literacy depends on the stage of education. Madhabi Chatterji (2006) showed negligible differences in reading ability in kindergarten and at the entrance to the primary school. Before the third year of primary school boys are achieving results comparable to girls, the first noticeable differences appear near to the 3rd grade (Husain & Millimet 2009). Some studies, however, showed that the variation in reading tests results may appear in the early stages of primary school (LoGerfo et al. 2006). At the beginning of primary school the effect equals to 0.12 standard deviation in favour of girls for the 90% of achievement distribution and ranges from 0.18 to 0.21 for the remaining part of the distribution. In 5th grade for top 25% of the reading test results the effect of gender gap becomes insignificant, and then again increases (Robinson & Lubienski 2011).

The evaluation of the gender gap in literacy, which is based only on the mean differences in reading and writing test results, can be misleading. Mean differences in test results between girls and boys may be still small, but the variances of the results in two groups may differ considerably. Large differences in test scores variances may translate into large differences in the results of boys and girls at both ends of the ability distribution – among the most and least talented students in the population. Variance of reading and writing ability for boys is typically higher than among girls by about 9% (Strand et al. 2006). Women have an advantage in each part of the reading ability distribution, but overrepresentation of boys is particularly visible at the left extreme of distribution, among the students with the worst test results (Robinson & Lubienski 2011).

Limitations of existing studies

Claudia Buchmann and colleagues (2008) emphasized that the debate on gender gap in literacy ability is not completed. As indicated previously, research on the gender differences between girls and boys are not conclusive. Many studies were based on mean performance differences between boys and girls, which is insufficient (Robinson & Lubienski 2011). In addition, the size of the effect clearly depends on the metric used. When the standardized results are used (as an indicator of the students achievements) gender gap is narrowing, whereas the analyses based on the scaled results indicate growth of the gap. Since, as noted above, the difference in the variances among boys and girls results is evident, the only proper metric used to comparison should be the effect size statistic. Moreover, most of the research was based on data collected in the Anglo-Saxon countries, especially on American student's samples. At the same time, patterns of gender inequalities in developing countries vary from those observed in the United States (Buchmann & Hannum 2001), so the results are not representative for other countries.

The aim of the study and methodology

The aim of the study was to examine gender gap in the results of large-scale, high-stakes tests in literacy in the Polish cultural context.

Student's achievements were estimated using results from Polish examinations after lower secondary school, from 2012 to 2014. In contrast to the majority of studies exploring gender gap in cognitive skills, in this study the data collected for the whole population (instead of just a sample) were used. The data were initially gathered by Central Examination Board. The size of population in the subsequent years was as follows:

- 393,766 students (49.19% girls and 50.81% boys) in 2012;
- 379,752 students (48.97% girls and 51.03% boys) in 2013;
- 362,755 students (48.96% girls and 51.04% boys) in 2014.

Student's ability level was estimated by using standardized total test scores and scaled scores (with Item Response Theory (IRT) modeling). Thereby, it was possible to compare the magnitude of the gender gap estimated in two different ways, since, as stated above, the size of the gender gap effect is sensitive to the metric used.

IRT relies on two basic assumptions (Hambleton, Swaminathan & Rogers 1991): a) the performance of a student on a particular test can be predicted by a set of latent traits (abilities) and b) the relationship between students' item performance and these latent traits can be described by the monotonically increasing function called item characteristic curve (ICC). This function indicates that for students with higher levels of latent trait (ability), the probability of the correct response to particular item increases.

The three most popular IRT models take their names from the number of item parameters (characteristics) estimated in the model. One-parameter logistic model (1PLM) is often called the Rasch model (Rasch 1960). In 1PLM only one item parameter – the difficulty (b parameter) is estimated. The more difficult particular item is, the higher level of latent trait (ability) is needed to answer this item correctly. The difficulty is also called location parameter, because it indicates the location of ICC relatively to ability scale. Lower b-values are associated with shift of ICC to the left part of the ability scale, suggesting that an item is easy. Two-parameter logistic model (2PLM; Birnbaum 1968) estimates also discrimination (a) parameter (in 1PLM model, for all items constant, equals to 1). Discrimination parameter indicates the slope of the ICC (how steep the ICC is for a particular item). The higher the discrimination parameter is, the better differentiation between students with different ability levels. Three-parameter logistic model (3PLM; Samejima 1969) allows for non-zero lower asymptote of ICC, and therefore the third parameter called the guessing or pseudo-guessing (c) is estimated. C parameter represents the probability of providing the correct

answer in the item by students with a very low ability level. Other models in IRT framework are the extensions of these three fundamental ones[1].

In this study, for the dichotomous items three-parameter logistic model (3PLM) was used to estimate students' reading and writing ability. For the vast majority of items 3PLM exhibited a good fit to the data. For the rest of items two-parameter logistic model (2PLM) was used. For polytomously scored items Graded Response Model (GRM; Samejima 1969) was used to estimate students' ability level and item parameters.

As students' ability level (θ) is computed on the basis of only a subset of the total possible item pool, it is estimated with a substantial amount of measurement error. To take this uncertainty associated with the estimates into account, and to compute unbiased group-level estimates, plausible values (PV) were used. Plausible values are multiple values representing the likely distribution of a student's ability and are generated through multiple imputations (Mislevy 1993). In this study five plausible values per individual were estimated.

To estimate the effect size of gender gap in reading and writing ability, procedures traditionally occurring in meta-analysis (MA) were used. The effect of gender gap was estimated using standardized mean difference (Hedges & Olkin 1985). To assess discrepancy of literacy achievements between boys and girls, variance ratio (VR) was used. This is the ratio of the male variance to the female variance (Feingold 1992). Thus, a VR greater than 1.00 indicates greater male variability, VR of 1.00 indicates homogeneity of variance across gender and VR less than 1.00 indicates greater female variability.

The gender gap on both ends of ability distribution (among the most and least talented students in the population) was defined as a ratio of girls to boys (and reverse) for the lowest and highest 10, 5 and 1 percentage of ability estimates.

Results

The effect size of gender gap in literacy achievements, measured as a standardized mean difference between girls and boys, is generally slightly higher for IRT ability estimates than for other metrics. For IRT estimates, standardized differences in means of girls and boys amount to 0.47, 0.46 and

[1] More can be read in e.g. Kondratek and Pokropek (2013).

0.46 in 2012, 2013 and 2014, respectively. For plausible values the gender gap effect size is slightly smaller, ranging from 0.42 (in 2014) to 0.44 (in 2012). Details about the effect sizes and their corresponding confidence intervals (CI) can be found in Table 1.

Table 1. Effect sizes of gender gap in literacy tests (examination after lower secondary school) 2012–14 with corresponding confidence intervals according to the metric used

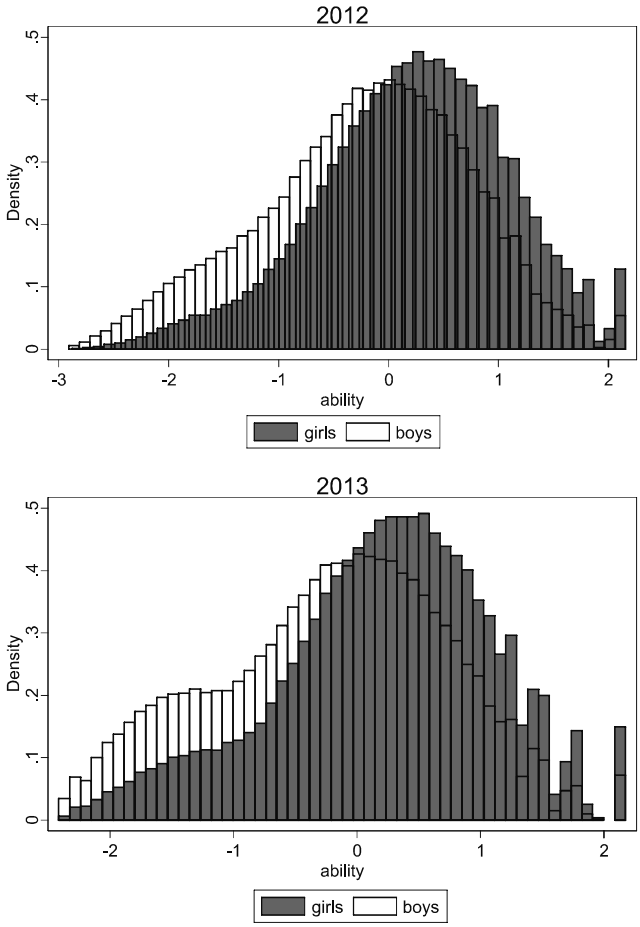
Year	Estimate	Effect size	Lower bound of CI	Upper bound of CI
2012	IRT (θ)	0.47	0.46	0.47
	PVs	0.44	0.43	0.44
	standardized total scores	0.45	0.44	0.46
2013	IRT (θ)	0.46	0.45	0.46
	PVs	0.43	0.42	0.43
	standardized total scores	0.42	0.42	0.43
2014	IRT (θ)	0.46	0.45	0.46
	PVs	0.42	0.42	0.42
	standardized total scores	0.43	0.42	0.43

* Final estimates of effect size for plausible values were computed by averaging effect sizes for five PVs.

As noted above, effect size defined as standardized mean difference is insufficient, since relying on average performance may translate into attenuation of gender gap effect magnitude. Examination of Graph 1 reveals evident differences between boys and girls in performance (estimated as θ in IRT models) on literacy test virtually for the entire ability distribution. Boys are overrepresented for the lower levels of ability (left part of the distribution) whereas girls dominate in the right part of the distribution.

As can be seen in Table 2, boys' ability estimates, regardless of whether they are specified as scaled (with IRT models) or standardized scores, are more variable than girls results for each year concerned. Taking into consideration IRT ability estimates, the variance ratio (VR) is the highest for 2012 examination and lowest in 2014. In percentage terms boys' results are 19% more variable than girls' in 2012 and 18% more variable in 2013. The magnitude of variance ratio is, as expected, sensitive to the metric used to estimate literacy ability, as standardized total test scores indicate the highest and PVs the lowest values of VR. According to standardized total test scores differences in ability variance among

boys and girls are meaningful, as boys' scores are even 23% more variable than girls (in 2012).

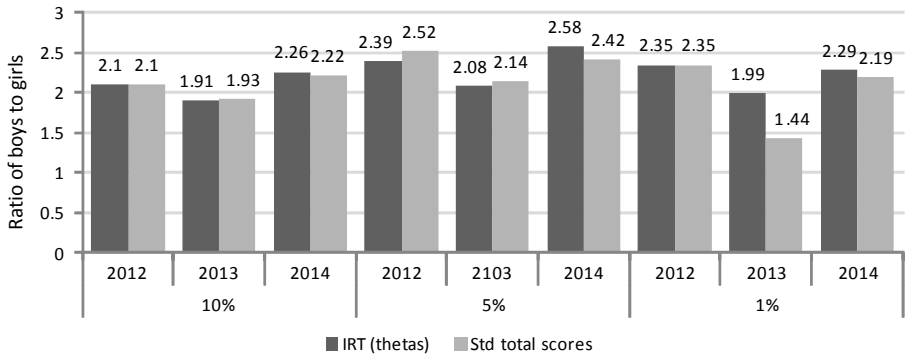


Graph 1. Distributions of literacy ability (estimated as θ in IRT models) 2012–14 according to gender

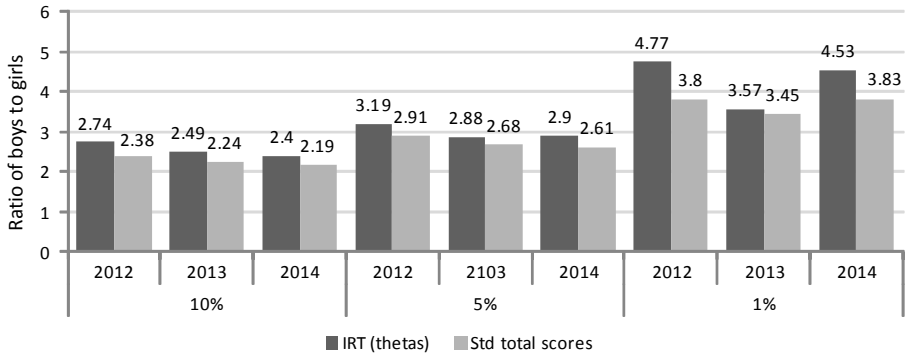
Table 2. Variance ratios (VRs) for the results of examination after lower secondary school in 2012–14 according to the metric used to estimate literacy ability

Year	IRT (θ)	PVs	Standardized total scores
2012	1.19	1.16	1.23
2013	1.18	1.16	1.19
2014	1.10	1.07	1.22

High differences in variance between boys and girls results may translate into considerable gender gap for the most and least talented parts of population and this hypothesis was examined. The ratios of girls' to boys' scoring above 90% are from 1.91 in 2013 to 2.26 in 2014 (for IRT ability estimates). The biggest gender gap can be observed in 2013, when 2.5 times more girls than boys achieved scores above 95% of distribution. Using standardized total test scores as ability estimates, similar results can be observed (except 2013), indicating twice the number of female among the most talented students in the population.



Graph 2. Ratios of girls to boys in 10, 5 and 1 percentage of the highest literacy ability estimates (θ in IRT models and standardized total test scores) in 2012–14



Graph 3. Ratios of boys to girls in 10, 5 and 1 percentage of the lowest literacy ability estimates (θ in IRT models and standardized total test scores) in 2012–14

As can be seen in Graph 3, even more evident gender inequalities can be observed at the lower tail of the ability distribution. For the 1% of the least talented students the overrepresentation of boys is clear. In 2012 in the group of lowest scoring students, there were nearly five times more boys than girls.

In 2013 this ratio was lower, but in 2014 once again exceeded 4.5. For both, the upper and lower tail of ability distribution, disparities in the number of girls and boys are most pronounced for 1% of the most and least talented students in the population.

Discussion

The study confirmed that the gender gap is still visible in the literacy ability for students at the end of the lower secondary school. As expected (Robinson & Lubienski 2011), using scaled scores as estimates of literacy ability produces slightly higher gender gap effect sizes in comparison to standardized scores. Regardless of the metric used to estimate literacy ability, according to the rule of thumb proposed by Jacob Cohen (1992), the magnitude of gender gap can be classified as moderate. In practical terms effect sizes obtained in this study indicate that the mean literacy score of girls is at the 70% of boys' literacy scores. Translating this gap into common language effect size, also known as the probability of superiority (Grissom & Kim 2005), leads to the conclusion that the probability that a randomly sampled girl will have a higher literacy ability level than a randomly sampled boy is about 33%. In terms of differences in variability, the distribution of scores for the girls overlaps only in 2/3 with the distribution of scores for boys, and nonoverlap amounts to 33%. This lack of overlap can be confirmed by visual examination of both distributions. Gender differences can be observed essentially in any part of the ability distribution, especially for both lower and upper tail.

The effect of gender gap is considerable and slightly higher (difference about half a standard deviation) than reported in Anglo-Saxon countries. This suggests that gender gap in reading and writing is strongly sensitive to the culture, as indicated by Johnson (1973–1974). The magnitude of the gender gap in education is strongly dependent on the socio-economic conditions in the particular country. Luigi Guiso and colleagues (2008) evidenced that gender differences in school achievements are related to Gender Gap Index (GGI). Gender Gap Index consists of four dimensions: health, educational attainment, economic participation and opportunity, and political empowerment. Thus, it reflects the global gender inequalities in key areas of social life. According to The Global Gender Gap Report (2013) results, the rank of Poland decreased from 44. in 2006 to 54. in 2013, indicating growing gender inequalities, relative to other 135 analyzed countries.

Taking into consideration these results, further studies should examine whether the gender gap in reading and writing skills widens over time.

Differences in variances of ability estimates in a group of girls and boys are meaningful, especially for IRT estimates and standardized total test scores, where the variance ratio suggests that boys' scores are about 20% more variable than girls' scores. These results are consistent with the trends found in other countries, however the variance ratio (VR) magnitude is about twice as high as VR reported in the other empirical studies (e.g. Strand et al. 2006).

Such big differences in variances of girls and boys performance may translate into considerable gender gap for the most and least talented parts of population. This hypothesis found a confirmation in this study. The highest gender gap can be observed for the lower tail of literacy ability distribution. Among the students with lowest 1 percentage of ability scores, there is over four times more boys than girls. In 2012 boys made up more than 80% of students in this part of distribution. This effect is higher than should be expected based on existing studies concerning assessment of gender gap in high-stakes tests (e.g. Hyde et al. 2008).

The possible explanations of gender gap in the literacy can be classified into a few categories. Some scientists underline that boys and girls differ in reading and writing achievements due to biological differences (Gunzelmann & Connell 2006). Left brain hemisphere, responsible for verbal activities, develops earlier for girls than for boys. Males' fine motor also develops later, leading to difficulties in mastering the mechanic activities needed for reading and writing (e.g. holding a pen). Another set of potential explanations is associated with socialization. Some studies (e.g. Wilhelm & Smith 2009) showed that boys may consider reading as a "feminine" activity. Therefore some students may reject reading, trying to confirm male identity, which may reduce boys' motivation to reading and lead to a self-fulfilling prophecy. Boys and girls also differ in learning strategies. Girls apply more strategies and use meta-knowledge, whereas boys are focused on information retrieval, preferring rapidity of responses over deep understanding of text (Smith & Wilhelm 2002). In recent years, PISA study showed (OECD 2009) that reading engagement is one of the best predictor of literacy achievements, sometimes even stronger than socioeconomic status. Boys are less engaged in reading than girls and moreover, among boys reading comprehension is affected by the content of text. The question whether gender gap in literacy among Polish students is caused more by biological, social or psychological factors should be the subject of further research and analysis.

The study is exploratory and only shed little light on the description and magnitude of gender gap in literacy achievements, measured by high-stakes tests (examination after lower secondary school). The possibility to assess the change of gender gap over time is limited, as the results include the data form examination in three sequential years. The conclusion based on analyzed data is that gender gap does not tend to narrow between 2012 and 2014, either in terms of mean effect size, or in terms of frequency of boys and girls among the most and least talented students. Considering the nature of effect size statistic, which is in principle a standardized metric, the comparison of gender gap over time is methodologically proper, however using equated test scores will increase results validity. Further research will be focused on the assessment of gender gap in other educational stages (primary and upper-secondary school), as well as changes in phenomenon over time, using Comparable Examination Results (CER)[2].

References

American Educational Research Association, American Psychological Association & National Council on Measurement in Education. 2014. *Standards for educational and psychological testing*, American Educational Research Association, Washington.

Birnbaum A. 1968. Some latent trait models and their use in inferring an examinee's ability, in: Lord F.M. & Novick M.R. (eds.), *Statistical theories of mental test scores*, Addison-Wesley, Reading, pp. 17–20.

Buchmann C., DiPrete T.A. & McDaniel A. 2008. Gender inequalities in education, *Annual Review of Sociology*, *34*, pp. 319–337.

Buchmann C. & Hannum E. 2001. Education and stratification in developing countries: A review of theories and empirical research, *Annual Review of Sociology*, *27*, pp. 77–102.

Chatterji M. 2006. Reading achievement gaps, correlates, and moderators of early reading achievement: Evidence from the early childhood longitudinal study (ECLS) kindergarten to first grade sample, *Journal of Educational Psychology*, *98*(3), pp. 489–507.

Chudowsky N. & Chudowsky V. 2010. *Are there differences in achievement between boys and girls?*, Center on Education Policy, Washington, DC.

Cohen J. 1992. A power primer, *Psychological Bulletin*, *112*, pp. 155–159.

Feingold A. 1992. Sex differences in variability in intellectual abilities: A new look at an old controversy, *Review of Educational Research*, *62*, pp. 61–84.

Grissom R.J. & Kim J.J. 2005. *Effect Sizes for Research: A Broad Practical Approach*, Lawrence Erlbaum Associates, Mahwah, NJ.

[2] More about CER can be found on: http://pwe.ibe.edu.pl

Gromkowska-Melosik A. 2011. *Edukacja i (nie)równość społeczna kobiet. Studium dynamiki dostępu*, Oficyna Wydawnicza Impuls, Kraków.

Guiso L., Monte F., Sapienza P. & Zingales L. 2008. Culture, Gender and Math, *Science*, *320*(5880), pp. 1164–1165.

Gunzelmann B., & Connell D. 2006. The new gender gap: Psychological, neurobiological, and educational perspectives, *Educational Horizons*, *84*(2), pp. 94–101.

Hambleton R.K., Swaminathan H. & Rogers H.J. 1991. *Fundamentals of Item Response Theory*, Sage, Newbury Park.

Hedges L.V. & Olkin I. 1985. *Statistical methods for meta-analysis*, Academic Press, Orlando.

Husain M. & Millimet D. 2009. The mythical "boy crisis"?, *Economics of Education Review*, *28*, pp. 38–48.

Hyde J.S. & Linn M.C. 1988. Gender differences in verbal ability: A meta-analysis, *Psychological Bulletin*, *104*, pp. 53–69.

Johnson D.D. 1973–1974. Sex differences in reading across cultures, *Reading Research Quarterly*, *9*(1), pp. 67–86.

Kane M. 2010. Validity and fairness, *Language Testing*, *27*(2), pp. 177–182.

Knickerbocker J.L. 1989. Sex differences in reading achievement: A review of recent research, *Ohio Reading Teacher*, *24*(2), pp. 33–42.

Kondratek B. & Pokropek A. 2013. IRT i pomiar edukacyjny, *Edukacja*, *124*(4), pp. 42–66.

Lietz P. 2006. A meta-analysis of gender differences in reading achievement at the secondary school level, *Studies in Educational Evaluation*, *32*, pp. 317–344.

Logan S. & Johnston R.S. 2009. Gender differences in reading ability and attitudes: Examining where these differences lie, *Journal of Research in Reading*, *32*(2), pp. 199–214.

Logan S. & Johnston R.S. 2010. Investigating gender differences in reading, *Educational Review*, *62*(2), pp. 175–187.

LoGerfo L., Nichols A. & Reardon S.F. 2006. *Achievement gains in elementary and high school*, Urban Institute, Washington, DC.

Messick S. 1989. Validity, in: Linn R.L. (ed.), *Educational measurement*, American Council on Education, Macmillan, New York, pp. 13–103.

Mislevy R.J. 1993. Should "multiple imputations" be treated as "multiple indicators"?, *Psychometrika*, *58*(1), pp. 79–85.

Mullis I.V.S., Martin M.O., Kennedy A.M. & Foy P. 2007. *PIRLS 2006 international report: IEA's progress in international reading literacy study in primary schools in 40 countries*, Boston College, Chestnut Hill, MA.

Organisation for Economic Co-operation and Development. 2009. *Equally prepared for life? How 15-year-old boys and girls perform in school*, Organisation for Economic Co-operation and Development, Paris.

Rampey B.D., Dion G.S. & Donahue P.L. 2009. *NAEP 2008 trends in academic progress* (NCES 2009-479), National Center for Education Statistics, Washington, DC.

Rasch G. 1960. *Probabilistic models for some intelligence and attainment tests*, Danish Institute for Educational Research, Copenhagen.

Robinson J.P. & Lubienski S.T. 2011. The development of gender achievement gaps in mathematics and reading during elementary and middle school: Examining direct cognitive assessments and teacher ratings, *American Educational Research Journal*, *48*(2), pp. 268–302.

Samejima F. 1969. *Estimation of latent ability using a response pattern of graded scores*, Psychometrika monograph (Vol. 17).

Skórska P. & Świst K. 2014. Wielkość efektu płci w wewnątrzszkolnych i zewnątrzszkolnych wskaźnikach osiągnięć ucznia, [in:] Niemierko B. & Szmigiel M.K. (eds.), *Diagnozy Edukacyjne. Dorobek i nowe zadania*, Grupa Tomami, Kraków, pp. 89–103.

Smith M.W. & Wilhelm J.D. 2002. *Reading don't fix no chevys: Literacy in the lives of young men*, Heinemann, Portsmouth, NH.

Strand S., Deary I.J. & Smith P. 2006. Sex differences in Cognitive Abilities Test scores: A UK National Picture, *British Journal of Educational Psychology*, *76*, pp. 463–480.

Wallentin M. 2009. Putative sex differences in verbal abilities and language cortex: A critical review, *Brain and Language*, *108*, pp. 175–183.

Wilhelm J. & Smith M.W. 2009. Boys and literacy: Complexity and multiplicity, in: Christenbury L., Bomer R. & Smagorinsky P. (eds.), *Handbook of adolescent literacy research*, Guilford Press, New York, pp. 360–371.

World Economic Forum. 2013. *The Global Gender Gap Report* (2013). Available: http://www3.weforum.org/docs/WEF_GenderGap_Report_201.

KWARTALNIK PEDAGOGICZNY
2015 NUMER 4(238)

MATERIAŁY I SPRAWOZDANIA Z BADAŃ

Ewa Maria Filipiak
Uniwersytet Kazimierza Wielkiego w Bydgoszczy

Zdolności wytwarzania konwergencyjnego i dywergencyjnego uczniów szkoły podstawowej i gimnazjum[1]

Summary

CONVERGENT AND DIVERGENT PRODUCTION ABILITIES OF PRIMARY AND SECONDARY SCHOOL STUDENTS[2]

The article presents the results of research regarding the diagnosis of primary and secondary school students' abilities in respect of two operations identified by Joy P. Guilford: convergent production and divergent production. The research used a set of *SOI-Learning Abilities Tests* by Mary Meeker and Robert Meeker. The research was conducted on 516 students (323 primary

[1] Badania przeprowadzono w roku 2009 w ramach projektu MNiSW Nr N 107 037 31/38/75 pt.: *Rozwój zdolności uczenia się uczniów szkoły podstawowej i gimnazjum*. Kierownik projektu: prof. dr hab. Ewa Filipiak. Zespół: prof. dr hab. Anna I. Brzezińska z Instytutu Psychologii UAM w Poznaniu, dr hab. Bogusława D. Gołębniak, prof. DSW we Wrocławiu, dr Halina Smolińska-Rębas i dr Ewa Lemańska-Lewandowska z UKW w Bydgoszczy. Wyniki badań w których poddano analizie poziom i właściwości zdolności uczenia się w zakresie operacji wyróżnionych przez Guilforda (poznawanie, pamięć, ocenianie, wytwarzanie dywergencyjne i konwergencyjne zaprezentowano w tekstach Filipiak 2009, 2011a, 2015).

[2] The research was conducted in 2009 as part of the MNiSW Nr N 107 037 31/38/75 project entitled: *Development of learning abilities in primary and secondary school students*. Project supervisor: prof. dr hab. Ewa Filipiak. Research team: prof. dr hab. Anna I. Brzezińska of Institute of Psychology of Adam Mickiewicz University in Poznań, dr hab. Bogusława D. Gołębniak, DSW professor in Wrocław, dr Halina Smolińska-Rębas and dr Ewa Lemańska-Lewandowska of Kazimierz Wielki University in Bydgoszcz. The results of research analysing the level and properties of learning abilities as regards operations identified by Guilford (cognition, memory, evaluation, divergent and convergent production) were presented in the works of Filipiak 2009, 2011a, 2015).

school students and 193 secondary school students; 259 students were boys). In divergent production tests students at all levels of school education (1st and 2nd stage of primary school and secondary school) experienced difficulties with tests requiring the application of non-standard solutions and new approaches. This tendency grows in the consecutive years of education. Very good results of primary school students as regards convergent production – which involves the ability to combine and synthesize information relevant for the purpose of solving a given problem – do not continue for secondary school students. The age and sex proved to be differentiating factors for the level and properties of learning abilities as regards the abovementioned operations. In all age groups girls scored better than boys. Third grade students achieved the most scores assessed as above average, outstanding and talented.

Key words: divergent production, convergent production, Guilford's structure of intellect, information.

red. Paulina Marchlik

Wprowadzenie

Świat *płynnej ponowoczesności*, który opisuje Zygmunt Bauman nieustannie nas zaskakuje. O ile w poprzedniej epoce doświadczaliśmy fazy niedoinformowania, dziś pochłania nas ocean informacji, w którym, jak pisze Z. Bauman (2011), nie da się już pływać ani nurkować, można jedynie dryfować lub surfować. „W niestabilnym świecie płynnej nowoczesności, w którym formy niemal nigdy nie zachowują swojego kształtu wystarczająco długo, by zakrzepnąć i dowieść, że można na nich polegać, lepiej iść, niż siedzieć, lepiej biec, niż iść, a jeszcze lepiej – surfować. Surfing wygrywa na lekkości i zwinności surfera; dobrze też, jeśli surferzy nie okażą się szczególnie wybredni w doborze fal i gotowi będą w każdej chwili zmienić upodobania" (Bauman 2011: 153). Świat, w którym przyszło nam i naszym uczniom żyć, utracił swoją stabilność. Jednym ze strategicznych celów szkoły jest przygotowanie pracowników przyszłości, zdolnych myśleć twórczo, rozwiązywać problemy, ale też przygotowanych do satysfakcjonującego, samodzielnego życia w dynamicznie zmieniającym się świecie. Pytanie, które warto tu postawić jest pytaniem o to **czy** w ogóle, a jeżeli tak, to **w jaki sposób** szkoła przygotowuje uczniów do życia w świecie przesyconym nadmiarem informacji. Fakt, że świat, w którym przyszło nam żyć współcześnie, stanowi imperium informacji jest niezaprzeczalny. Wyzwaniem jest przygotowanie jednostek do rozumnego i realizowanego z poczuciem sprawstwa wejścia w rolę „kognitariusza" (określenie Józefa Kozieleckiego 1998), czyli takiego pracownika przyszłości, który tworzy, przechowuje, wykorzystuje i upowszechnia infor-

mację i wiedzę. Jak z tym wyzwaniem radzi sobie współczesna szkoła? Czy jest adekwatna kulturowo? Niepokój o wydolność i kondycję polskiej szkoły od wielu już lat jest przedmiotem rozważań podejmowanych przez pedagogów. Niewydolność kształceniową i wychowawczą szkoły, hamowanie inicjatywy intelektualnej, blokowanie rozwoju myślenia uczniów i nauczycieli, niewrażliwość szkoły na wydarzenia zachodzące w świecie odsłania Dorota Klus Stańska (2008a, 2008b). Stanisław Dylak (2008) przestrzega przed przekonaniem (nauczycieli i uczniów) o posiadaniu pewnej, doskonałej wiedzy, które może być równie szkodliwe, jak brak wiedzy w ogóle. We współczesnej polskiej szkole, która niewątpliwie się zmienia, nadal mało jest przestrzeni na samodzielne zdobywanie wiedzy, uczenie się problemowe. „Pewność wiedzy nauczycielskiej, zarówno w sensie jej niewzruszalności, jak i obecności i dostępności, nie sprzyja kształtowaniu u uczniów dyspozycji do budowania, poszukiwania i samodzielnego tworzenia. Może nawet dzieje się gorzej – to ich rozleniwia" (Dylak 2008: 72). Bauman zauważa, że receptą na sukces w świecie nowoczesnym jest nie tylko ważna umiejętność funkcjonowania w świecie przesyconym informacją, ale „bycie sobą i bycie niepodobnym do innych". „To różnica a nie tożsamość, sprzedaje się najlepiej. Dysponowanie wiedzą i umiejętnościami wymaganymi na danym stanowisku" i prezentowanymi już przez tych, którzy zajmowali je wcześniej lub starają się o nie obecnie, to dzisiaj za mało. To raczej słabość niż atut. Zamiast tego potrzeba niezwykłych pomysłów, wyjątkowych i niespotykanych wcześniej, a przede wszystkim umiejętności chadzania własnymi ścieżkami. Tego rodzaju wiedzy (a właściwie inspiracji) pragną ludzie żyjący w epoce płynnej nowoczesności" (Bauman 2011: 159). W tym świecie ponowoczesnym nie ulega wątpliwości, że szczególnie ważne stają się zdolności twórcze jednostki. Współczesna dydaktyka podtrzymując tendencje sygnalizowane już przez pedagogów w minionym wieku (!) nadal wielkie znaczenie przypisuje kształceniu dla twórczości, organizowaniu sytuacji dydaktycznej ukierunkowanej na rozwiązywanie problemów, „wychodzeniu poza dostarczone informacje" (Bandura 1975; Nosal 1992; Nęcka 1994; Sękowski 2005; Szmidt 2007; Limont 1994, 2012; Szymański 1987; Góralski 1990; Pachociński 1997; Popek 2000; Matczak 2001; Filipiak 2012). Te deklaracje „kształcenia dla twórczości" niestety nie tak łatwo przenikają do szkolnej codzienności. W artykule poddano analizie zdolności wytwarzania dywergencyjnego i konwergencyjnego uczniów szkoły podstawowej i gimnazjum.

Problem badawczy

Punktem wyjścia w badaniach było przyjęcie rozumienia zdolności uczenia się w ujęciu J.P. Guilforda, który przedstawił trójwymiarową koncepcję struktury intelektu znaną jako Model Struktury Intelektu (MSI). Kontrowersyjny, morfologiczny (jak nazywał go sam autor) model struktury intelektu, który przypomina tablicę pierwiastków Mendelejewa „służył jako podstawa do generowania hipotez na bazie których, posługując się wspomnianą metodą analizy czynnikowej, wyodrębniano coraz to nowe czynniki" (Strelau 2014: 81). Model ten w dużym stopniu ukierunkował badania nad zdolnościami. Zdaniem Guilforda (1978) każdą zdolność można opisać za pomocą trzech aspektów: operacji, treści oraz wytworów. Model MSI obejmuje 120 sześcianów. Każdy z nich reprezentuje unikalny rodzaj zdolności. Każda komórka w sześcianie jest specyficzną kombinacją jednego rodzaju operacji, jednego rodzaju treści oraz jednego rodzaju wytworu (Guilford 1978)[3].

Operacje dotyczą procesów przetwarzania informacji. Obejmują 5 kategorii, które opisują specyficzny sposób funkcjonowania umysłu jednostki. Są to: poznawanie, pamięć, ocenianie, wytwarzanie konwergencyjne i wytwarzanie dywergencyjne. Kolejność wyodrębnionych kategorii operacji ma pewne logiczne uzasadnienie, ale autor zaznacza że niekoniecznie musi być właśnie taka:

> Jeśli chodzi o operacje, poznawanie jest podstawowe dla wszystkich pozostałych rodzajów, dlatego umieszczone jest jako pierwsze. Jeżeli nie ma poznawania, nie ma i zapamiętywania, jeżeli nie ma zapamiętywania, nie ma wytwarzania, ponieważ rzeczy wytwarzane pochodzą przede wszystkim z zasobów pamięci. Jeżeli brak jest i poznawania, i wytwarzania, nie ma wówczas oceniania. Od przodu do tyłu modelu występuje zatem wzrastająca zależność danego rodzaju operacji od innych (Guilford 1978: 141).

Człowiek dokonuje operacji umysłowych przetwarzając dwojakiego rodzaju informacje: treści i wytwory.

Treści jak wskazuje sama nazwa stanowią treściowy aspekt informacji rozróżnianych przez organizm. Autor modelu wyróżnił następujące kategorie treści: figuralne, symboliczne, semantyczne i behawioralne (Guilford 1978).

Wytwory procesu poznawczego to formy, w jakiej występują informacje. Są to kategorie skupiające bodźce figuralne, symboliczne i semantyczne. Guilford

[3] W ostatecznej wersji swojej teorii w 1988 roku Guilford wyróżnił 180 czynników (6x5x6), szerzej: Strelau 2014.

wyodrębnił sześć rodzajów wytworów: jednostki, klasy, relacje, systemy, przekształcenia i implikacje.

W modelu MSI Guilforda szczególne znaczenie ma **informacja**. Autor określa jej naturę przyjmując psychologiczny punkt widzenia. Wprowadza rozróżnienie między klasycznym pojęciem informacji a rozumieniem psychologicznym informacji w przyjętym modelu, akcentując stopień jej pewności oraz rozróżnienie terminów: wiedza i informacja (Guilford 1978).

W projekcie badawczym realizowanym w ramach projektu MNiSW Nr N 107 037 31/38/75 (por. przyp. 2) analizie poddano wszystkie kategorie operacji (szerzej Filipiak 2009, 2011a, 2015). W niniejszym artykule przedstawiono analizę wyników badań w odniesieniu do dwóch wyróżnionych przez Guilforda zdolności wytwarzania, którym przypisuje bardzo ważną rolę. Są to: wytwarzanie konwergencyjne i wytwarzanie dywergencyjne.

Wytwarzanie dywergencyjne (*divergent production*) jest to proces przetwarzania informacji, który prowadzi do uzyskania wielu możliwych rozwiązań. Można powiedzieć, że jest ukierunkowany na generowanie informacji z naciskiem na różnorodność i jakość wygenerowanych produktów. W tej grupie zdolności Guilford (1978) wyróżnił kilka zdolności objętych wspólnym mianem *płynności* wytwarzania oraz zdolności objętych mianem *giętkości* wytwarzania (por. też Pietrasiński 1969, 1976; Strelau 2014; Nęcka 2002; Filipiak 2002).

Płynność jest rozumiana jako zdolność do wytwarzania w krótkim czasie wielu pomysłów, słów, idei, sentencji. Jej wskaźnikiem jest liczba wygenerowanych pomysłów. Guilford (1978) wyróżnił kilka rodzajów płynności: słowną, ideacyjną i skojarzeniową. Giętkość jest rozumiana jako gotowość do zmiany kierunku myślenia czy też modyfikowania informacji. Wskaźnikiem giętkości jest liczba wymienionych różnych kategorii zastosowań w której mieszczą się wygenerowane pomysły. Jednostki, które wykazują wysokie zdolności wytwarzania dywergencyjnego wymykają się stereotypom i konwencjonalnym „trafnym odpowiedziom". Nie znaczy to jednak, że ich wytwarzanie jest pozbawione dyscypliny wytwarzania, poszukiwania możliwych oryginalnych rozwiązań i troski o jakość wytworów. Pomiar wytwarzania dywergencyjnego obejmuje także *oryginalność* rozumianą jako zdolność do wytwarzania odpowiedzi rzadko uzyskiwanych lub niezwykłych oraz *staranność* (zdolność zamykania i uzupełniania struktur).

Wytwarzanie konwergencyjne (*convergent production*) jest to proces znajdowania prawidłowego rozwiązania problemu na drodze wyprowadzania dedukcji logicznej z podanej informacji. Jest związane ściśle ze zdolnością wyprowadzania wniosków. To właśnie wytwarzanie konwergencyjne staje się dominujące, jeżeli wkład informacji, na których jednostka „pracuje" wystarcza do ustalenia unikal-

nej odpowiedzi. W wytwarzaniu konwergencyjnym nie ma takiej swobody wytwarzania, jaką daje wytwarzanie dywergencyjne (Guilford 1978).

Można wskazać zarówno podobieństwa, jak i różnice między zdolnościami wytwarzania konwergencyjnego i wytwarzania dywergencyjnego (opracowano na podstawie Guilforda, 1978).

Podobieństwa (cechy wspólne) między wytwarzaniem dywergencyjnym i wytwarzaniem konwergencyjnym są następujące:

- Obie operacje wymagają generowania nowych informacji z informacji danych,
- Wygenerowane informacje są w dużym stopniu uzależnione od zasobów pamięci,
- Obie kategorie operacji umożliwiają wykorzystywanie w rozwiązywaniu problemu/zadania nowych informacji wejściowych, które są dostępne jednostce „pod ręką" albo mogą być dostępne w zasięgu poszukiwań,
- Zarówno wytwarzanie dywergencyjne, jak i konwergencyjne wykorzystują transfer oparty na przypominaniu.

Różnice między porównywanymi operacjami dotyczą: 1. sytuacji problemowej, 2. zakresu poszukiwań, 3. kryteriów powodzenia, 4. specyfiki celu i drogi wytwarzania.

1. **Sytuacja problemowa**. W przypadku wytwarzania konwergencyjnego problem może być (i z reguły jest) ściśle strukturalizowany natomiast w sytuacji wytwarzania dywergencyjnego wymagania dotyczące rozwiązania problemu nie muszą być bardzo dokładnie i sztywno określone. Rozwiązując problem jednostka poszukuje odpowiedzi posługując się metodą prób i błędów, stosuje naprzemiennie operacje wytwarzania dywergencyjnego i oceniania.
2. **Zakres poszukiwań**. Zakres poszukiwań w wytwarzaniu konwergencyjnym jest wąski a liczba poprawnych odpowiedzi jest ograniczona w przeciwieństwie do wytwarzania dywergencyjnego, gdzie zakres poszukiwań jest szeroki, liczba możliwych odpowiedzi jest duża i „rozchodzi się" w różnych kierunkach.
3. **Kryteria powodzenia**. W odniesieniu do wytwarzania konwergencyjnego przyjmuje się ostrzejsze, ścisłe i rygorystyczne kryteria powodzenia. W przypadku wytwarzania dywergencyjnego kryteria te są niewyraźnie, swobodne, ceni się różnorodność i liczbę niestandardowych odpowiedzi (Guilford 1978).
4. **Cel i droga wytwarzania**. Wytwarzanie konwergencyjne ukierunkowane jest na generowanie **konieczności** logicznych, natomiast fundamentem wytwarzania dywergencyjnego jest dedukcja a wraz z nią generowanie logicznych **możliwości**.

Celem przeprowadzonych badań było dokonanie diagnozy właściwości wybranych zdolności uczenia się uczniów szkoły podstawowej i gimnazjum. Sformułowano dwa problemy badawcze:

1. Jakie są właściwości zdolności uczenia się uczniów szkoły podstawowej w zakresie wybranych operacji wyróżnionych przez J.P. Guilforda: wytwarzania konwergencyjnego i wytwarzania dywergencyjnego[4]?
2. Czy różnice indywidualne u uczniów w zakresie wytwarzania dywergencyjnego i konwergencyjnego związane są z płcią i wiekiem?

Narzędzie badań

W badaniach zastosowano test zdolności uczenia się SOI-LA (*Structure of Intellect Learning Abilities Test*) (Meeker i in. 1985). Testy te, opracowane na podstawie wieloczynnikowego modelu inteligencji Guilforda, obejmują 26 podtestów (Filipiak 2015). Każdy z podtestów bada jeden z czynników Guilforda. Jest określony trzyliterowym kodem. W badaniach wykorzystano 3 testy do diagnozy zdolności wytwarzania dywergencyjnego i 4 testy do diagnozy zdolności wytwarzania konwergencyjnego (por. tabela 1 i 2).

Tabela 1. Podtesty formularzy SOI-La wykorzystane w badaniach

WYTWARZANIE KONWERGENCYJNE	
Konwergencyjne wytwarzanie jednostek figuralnych	NFU
Konwergencyjne wytwarzanie systemów symbolicznych	NSS
Konwergencyjne wytwarzanie przekształceń symbolicznych	NST
Konwergencyjne wytwarzanie implikacji symbolicznych	NSI
WYTWARZANIE DYWERGENCYJNE	
Dywergencyjne wytwarzanie jednostek figuralnych	DFU
Dywergencyjne wytwarzanie jednostek semantycznych	DMU
Dywergencyjne wytwarzanie relacji symbolicznych	DSR

Źródło: opracowano na podstawie Meeker 1985.

[4] W badaniach realizowanych w projekcie MNiSW Nr N 107 037 31/38/75 oceniono i scharakteryzowano wszystkie operacje wyróżnione przez J.P. Guilforda. Dane te zawiera raport z badań Filipiak 2009. Wyniki zaprezentowane zostały w artykułach: Filipiak 2012, 2015. W tym artykule szerzej został zaprezentowany wątek operacji wytwarzania dywergencyjnego i wytwarzania konwergencyjnego ze względu na niepokojąco niskie rezultaty jakie uzyskali w tych testach uczniowie.

Tabela 2. Podtesty SOI-La podzielone na kategorie zdolności szkolnych

Zdolności szkolne	Podtest SOI-LA
Zdolności czytania – czytanie zaawansowane	NST
Zdolności arytmetyczne – stosowanie prawidłowości matematycznych	NSS
Osiągnięcia matematyczne – umiejętność logicznego myślenia	NSI
Zdolności pisania – gotowość psychoruchowa	NFU
Kreatywność w zakresie figur i przedmiotów	DFU
Kreatywność w zakresie prawidłowości matematycznych	DSR
Kreatywność w zakresie słów i pojęć	DMU

Źródło: opracowano na podstawie Meeker 1985.

Po przeprowadzeniu testów formularze SOI-LA były ręcznie oceniane przez członka zespołu badawczego przeszkolonego i mającego doświadczenie badawcze w projektach z wykorzystaniem testów opartych o teorię Guilforda (Filipiak 2009, 2015).

Osoby badane

Badania przeprowadzono w środowisku wielkomiejskim (Bydgoszcz). Wytypowano szkoły, w których uczniowie osiągali wyniki średnie w testach osiągnięć szkolnych. W badaniach wzięło udział 569 uczniów (323 uczniów SP i 193 uczniów gimnazjum). Kryterium doboru osób do badań stanowiły dwie zmienne: wiek/klasa i płeć ucznia. Wyodrębniono trzy grupy wiekowe w szkole podstawowej: uczniowie klas III, IV, VI oraz dwie grupy wiekowe uczniów klas gimnazjalnych (I i III) (por. Filipiak 2009, 2011a, 2015), (tabela 3).

Tabela 3. Liczba badanych uczniów klas III i VI szkół podstawowych

Grupa: 0 – M; 1 – K	Klasa		Liczba
SZKOŁA PODSTAWOWA			
0 0	III	M	50
0 1		K	48
1 0	IV	M	
1 1		K	
2 0	VI	M	53
2 1		K	44
Razem szkoła podstawowa			162(M) + 161(K) = 323 (SP)

Grupa: 0 – M; 1 – K	Klasa		Liczba
GIMNAZJUM			
3 0	I	M	42
3 1		K	47
4 0	III	M	55
4 1		K	49
Razem gimnazjum			97(M) + 96(K) = 193(G)
RAZEM			259 (M) + 257 (K) = 516 (SP+G)

Źródło: badania własne (Filipiak 2009).

Organizacja badania

Badania zdolności uczenia się z wykorzystaniem Testów SOI-La przeprowadzono zbiorowo – w grupach 20–25 osobowych. Czas wypełniania każdego z podtestów jest limitowany. Prowadząc badania w grupie dzieci ze szkoły podstawowej, testy przeprowadzono na czterech spotkaniach 45 minutowych. Badania realizowane w gimnazjum były prowadzone w dwóch sesjach 90 minutowych.

Wyniki badań

Poziom i właściwości zdolności uczenia się w zakresie operacji wytwarzania konwergencyjnego i dywergencyjnego uczniów szkoły podstawowej i gimnazjum

Rezultaty (średnie wyniki) uzyskane przez uczniów w testach SOI-LA przedstawiono w tabeli 4 oraz na wykresach (por. Filipiak 2009, 2015). Na podstawie tych wyników wykreślono dla każdej grupy profile, odnosząc je do 7 poziomów wyróżnionych przez Meeker, Meekera i Roid (1985): „uzdolniony", „wybitny", „powyżej przeciętnej", „przeciętny", „poniżej przeciętnej", „ograniczający" oraz „upośledzający".

Tabela 4. Zdolności uczenia się uczniów szkoły podstawowej i gimnazjum – operacje wytwarzania konwergencyjnego i wytwarzania dywergencyjnego (wartości średnie dla grup)

Klasa	Płeć	Zdolności uczenia się						
		Wytwarzanie konwergencyjne				Wytwarzanie dywergencyjne		
		NFU	NSS	NST	NSI	DFU	DMU	DSR
SZKOŁA PODSTAWOWA								
III	M	20,3	2,6	66,4	12,0	15,0	26,8	38,3
	K	27,7	2,7	82,5	13,2	19,3	41,2	43,3
IV	M	19,1	2,8	72,5	11,5	16,8	36,6	43,7
	K	21,9	2,8	93,4	15,3	19,4	55,9	50,2
VI	M	22,1	3,1	97,1	14,3	17,5	46,3	50,7
	K	23,9	3,1	108,1	17,7	17,2	59,4	49,1
GIMNAZJUM								
I	M	28,0	5,1	107,3	16,5	14,1	47,4	45,7
	K	30,7	3,3	112,3	18,0	14,8	64,6	44,1
III	M	27,4	5,3	115,3	17,5	17,3	47,1	53,5
	K	28,9	6,2	144,7	19,6	18,8	68,0	60,3

Źródło: badania własne, Filipiak 2009.

Wytwarzanie dywergencyjne

Zdolności oceniane przez podtesty wytwarzania dywergencyjnego mają związek z rozwojem takich umiejętności szkolnych, jak: kreatywność w zakresie figur i przedmiotów (DFU), kreatywność w zakresie prawidłowości matematycznych (DSR) oraz kreatywność w zakresie zdolności językowych, słów i pojęć (DMU). W testach wytwarzania dywergencyjnego uczniowie szkoły podstawowej i gimnazjum uzyskali niepokojąco słabe rezultaty. Najniższe wyniki uzyskali uczniowie w teście DFU (dywergencyjnego wytwarzania jednostek figuralnych). Test ten badał zdolność komunikowania się za pomocą wyobrażeń figuralnych, zdolność do wykorzystywania dwuznacznych bodźców w kreatywny sposób. Podczas rozwiązywania tego testu uczniowie mogli wykazać się talentem do tworzenia komiksów, projektowania, szkicowania. Niestety rezultaty aż 8 (z 10) grup kształtują się na poziomie określonym jako „ograniczający". Są to wyniki

uczniów ze wszystkich wyodrębnionych ze względu na wiek x płeć grup szkoły podstawowej oraz uczniów III klasy gimnazjum. Rezultaty uczniów I klasy gimnazjum sklasyfikowano na poziomie „upośledzającym". Słabe wyniki uzyskane przez uczniów w tym teście mogą sugerować brak płynności w zakresie pomysłowości (twórczego wytwarzania). Nieco korzystniejsze są wyniki badań uczniów w teście DMU – dywergencyjnego wytwarzania jednostek semantycznych. Test ten diagnozował zdolność kreatywnego pisania i tworzenia oryginalnych pojęć w ograniczonym czasie. Najlepsze rezultaty – określone jako „powyżej przeciętnej" – w tym teście uzyskały dziewczynki z klasy III i IV szkoły podstawowej. Uczniowie z klasy III gimnazjum usytuowali się na poziomie „przeciętnym". Poziom „poniżej przeciętnej" charakteryzuje wyniki chłopców z klasy III i IV szkoły podstawowej, dziewcząt z klasy VI oraz dziewcząt z klasy I gimnazjum. Najniżej, bo na poziomie „ograniczającym" sklasyfikowano rezultaty uzyskane w tym teście przez **chłopców** z klasy VI szkoły podstawowej, I i III gimnazjum. W teście dywergencyjnego wytwarzania jednostek semantycznych w każdej grupie wiekowej zdecydowanie lepsze rezultaty uzyskały dziewczynki niż chłopcy (por. tabela 4).

Trzeci z testów **(DSR)** diagnozował dywergencyjne wytwarzanie relacji symbolicznych. Test ten można określić inaczej jako kreatywne, symboliczne rozwiązywanie problemów. Do wypełnienia tego testu niezbędna była znajomość pojęć matematycznych przez uczniów. Dobre rezultaty sklasyfikowane na poziomie „przeciętnym" uzyskały dzieci z klasy III i IV szkoły podstawowej. Przy czym wyniki dziewcząt były znacznie lepsze niż chłopców. Na tym poziomie sytuują się także rezultaty uzyskane przez dziewczęta z klasy III gimnazjum. Poziom „poniżej przeciętnej" cechuje rezultaty uczniów z klasy VI oraz chłopców z klasy III gimnazjum. Najniżej w teście zostały ocenione zdolności uczniów z klasy I gimnazjum. Wskazuje to prawdopodobne na ich trudności w przyswajaniu pojęć matematycznych, brak odwagi w poszukiwaniu i odkrywaniu następstw ich zastosowania w ćwiczeniach.

Wytwarzanie konwergencyjne

Wytwarzanie konwergencyjne jest zdolnością do rozwiązywania problemów i trzymania się zasad w celu wygenerowania informacji z informacji danej, przy wykorzystaniu z góry określonej odpowiedzi. Zdolności oceniane przez podtesty wytwarzania konwergencyjnego mają związek z rozwojem umiejętności szkolnych: umiejętności czytania (NST), pisania (NFU), zdolności arytmetycznych (NSS, NSI)

(por. tabela 2). Bardzo dobre są rezultaty jakie odnotowano w testach związanych z zdolnościami arytmetycznymi u dzieci kończących I poziom edukacji szkolnej tj. uczniów klas III i uczniów klas IV. Dotyczy to testów NSS oraz testu NSI. Pierwszy z testów **konwergencyjnego wytwarzania systemów symbolicznych (NSS)** wymaga przy jego wypełnianiu doskonałego opanowania podstaw arytmetyki. Diagnozuje zdolność rozwiązywania złożonych, niewerbalnych zadań matematycznych przy jednoczesnym wykorzystaniu zasad matematycznych. W teście tym, poziom „utalentowany" odnosi się do wyników badań zdolności dzieci z klasy III, a poziom „wybitny" charakteryzuje zdolności dzieci z klasy IV. Na poziomie już tylko „przeciętnym" sytuują się rezultaty uczniów z klas VI (i dziewcząt z klasy III gimnazjum). W kolejnych latach edukacji obserwujemy dalszy regres. Rezultaty chłopców z klas I i III gimnazjum umiejscowiono na poziomie „poniżej przeciętnej" a dziewcząt z klasy I – na niekorzystnym i wymagającym interwencji poziomie „ograniczającym".

Podobną tendencję można zaobserwować w odniesieniu do testu diagnozującego zdolność **konwergencyjnego wytwarzania implikacji symbolicznych (NSI)**. Jest to test logicznego myślenia i rozumowania. Wymaga zdolności wyciągania wniosków symbolicznych na podstawie danych informacji symbolicznych, które wcześniej nie były znane osobie. Te zdolności zależą od opanowania podstawowych zdolności, takich jak: poznawanie, zapamiętywanie i ocenianie w wymiarze symbolicznym i semantycznym. W teście tym bardzo dobre rezultaty uzyskali uczniowie szkoły podstawowej. Sugeruje to dobre fundamenty dla dalszego kształcenia matematycznego oraz możliwości rozwijania krytycznego myślenia w naukach społecznych. Poziom „wybitny" (niżej jednak niż w teście NSS) uzyskali uczniowie klas III szkoły podstawowej i dziewczynki z klasy IV. Dobry, „powyżej przeciętnej" poziom prezentują chłopcy z klasy IV i dziewczęta z klasy VI szkoły podstawowej. Na poziomie przeciętnym określono zdolności w tym teście chłopców z klasy VI, dziewcząt z klas I i III gimnazjum. Niepokojąco niskie („poniżej przeciętnej") określono rezultaty chłopców z klas I i III gimnazjum. Uczniowie ci mogą mieć trudności w zrozumieniu nowych relacji wprowadzanych do systemu.

Test **konwergencyjnego wytwarzania jednostek figuralnych (NFU)** diagnozuje zdolności psychoruchowe. Na podstawie jego wyników można stwierdzić, że w zasadzie szybkość pisania diagnozowana u uczniów szkoły podstawowej i gimnazjum odpowiada wymaganiom stawianym przez szkołę. Uczniowie są gotowi do odtwarzania figur w odpowiednim tempie. Najlepsze rezultaty w tym teście uzyskały dziewczynki z klasy III szkoły podstawowej („wybitne") a naj-

niższe należą do chłopców z klasy IV („poniżej przeciętnej"). Rezultaty pozostałych uczniów w większości sytuują się na poziomie „przeciętnym".

Czwarty z testów oceniał zdolność **konwergencyjnego wytwarzania przekształceń symbolicznych (NST)**. Diagnozował szybkość rozpoznawania słów i pozwolił określić czy uczeń nadąża z zadaniami związanymi z czytaniem. W tym teście tylko jedna z grup dzieci uczestniczących w badaniach uzyskała poziom określony jako „powyżej przeciętnej". Były to dziewczynki z klasy III szkoły podstawowej. Niekorzystne, poniżej przeciętnej rezultaty określono w odniesieniu do grupy chłopców z klasy IV, uczniów klasy VI, kończących szkołę podstawową oraz uczniów klasy I gimnazjum. Uczniowie ci mogą mieć trudności z dokańczaniem realizowanych zadań oraz wolne tempo czytania.

Wiek i płeć jako czynniki różnicujące zdolności uczenia się uczniów

Różnice indywidualne w zakresie poziomu i właściwości zdolności uczenia się u dzieci są związane zarówno z płcią jak i z wiekiem uczniów (tabela 5).

Tabela 5. Poziom istotności statystycznej – klasa, płeć i ich interakcja

Efekt czynnika	Lambda Wilksa	R Rao	df1	df2	Poziom p	Interakcja
1	0,11	14,32	105,00	2630,00	< 0,001	Istotny wpływ klasy
2	0,69	11,36	21,00	537,00	< 0,001	Istotny wpływ płci
1x2	0,72	1,75	105,00	2630,00	< 0,001	Istotny wpływ interakcji klasa x płeć

Źródło: badania własne, Filipiak (2009); Filipiak (2015).

W każdej grupie wiekowej dziewczynki uzyskiwały wyższe rezultaty niż chłopcy. Rezultaty młodszych uczniów (szkoła podstawowa) są korzystniejsze niż uczniów starszych (gimnazjum).

Wnioski. Rekomendacje dla praktyki edukacyjnej

Badania pokazują, że w zakresie wyodrębnionych wśród operacji występujących w modelu Guilforda zdolności wytwarzania dywergencyjnego i konwergencyjnego uczniowie szkoły podstawowej i gimnazjum wykazują zróżnicowany poziom.

Zdecydowanie korzystniej prezentują się rezultaty uczniów w zakresie zdolności wytwarzania konwergencyjnego niż dywergencyjnego. Wskazuje to, że uczniowie zarówno szkoły podstawowej, jak i gimnazjum potrafią sobie radzić w sytuacjach typowych, w których wąski jest zakres poszukiwań a liczba poprawnych odpowiedzi jest ograniczona. Być może jest to też konsekwencją funkcjonowania w szkole, w której uczy się „pod testy". Bardzo dobre rezultaty dzieci ze szkoły podstawowej, zwłaszcza uczniów kończących I etap edukacji szkolnej w zakresie zdolności arytmetycznych, w teście, w którym dzieci wykazały się umiejętnością rozwiązywania złożonych, niewerbalnych zadań matematycznych przy jednoczesnym wykorzystaniu zasad matematycznych mogłyby napawać optymizmem. Niestety tak dobrze zapowiadający się rozwój zdolności dzieci kończących I etap edukacji szkolnej zostaje skutecznie wyhamowywany w kolejnych latach edukacji (Filipiak 2011a, 2015). Poziom „utalentowany" dziewczynek z klasy III zostaje zastąpiony poziomem „ograniczającym" dziewcząt w klasie I gimnazjum w teście Konwergencyjnego wytwarzanie systemów symbolicznych. Czy jest to zatem potwierdzenie wielokrotnie powtarzanej tezy Doroty Klus-Stańskiej (2008) o „amputowanych kompetencjach"?

> [...] szkoła tak rozwija określone kompetencje i takie kompetencje rozwija, że zamiast wspierać rozwój umysłowy ucznia, zaniedbuje i ogranicza rozwój kompetencji dla niego istotnych (Klus-Stańska 2008: 43).

Najbardziej niepokojące wyniki odnotowano w testach wytwarzania dywergencyjnego. Wyłaniający się z nich obraz ucznia szkoły podstawowej i gimnazjum wywołuje obawę. Na podstawie analizy wyników badań w zakresie zdolności wytwarzania konwergencyjnego i dywergencyjnego można wnioskować, że uczniowie nie są kreatywni. Najwięcej problemów mają z kreatywnością w zakresie wytwarzania figuralnego. Dobre zadatki pączkujących zdolności w tym zakresie (znajdujące się w strefie najbliższego rozwoju), ujawniane przez dziewczynki z klas III nie zostały rozwinięte w kolejnych latach edukacji. Może to w przyszłości skutkować trudnością działania jednostek w sytuacjach niestereotypowych oraz narastającymi trudnościami procesu przetwarzania informacji. Badani uczniowie potrafią szybko pisać i przepisywać, ale nie potrafią myśleć twórczo i problemowo. Mają podstawy – fundament dla rozwijania zdolności myślenia matematycznego, ale niestety zostaje on zburzony i niewykorzystany dla dalszych działań. W kolejnych latach edukacji uczniowie nabywają trudności w przyswajaniu pojęć matematycznych, gubią odwagę w odkrywaniu następstw, relacji i we wnioskowaniu.

Trudno jednoznacznie winić szkołę za obserwowany regres w zakresie analizowanych zdolności wytwarzania dywergencyjnego i konwergencyjnego. Nie można jednak zaprzeczyć, że innowacyjności w działaniu i rozwiązywaniu problemów nie sprzyjają sytuacje edukacyjne w szkole „pracującej pod testy", które nie uwzględniają zadań odkrywających prawidłowości matematyczne. Wspieranie rozwoju myślenia uczniów jest jednak możliwe dzięki tworzeniu bazy dla konstruowania symbolicznej wiedzy matematycznej i poprzez różnorodność oraz bogactwo zadań problemowych. Wymaga jednak zmiany filozofii edukacyjnej nauczyciela, który organizując proces nauczania – uczenia się zdaje sobie sprawę, jak ważny jest sam proces konstruowania wiedzy, strategie dochodzenia do wyniku/wytworu a nie sam wytwór (Dąbrowski 2008; Klus-Stańska; Kalinowska 2004). Od wielu już lat propagowane są programy wychowania do twórczości, które niestety z trudem przenikają do codzienności szkolnej. Myślenie (twórcze) może być przedmiotem nauczania tak samo ważnym jak inne przedmioty szkolne[5] (por. Nęcka 2002: 202). Przykładem takiego podejścia jest *Odyseja umysłu*, lekcje twórczości w projekcie Szmidt (i in. 2003) *Porządek i przygoda*, projekty Wiesławy Limont wykorzystujące zasady synektyki (1994), Nęcki (i in. 2005), Góralski (1990) i inni.

Pozostaje pytanie, na które nadal musimy poszukiwać odpowiedzi: Jakie warunki powinniśmy spełnić, aby stworzyć stymulującą kulturę szkolną, która umożliwi jednostkom skuteczne wykorzystanie własnego potencjału i potencjału szeroko rozumianej kultury? Jaka szkoła przygotuje do życia w świecie przesyconym informacjami?

Bibliografia

Bandura L. 1975. *Próba rozwijania myślenia twórczego w szkole podstawowej*, „Kwartalnik Pedagogiczny", nr 3, s. 97–110.

Bauman Z. 2011. *44 listy ze świata płynnej nowoczesności*, tłum. T. Kunz, Wydawnictwo Literackie, Kraków.

Dąbrowski M. 2008. *Pozwólmy dzieciom myśleć*, CKE, Warszawa.

Dobrołowicz W. 1995. *Psychodydaktyka kreatywności*, WSPS, Warszawa.

Dylak S. 2008. *Szkoła – jaka jest, każdy widzi, jak ma być, każdy wie, ale jaka mogłaby być, podpowiada Neil Postman*, [w:] *Dokąd zmierza polska szkoła*, red. D. Klus-Stańska, Wydawnictwo Akademickie Żak, Warszawa, s. 61–83.

[5] W Katedrze Dydaktyki i Studiów nad Kulturą Edukacji, w Instytucie Pedagogiki Uniwersytetu Kazimierza Wielkiego w Bydgoszczy od kilku już lat prowadzony jest autorski przedmiot Rozwijanie zdolności uczenia się.

Filipiak E. 2002. *Konteksty rozwoju aktywności językowej dzieci w wieku wczesnoszkolnym*, Wydawnictwo Uczelniane Akademii Bydgoskiej, Bydgoszcz.

Filipiak E. 2009. *Rozwój zdolności uczenia się uczniów szkoły podstawowej i gimnazjum. Raport z badań*, Uniwersytet Kazimierza Wielkiego (maszynopis niepublikowany, przygotowany w ramach grantu MNiSW Nr N 107 037 31/3875), Bydgoszcz.

Filipiak E. 2011a. *Produkty kultury uczenia się uczniów szkoły podstawowej i gimnazjum*, „Forum Oświatowe", 3, s. 159–185.

Filipiak E. 2011b. *Motywacja i strategie uczenia się uczniów szkoły podstawowej i gimnazjum*, „Edukacja", 2, s. 28–36.

Filipiak E. 2012. *Rozwijanie zdolności uczenia się*, GWP, Sopot.

Filipiak E. 2015. *Zdolności uczenia się absolwentów I i II etapu edukacyjnego*, „Edukacja", 1.

Guilford J.P. 1978. *Natura inteligencji człowieka*, tłum. B. Czarniawska, W. Kozłowski, J. Radzicki, PWN, Warszawa.

Góralski A. 1990. *Być nowatorem. Poradnik twórczego myślenia*, PWN, Warszawa.

Kin-Wiśniewska M. 2008. *Zdolności poznawcze dzieci w wieku 7–10 lat*, „Problemy Wczesnej Edukacji", 1, s. 127–139.

Klus-Stańska D., Kalinowska A. 2004. *Rozwijanie myślenia matematycznego młodszych uczniów*, Wydawnictwa Akademickie Żak, Warszawa.

Klus-Stańska D. 2008a. *Mitologia transmisji wiedzy, czyli o konieczności szukania alternatyw dla szkoły, która amputuje rozum*, „Problemy Wczesnej Edukacji", 2, s. 35–45.

Klus-Stańska D. red. 2008b. *Dokąd zmierza polska szkoła*, Wydawnictwo Akademickie Żak, Warszawa.

Kozielecki M. 1998. *List o edukacji*, „Forum Oświatowe", 2 (19), s. 103–106.

Ledzińska M. 2008. *Fakty i mity na temat uczniów zdolnych*, „Problemy Wczesnej Edukacji", nr 1, s. 49–59.

Lewowicki T. 1986. *Kształcenie uczniów zdolnych*, WSiP, Warszawa.

Limont W. 1994. *Synektyka a zdolności twórcze*, Wydawnictwo Uniwersytetu Mikołaja Kopernika, Toruń.

Meeker M., Meeker R. 1975. *S.O.I. Learning Abilities Test examiner's manual*, SOI Institute, El Segundo.

Meeker M., Meeker R., Roid G.H. 1985. *Structure of Intellect Learning Abilities Test (SOI-LA)*, Western Psychological Services, Melton South.

Meeker M., Meeker R., Roid G.H. 1985. *Structure of Intellect Learning Abilities Test (SOI-LA).* Instrukcja do badań. Formularz A i B. (niepublikowana wersja polska dostępna w Instytucie Pedagogiki UKW; tłum. J. Puchalska. Oprac. naukowe i merytoryczne: E. Filipiak).

Nęcka E. 2002. *Psychologia twórczości*, GWP, Sopot.

Nęcka E. 2009. *Inteligencja jest procesem*, [w:] *Nowe idee w psychologii. Psychologia XXI wieku*, red. J. Kozielecki, GWP, Sopot, s. 21–40.

Nęcka E., Orzechowski J., Słabosz A., Szymura B. 2005. *Trening twórczości*, GWP, Sopot.

Nosal C. red. 1992. *Twórcze przetwarzanie informacji*, Drukarnia Agencji DELTA, Wrocław.

Pietrasiński Z. 1969. *Myślenie twórcze*, PZWS, Warszawa.

Pietrasiński Z. 1976. *Zdolności*, [w:] *Psychologia*, T. Tomaszewski, PWN, Warszawa, s. 735–762.

Sękowski A. red. 2005. *Psychologia zdolności. Współczesne kierunki badań*, WN PWN, Warszawa.

Strelau J. 1978. *Przedmowa*, [w:] J.P. Guilford, *Natura inteligencji człowieka*, tłum. B. Czarniawska, W. Kozłowski, J. Radzicki, PWN, Warszawa, s. 5–20.

Strelau J. 2014. *Różnice indywidualne. Historia – Determinanty – Zastosowania*, Wydawnictwo Naukowe Scholar, Warszawa.

Szmidt K. 2003. *Koncepcja rozwijania zdolności twórczych uczniów klasy III do programu kształcenia zintegrowanego Przygoda z klasą*, WSiP, Warszawa.

Szmidt K.J. 2007. *Pedagogika twórczości*, GWP, Sopot.

Szymański M.S. 1987. *Twórczość i style poznawcze uczniów*, WSiP, Warszawa.

KWARTALNIK PEDAGOGICZNY
2015 NUMER 4(238)

Barbara Ciżkowicz
Uniwersytet Kazimierza Wielkiego w Bydgoszczy

Zastosowanie SEM[1] do oceny różnic w uwarunkowaniach osiągnięć matematycznych polskich 15-latków w badaniach PISA 2003 i 2012

Summary

THE APPLICATION OF SEM TO ASSESS DIFFERENCES IN THE CONDITIONS OF MATHEMATICS ACHIEVEMENTS OF POLISH 15-YEAR-OLDS IN PISA 2003 AND 2012

The article compares the learning environments in the years 2003 and 2012 in order to clarify the reasons for the significant increase in the mathematical skills level of Polish 15-year-olds in those years as pointed by PISA studies. We studied the effect of school and family factors as well as students' non-cognitive achievements on mathematics scores in PISA. In addition, SEM was used in subgroups in order to check whether the role of investigated factors changed over time. The research model adopted allowed to explain a large part of variability of students' mathematical achievements.

Analyses carried out on a sample of 8,966 people revealed significant, positive changes in the students' family environment: conditions at home improved and parents' educational level increased. The impact of home factor on non-cognitive student achievement intensified. School principals pointed to the favourable changes in the school environment. In turn, students' attituds towards school became less optimistic and their sense of belonging weakened. Also the non-cognitive performance of pupils deteriorated over time.

Key words: achievement in mathematics, motivation, anxiety, self-esteem, the impact of family and school, structural equation modelling.

red. Paulina Marchlik

[1] SEM (ang. *Structural Equation Modelling*) – modelowanie równań strukturalnych.

Diagnoza edukacyjna, rozumiana jako „rozpoznawanie warunków, przebiegu i wyników uczenia się", stanowi niezbędne ogniwo w strukturze intencjonalnych działań edukacyjnych na rzecz rozwoju dzieci, młodzieży i osób dorosłych (Niemierko 2007: 365). Jej wysoka jakość oraz wykorzystanie informacji zwrotnej w doskonaleniu procesu kształcenia jest warunkiem koniecznym skuteczności tych działań. Diagnoza z bliska prowadzona w praktyce nauczycielskiej była przez lata głównym źródłem informacji o uczniu (Niemierko 2013; Szyling 2013). Zachodzące zmiany w systemie edukacji spowodowały potrzebę wzbogacenia diagnozy z bliska poprzez diagnozę z daleka. Od 1999 roku w polskich szkołach zaczęto wprowadzać egzaminy zewnętrzne tworzące w skali kraju wspólny układ odniesienia dla uczniów. Polska uczestniczy też w międzynarodowych badaniach umiejętności uczniów. Należą do nich badania PIRLS i TIMSS obejmujące uczniów 10-letnich oraz PISA dotyczące młodzieży 15-letniej (Jakubowski 2013; Ciżkowicz 2013a; Ciżkowicz 2013b). Te dwie perspektywy diagnozy z bliska i z daleka pozostają względem siebie komplementarne dostarczając wzajemnie uzupełniających się informacji. Pierwsza z nich przeprowadzana przez nauczyciela dotyczy dyspozycji ucznia, w których kształtowaniu nauczyciel uczestniczy; druga pozwala na ocenę efektywności kształcenia na tle szerszego układu odniesienia.

Badania PISA

Program Międzynarodowej Oceny Umiejętności Uczniów (PISA *Programme for International Student Assessment*), zapoczątkowany przez OECD (*Organisation for Economic Co-operation and Development*) w 2000 roku, dotyczy diagnozy kluczowych kompetencji uczniów zbliżających się do końca obowiązkowej edukacji. Badania te mają na celu sprawdzenie stopnia przygotowania młodzieży do życia poza szkołą (PISA 2012a). W odstępach trzyletnich diagnozowane są umiejętności uczniów w trzech dziedzinach: matematyce, czytaniu i interpretacji oraz rozumowaniu naukowym. W każdym badaniu jedna z dziedzin jest dziedziną wiodącą. Zastosowana metodologia pozwala nie tylko określać poziom umiejętności młodzieży, ale też śledzić występujące w nim zmiany.

W roku 2012 w badaniach PISA umiejętności matematyczne były po raz drugi dziedziną wiodącą (poprzednio w 2003 roku). Średni wynik polskich piętnastolatków z matematyki (518; SE = 3,40) zapewnił im pierwszą lokatę wśród krajów UE, na równi z uczniami Holandii, Estonii i Finlandii (PISA 2012b). Odnotowano to, jako znaczący sukces, szczególnie w kontekście badań z roku 2003,

w których polscy uczniowie uzyskali z matematyki wynik (490; SE = 2,96) poniżej średniej w krajach OECD (PISA 2003a). Szczegółowe analizy wyników przedstawione zostały w Raporcie krajowym (OECD 2012a).

Oprócz diagnozy umiejętności uczniów w PISA, przeprowadzane są badania kwestionariuszowe dostarczające wiedzy na temat szerokiego spektrum zmiennych związanych z procesem kształcenia i wychowania. Zmienne te dotyczą rodziny, szkoły oraz ucznia (Strelau 2000; OECD 2003; OECD 2012b). Zgromadzone dane empiryczne pozwalają na porównania uwarunkowań procesu kształcenia krajów uczestniczących w badaniach.

Wyjaśnienie podstawowych pojęć

W wychowaniu rozumianym, jako świadomie organizowanej działalności społecznej mającej na celu wywołanie zamierzonych zmian w osobowości wychowanka w sferze poznawczo-instrumentalnej oraz emocjonalno-motywacyjnej (Okoń 2007), naczelną rolę odgrywa rodzina i szkoła należące do najważniejszych środowisk wychowawczych (Fleming 2010; Kyunghee 2011).

Rodzina jest naturalnym środowiskiem życia i wychowania młodego pokolenia. Kształtowanie osobowości, systemu wartości, postaw społecznych dzieci i młodzieży pozostaje pod wpływem oddziaływań środowiska rodzinnego. Wychowawcze funkcje rodziny powinny być ściśle zsynchronizowane z działaniami szkoły (Izdebska 1993). Znaczenie środowiska rodzinnego dla osiągnięć szkolnych potwierdzają liczne badania (Connolly 2006, Gil-Flores i in. 2011). Podkreślany jest wpływ zasobów rodziny i kapitału kulturowego na poziom wykształcenia dzieci. Wskazywane są istotne związki między osiągnięciami dzieci i dochodem rodziny, statusem zawodowym i poziomem wykształcenia rodziców (Smith i in. 1997; Shonkoff, Phillips 2000; Bradshaw 2002; Feinstein i in. 2008; Strelitz, Lister 2008). W krajach, w których uczniowie osiągają wysokie wyniki z matematyki (Korea, Tajwan, Japonia) czynnik rodziny wykazuje silny związek z tymi osiągnięciami (Liu i in. 2006). Natomiast wpływ czynnika dziedziczenia na poziom wykształcenia wskazywany w różnych doniesieniach badawczych waha się od 18% do 77% (Branigan i in. 2013).

Szkoła jest miejscem intensywnych interakcji społecznych. Służą one w znacznej mierze wywoływaniu zamierzonych zmian w uczniu. Podkreślany jest wpływ organizacji instytucjonalnej szkoły na osiągnięcia uczniów (Greenwald i in. 1996). Wskazywany jest też negatywny wpływ wielkości szkoły na osiągnięcia

matematyczne (Lee i in. 1997). Mniej jednoznaczne wnioski dotyczą wielkości klasy. Azjatyccy badacze dowodzą, że zmniejszanie wielkości klasy nie poprawia jakości nauczania-uczenia się (Wang, Lin 2005). Istotna dla realizacji procesu kształcenia jest również atmosfera w klasie i w szkole. Podmiotowej interakcji, w której kontakty nauczyciel-uczeń polegają na współpracy i dialogu, towarzyszy wzajemne poszanowanie. Taki model interakcji między nauczycielem i uczniami zazwyczaj przenosi się na relacje między nauczycielami oraz między uczniami, a oferta wychowawcza dostosowana do oczekiwań i aspiracji uczniów pozytywnie oddziałuje na motywację do uczenia się (Strelau 2000).

Pozapoznawcze osiągnięcia uczniów

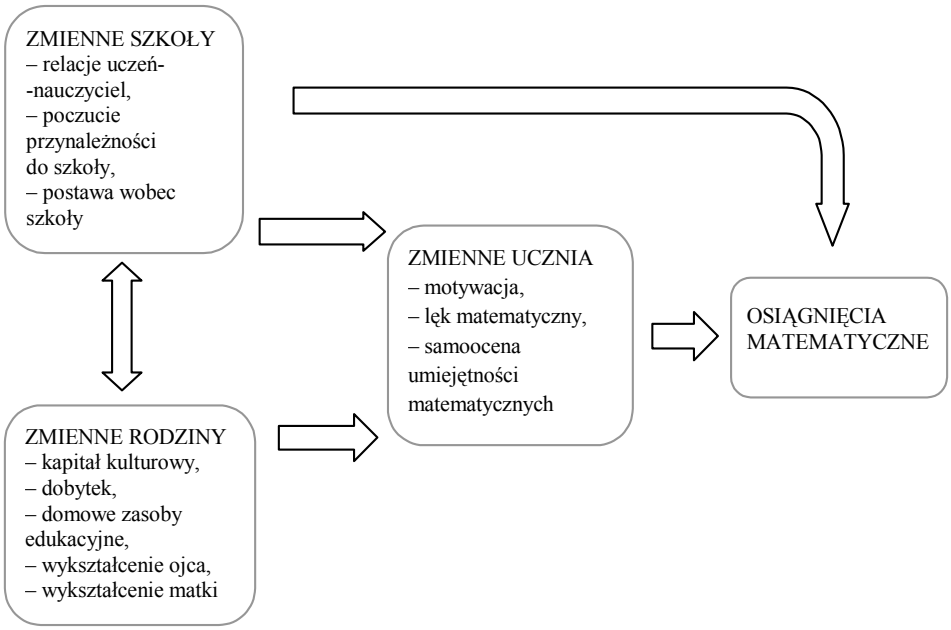
W uczeniu się matematyki powszechnie uznaje się znaczący wpływ czynnika o nazwie *lęk matematyczny* (Hembree 1990). Badania wskazują na związek matematycznego lęku z obawą o matematyczną porażkę (Richardson, Woolfolk 1980) oraz z unikaniem matematyki i zadań matematycznych (Krinzinger i in. 2009).

Samoocena uzdolnień matematycznych jest mediatorem związku między lękiem matematycznym a osiągnięciami matematycznymi. Poczucie zagrożenia okazuje się nieistotne dla niskich osiągnięć szkolnych, jeśli uwzględnia się samoocenę własnych zdolności (Strelau 2000).

Motywacja jest względnie stałą tendencją człowieka do realizowania określonych celów, zadań życiowych i wartości (Strelau 2000). O motywacji wewnętrznej w uczeniu się mówimy, jeśli czynność ta sprawia przyjemność i pobudza na tyle ciekawość ucznia, że niepotrzebna jest dodatkowa gratyfikacja, by tę czynność uczeń wykonywał (Deci, Ryan 1985). Motywacja zewnętrzna jest działaniem podejmowanym wskutek zewnętrznych wzmocnień. Badania pokazują, że motywacja wewnętrzna pozostaje pod wpływem motywacji zewnętrznej. Zainteresowania i twórczość w danej dziedzinie można ćwiczyć i wzmacniać nagradzaniem (Eisenberger, Selbst 1994; Eisenberge, Cameron 1996; Eisenberge i in. 1999).

Celem prezentowanych analiz jest wyłonienie, z grupy zmiennych kontrolowanych w PISA, tych zmiennych związanych z rodziną, szkołą i uczniem, które w sposób znaczący wyjaśniają poziom umiejętności matematycznych polskich 15-latków. Zmienne oraz założone relacje między nimi prezentuje model badawczy (ryc. 1). Ponadto analizy mają ujawnić istotne zmiany, które zaszły w poziomie tych zmiennych oraz sile związków między zmiennymi po dziewięciu latach,

czyli po czasie kiedy matematyka stała się ponownie przedmiotem wiodącym w badaniach PISA.



Ryc.1. Model badawczy

Metoda

Badani

Analizie poddano dane dotyczące polskich uczniów uczestniczących w badaniach PISA 2003 i 2012. Zostały one zaczerpnięte z baz danych PISA (OECD 2003; PISA, 2012b). Próby są próbami losowymi. Schemat doboru ma charakter dwustopniowego doboru warstwowego (OECD 2012a). W badaniach PISA w 2003 roku uczestniczyło 4359 polskich 15-latków, wśród których było 50,2% dziewcząt. Zdecydowana większość z nich (99,3%) z nich uczęszczało do szkół publicznych. Badaniami przeprowadzanymi w 2012 roku objęto 4607 osób, z czego 51,2% stanowiły dziewczęta. W porównaniu z rokiem 2003 zmniejszył się nieco odsetek uczniów reprezentujących szkoły publiczne (97%).

Pomiar

Umiejętności matematyczne są zmienną zależną w prezentowanych analizach. Jej poziom badany jest za pomocą wystandaryzowanych testów umiejętności, w których wyodrębniono pięć wskaźników empirycznych dotyczących różnych grup umiejętności począwszy od prostych, odtwórczych, po myślenie strategiczne oraz rozumowanie i argumentację (α Cronbacha = 0,98). W badaniach PISA poziom umiejętności uczniów wyrażany jest na skali akademickiej (M = 500; SD = 100).

Dane dotyczące zmiennych objaśniających zostały zaczerpnięte z badań kwestionariuszowych ucznia (PISA 2003b; OECD, 2012b). Zmienne te można podzielić na czynniki związane ze środowiskiem rodzinnym, szkołą i uczniem.

Środowisko rodzinne

Kapitał kulturowy jest zmienną, której wskaźnikiem empirycznym jest wielkość domowego księgozbioru. Zmienna ta przyjmuje wartości od 1 do 6. Wartość jeden – oznacza księgozbiór zawierający od 0 do 10 książek; wartość dwa – od 11 do 25 książek. Dla kolejnych wartości górne granice przedziałów są odpowiednio do: 100; 200; 500; więcej niż 500 książek.

Wykształcenie rodziców – na podstawie dokładniejszych danych dotyczących wykształcenia rodzice zostali podzieleni na grupę osób z wykształceniem średnim lub niższym i drugą – z wykształceniem wyższym.

Dobrobyt rodziny w badaniach PISA 2012 obejmuje 22 wskaźniki. Jednak ze względu na łączenie zbiorów uwzględniono tylko 11 wskaźników (np. „posiadanie biurka do nauki", „własnego pokoju", „komputera", „Internetu", „programów edukacyjnych" itd.), które występowały również w 2003 roku.

Domowe zasoby edukacyjne są składową *dobrobyt rodziny* i są badane za pomocą podzbioru (7 z 11) wskaźników *dobrobytu rodziny*.

Szkoła

Relacje uczeń-nauczyciel badane są skalą składającą się z pięciu pozycji (np. „uczniowie mają dobre relacje z większością nauczycieli"). Uczniowie udzielali odpowiedzi na czterostopniowej skali od „całkowicie się nie zgadzam" do „całkowicie się zgadzam". Odpowiedzi są kodowane 1 – 4, przy czym im wyższa

wartość tej zmiennej, tym relacje są lepsze. Rzetelność pomiaru jest zadowalająca (α Cronbacha = 0,79).

Skala Poczucie Przynależności do Szkoły – w roku 2012 zawierała dziewięć pozycji. W analizie uwzględniono sześć pierwszych (np. „czuję się obco w szkole"), które występowały również w 2003 roku. Sposób udzielania odpowiedzi oraz ich kodowanie są takie, jak w zmiennej *relacje uczeń-nauczyciel*. Im wyższa jest wartość wskaźnika, tym *poczucie przynależności do szkoły* jest większe (α Cronbacha = 0,77).

Postawa wobec szkoły badana jest skalą składającą się z czterech pozycji (np. „szkoła to strata czasu"). Sposób udzielania odpowiedzi oraz ich kodowanie są takie, jak w zmiennej *relacje uczeń-nauczyciel*. Im wyższy wynik, tym gorsza *postawa wobec szkoły* (α Cronbacha = 0,71).

Uczeń

W badaniach PISA kontrolowane są również pozapoznawcze osiągnięcia uczniów, wśród których uwzględniono *motywację do uczenia się matematyki*, *lęk matematyczny* oraz *samoocenę zdolności matematycznych*.

Motywacja do uczenia się matematyki badana jest ośmiopozycyjną skalą. Występują w niej dwie podskale (po cztery pozycje każda), z których jedna dotyczy *motywacji wewnętrznej* (np. „nie mogę doczekać się lekcji matematyki"), druga – *motywacji zewnętrznej* (np. „będę uczył się wielu rzeczy z matematyki, które pomogą mi dostać pracę"). Odpowiedzi udzielano na czterostopniowej skali szacunkowej. Zakres możliwych wartości dla każdej z podskal mieści się w przedziale od 4 do 16 punktów. Zgodność wewnętrzna jest wysoka: α Cronbacha = 0,88 dla motywacji wewnętrznej i 0,90 dla motywacji zewnętrznej.

Lęk matematyczny – wskaźnikami empirycznymi lęku jest suma punktów uzyskanych za pięć, z dziesięciu pozycji dotyczących uczenia się matematyki (np. „jestem bardzo podenerwowany rozwiązując problemy matematyczne"). Odpowiedzi udzielano na czterostopniowej skali szacunkowej, stąd wyniki mogą się zawierać między 5 a 20. Pomiar lęku jest wysoce rzetelny (α Cronbacha = 0,87).

Samoocena zdolności matematycznych mierzona jest sumą punktów uzyskanych w pozostałych pięciu pozycjach (np. „uczę się matematyki szybko"). Jej poziom może przybierać wartości z przedziału od 5 do 20. Rzetelność pomiaru jest wysoka (α Cronbacha = 0,91).

Procedura

W analizach, oprócz wyników uczniów z matematyki, uwzględniono te zmienne kwestionariuszowe, które występowały zarówno w badaniach w roku 2003, jak i 2012, i te które, zgodnie z literaturą przedmiotu, powinny być istotnymi predyktorami osiągnięć matematycznych. Wszystkie dane zaczerpnięto z baz udostępnianych przez OECD PISA.

Analiza danych

Zgodnie z przyjętym celem badań dane poddano analizie korelacyjnej oraz zastosowano modelowanie równań strukturalnych (SEM). Analizę korelacyjną przeprowadzono w pakiecie SPSS 21.0. Relacje przyczynowo-skutkowe między zmiennymi objaśniającymi i umiejętnościami matematycznymi uczniów weryfikowano przy użyciu SEM w pakiecie AMOS 21. SEM umożliwia ocenę dopasowania modelu jednocześnie do kilku podgrup.

W celu ustalenia czy związki między zmiennymi uwzględnionymi w modelu badawczym kształtują się inaczej w roku 2003 i w 2012 zastosowano modelowanie równań strukturalnych. W przeprowadzanych analizach dwie podgrupy, do których dopasowywano model wyznaczone zostały przez dane z dwóch kolejnych badań PISA, w których matematyka była dziedziną wiodącą (2003 i 2012). Założono, że struktura modelu jest taka sama w podgrupach, a różnica dotyczy jedynie wartości współczynników ścieżkowych. Przed przystąpieniem do oszacowania poprawności modelu sprawdzono założenia.

Wyniki

Szkoła

W danych PISA nie znalazło potwierdzenia uzyskiwanie wyższych osiągnięć z matematyki przez uczniów uczących się w małolicznych klasach. W roku 2012 blisko połowa uczniów (43%) uczęszczała od klas 21–25 osobowych, w których średni wynik testu z matematyki jest równy 507 (SE = 3,5). Wyższe wyniki osiągnęli uczniowie z klas liczniejszych (26–30; M = 529; SE = 5,9 oraz 31–35; M = 618; SE = 25,6). Dopiero w klasach o liczbie uczniów większej od 35 dają się odnotować niższe osiągnięcia matematyczne (M = 486; SE = 39). Jednak

w klasach takich uczy się zaledwie 0,7% badanych. Na uwagę zasługuje duży błąd standardowy średniego wyniku z matematyki w klasach o dużej liczbie uczniów (powyżej 30 osób). Świadczy to o znaczących różnicach w osiągnięciach z matematyki uczniów uczęszczających do tych klas.

Uczniowie uczestniczący w badaniach PISA 2003 i 2012 oceniali relacje łączące ich z nauczycielami w szkole (tab. 1). W opinii uczniów w 2012 roku nauczyciele mają lepsze „relacje z uczniami", są bardziej „zainteresowani samopoczuciem uczniów" oraz częściej „świadczą pomoc potrzebującym jej osobom". Nieco gorzej niż uczniowie w 2003 roku ocenili oni tylko „sprawiedliwe traktowanie uczniów".

Tabela 1. Relacje uczeń-nauczyciel w ocenie ucznia

Stwierdzenia	Rok 2003 (%)	Rok 2012 (%)
1. Dobre relacje uczeń-nauczyciel	67,1	74,4
2. Nauczyciele zainteresowani samopoczuciem uczniów	47,7	54,4
3. Nauczyciele słuchający co uczeń mam do powiedzenia	62,1	62,7
4. Nauczyciele świadczący pomoc potrzebującym uczniom	68,0	75,5
5. Nauczyciele traktujący sprawiedliwie ucznia	68,9	66,2

Wskaźnikami *poczucia przynależności do szkoły* jest postrzeganie przez uczniów relacji rówieśniczych oraz ocena własnego funkcjonowania w środowisku rówieśniczym (tab. 2). Wszelkie zmiany, które zaszły w 2012 roku względem 2003 okazują się niekorzystne. Większa część uczniów „czuje się obco i niezręcznie w szkole", ma problemy z „nawiązaniem znajomości" i „doświadcza uczucia osamotnienia". Prawie o 8% zmalała liczba uczniów przekonanych, że „w szkole są osobami lubianymi".

Tabela 2. Poczucie przynależności do szkoły

Stwierdzenia	Rok 2003 (%)	Rok 2012 (%)
1. Czuję się obco w szkole	8,2	10,1
2. Łatwo nawiązuję znajomości w szkole	88,1	86,8
3. Czuję, że jestem częścią szkoły	76,7	76,2
4. Czuję się niezręcznie i czuję, że nie pasuję do mojej szkoły	9,9	10,8
5. Wydaje mi się, że inni uczniowie mnie lubią	92,7	84,9
6. Czuję się w szkole osamotniony	8,4	8,7

Postawa uczniów wobec szkoły uległa też niekorzystnym zmianom (tab. 3). W 2012 roku prawie podwoił się odsetek uczniów, którzy w 2003 roku byli zdania, że „szkoła to strata czasu". Wzrosła też liczba uczniów uważających, że „szkoła nie przygotowuje ich do dorosłego życia". Zmalało natomiast przekonania, że w szkole „uczą się rzeczy praktycznych i ćwiczą procesy decyzyjne".

Tabela 3. Postawa wobec szkoły

Stwierdzenia	Rok 2003 (%)	Rok 2012 (%)
1. Szkoła nie przygotowuje do dorosłego życia	31,7	38,4
2. Szkoła to strata czasu	11,0	21,2
3. Szkoła pomaga zdobyć pewność siebie przy podejmowaniu decyzji	76,4	70,9
4. Szkoła uczy rzeczy przydatnych w pracy zawodowej	80,3	70,7

Środowisko rodzinne

W roku 2012 zmniejszył się odsetek uczniów zamieszkujących z obojgiem rodziców (o 6,5%). O 4,1% matek i 6% ojców mniej niż w 2003 roku mieszka na co dzień z dziećmi. Dane dotyczące zamieszkiwania rodziców z dziećmi zawiera tabela 4.

Tabela 4. Z kim na co dzień mieszkają badani uczniowie

Rok 2003 (N = 4377)	ojciec (%)		ogół	Rok 2012 (N = 4545)	ojciec (%)		ogół
matka (%)	nie	tak		matka (%)	nie	tak	
nie	1,1	1,4	2,5	nie	4,6	2,0	6,6
tak	11,8	85,8	97,5	tak	14,1	79,3	93,4
ogół	12,8	87,2	100,0	ogół	18,8	81,2	100,0

Odsetek rodziców posiadających wykształcenie wyższe (co najmniej licencjackie) zestawiono w tabeli 5. W roku 2012 odsetek rodziców legitymujących się wyższym wykształceniem wzrósł w porównaniu z 2003 rokiem, przy czym wśród matek o 6,8%, a wśród ojców o 2,6%.

Tabela 5. Rodzice z wykształceniem wyższym

Rok badania	Wykształcenie wyższe	
	matka (%)	ojciec (%)
2003 (N = 4383)	16,5	14,4
2012 (N = 4607)	23,3	17,0

Znacznej poprawie uległa sytuacja zawodowa rodziców (tab. 6). Zarówno wśród matek, jak i wśród ojców zwiększyła się o ok. 13% liczba osób zatrudnionych na pełnym etacie. W obu grupach zmalała też liczba bezrobotnych poszukujących pracy (o ponad 8%).

Tabela 6. Sytuacja zawodowa rodziców

Rodzice	Rok badania	Pełny etat	Część etatu	Bezrobotna/y	Inne
Matka (%)	2003 (N = 4329)	50,5	7,8	16,2	25,5
	2012 (N = 4497)	63,1	7,5	7,8	21,6
Ojciec (%)	2003 (N = 4179)	65,1	7,4	12,5	15,1
	2012 (N = 4300)	78,9	8,7	4,4	8,0

Kapitał kulturowy mierzony liczbą książek w domowej bibliotece zmalał nieznacznie. W domach dominowały biblioteki zawierające od 26 do 100 książek i występowały one ze zbliżoną częstością (0,34; 0,33). W roku 2003 średnia liczba książek to 167, a w 2012 – 156. Zmiany te mogą mieć związek ze znacznym upowszechnianiem się i-booków.

Warunki domowe uczniów uległy poprawie. Wzrosła liczba uczniów posiadających własny pokój (74%; 83%) i biurko do nauki (91%; 98%). Zwiększył się ich dostęp do słowników i książek pomocnych w nauce (92%; 96%). Największe zmiany ujawniły się w dostępie do komputera (60%; 97%), Internetu (34%; 95%) i oprogramowania edukacyjnego (47%; 71%). Nieznacznie zmalała liczba osób posiadających literaturę klasyczną (82%; 79%) i poezję (52%; 49%).

W tabeli 7 zamieszczono statystyki opisowe dla zmiennych ilościowych uwzględnionych w analizach. Ponadto dokonano porównania wartości tych zmiennych w roku 2003 i 2012.

Wartości współczynników asymetrii [–0,99 : 0,42] i kurtozy [–0,68 : 1,24] oraz liczebności prób uzasadniają zastosowanie testów parametrycznych dla wszystkich zmiennych oprócz *domowych zasobów edukacyjnych* (As = –2,5; K = 9,3 w 2012).

Tabela 7. Statystyki opisowe zmiennych, test istotności różnic średnich i wielkość efektu

Zmienne	Rok badań	N	M	S	A_S	K	p	d Cohena
umiejętności matematyczne	2003	4383	489,0	86,4	0,02	–0,17	0,001	0,36
	2012	4607	520,5	88,4	0,12	–0,37		
motywacja wewnętrzna	2003	4294	9,3	2,8	0,14	–0,17	0,001	0,25
	2012	3035	8,6	2,9	0,25	–0,37		
motywacja zewnętrzna	2003	4315	11,7	2,3	–0,66	1,07	0,001	0,19
	2012	4049	11,2	2,9	–0,51	0,19		
lęk	2003	4299	11,9	3,1	–0,09	0,09	0,347	0,03
	2012	3000	11,8	3,6	0,13	–0,33		
samoocena	2003	4279	12,1	3,4	0,24	–0,29	0,04	0,03
	2012	2987	12,0	3,9	0,22	0,68		
domowe zasoby edukacyjne	2003	4380	5,1	1,6	–0,51	–0,49	0,001*	
	2012	4525	6,5	0,8	–2,50	9,34		
dobytek	2003	4380	7,6	2,4	–0,51	–0,28	0,001	0,74
	2012	4571	9,1	1,6	–0,99	1,24		
relacje uczeń-naucyciel	2003	4365	13,3	2,8	0,24	1,23	0,001	0,14
	2012	3036	13,7	2,9	0,22	0,59		
przynależność do szkoły	2003	4201	18,9	2,7	–0,53	1,04	0,113	0,06
	2012	2991	18,8	3,0	–0,60	1,10		
postawa wobec szkoły	2003	4205	8,0	2,2	0,41	0,32	0,001	0,40
	2012	3013	8,9	2,3	0,42	0,41		

d Cohena – wielkość efektu, * – test U Manna-Whitneya

Z analizy wynika, że znacznemu wzrostowi osiągnięć matematycznych nie towarzyszą pozytywne zmiany ani w uczniu, ani w jego obrazie szkoły. W roku 2012 obniżyła się istotnie *motywacja uczniów*. *Samoocena umiejętności matematycznych* oraz *poziom lęku matematycznego* (d Cohena = 0,03) pozostają na niezmienionym poziomie. Nie zmieniło się też *poczucie przynależności do szkoły*. Uległa pogorszeniu *postawa uczniów wobec szkoły* (im wyższa średnia tym gorszy obraz). Poprawiły się jedynie *relacje uczniów z nauczycielami*. Największe, korzystne zmiany zaszły w *domowych zasobach edukacyjnych* oraz w *dobytku domowym* (tab. 7 d Cohena).

W celu weryfikacji przyjętego modelu badawczego (ryc. 1) zastosowano modelowanie równań strukturalnych (SEM). Dopasowanie modelu do danych PISA 2003 i 2012 szacowano stosując statystykę χ^2, RMSEA oraz indeksy dopasowania CFI, IFI, TLI, NFI. Do estymacji parametrów modeli zastosowano metodę największej wiarygodności (ML *maximum likelihood*).

Testowaniu poddano trzy modele. W każdym z nich zmiennymi egzogenicznymi były zmienne rodziny i szkoły, a zmiennymi endogenicznymi – osiągnięcia pozapoznawcze i *umiejętności matematyczne* ucznia. Testy dopasowania dla przyjętego modelu (por. ryc. 1) dały następujące wartości: $\chi^2 = 12664$ df = 616 $p < 0{,}001$; CFI = 0,88; IFI = 0,88; NFI = 0,91; TLI = 0,86; RMSEA = 0,047 (90% CI: 0,046, 0,047). Pierwszy krok doprowadził do usunięcia z modelu zmiennej *domowe zasoby edukacyjne*, która jest częścią składową *dobytku*. W efekcie uzyskano następujące wartości wskaźników dopasowania: $\chi^2 = 8620{,}6$ df = 520 $p < 0{,}001$; CFI = 0,91; IFI = 0,91; NFI = 0,91; TLI = 0,89; RMSEA = 0,042 (90% CI: 0,041, 0,042). W drugim kroku wyeliminowana została zmienna *relacje uczeń-nauczyciel*. W efekcie zmienna *dom* posiada cztery wskaźniki (*kapitał kulturowy*, *dobytek*, *wykształcenie ojca* i *wykształcenie matki*), a *szkoła* – dwa (*postrzeganie szkoły* i *postawa wobec szkoły*). Testy dopasowania dla przyjętego modelu dały następujące wartości: $\chi^2 = 8632$ df = 566 $p < 0{,}001$; CFI = 0,92; IFI = 0,92; NFI = 0,91; TLI = 0,90; RMSEA = 0,040 (90% CI: 0,039, 0,041). Statystyka χ^2 jest istotna statystycznie. Uwzględniając jednak duże liczebności porównywanych grup ($N_1 = 4383$; i $N_2 = 4607$) nie jest to dostatecznym powodem odrzucenia modelu. Wartości pozostałych indeksów dopasowania pozwalają uznać model za dobrze dopasowany do danych (Kline 2011).

W tabeli 8 przedstawiono współczynniki ścieżkowe testowanego modelu uzyskane odpowiednio dla danych w badaniach PISA w 2003 oraz w 2012 roku. Zastosowanie modelowania strukturalnego w podgrupach pozwala ocenić różnice w zależnościach między zmiennymi zawartymi w modelu. Dlatego też oprócz współczynników ścieżkowych wyznaczono wartości testu istotności różnic między współczynnikami ścieżkowymi w latach 2003 i 2012 (CR *Critical ratios for differences*). Istotność statystyki CR oznacza, że współczynnik ścieżkowy w latach 2003 i 2012 uległ znaczącym zmianom.

Szkoła w znaczącym stopniu wpływa na *lęk matematyczny, motywację zewnętrzną* i *wewnętrzną* (por. tabela 8), przy czym im większe poczucie przynależności do szkoły i lepsza postawa wobec szkoły tym niższy lęk i wyższa motywacja zewnętrzna. Jednak bezpośredni wpływ szkoły na osiągnięcia matematyczne uczniów jest nieistotny. Dotyczy to współczynników ścieżkowych wyznaczonych zarówno w badaniach w 2003, jak i w 2012 roku. Badania

Tabela 8. Współczynniki ścieżkowe testowanego modelu oraz CR dla różnic między 2003 i 2012 rokiem

Ścieżka	Rok badań 2003		Rok badań 2012		CR dla różnic	p dla różnic
	estymowane współczynniki ścieżek	SE	estymowane współczynniki ścieżek	SE		
lęk←szk	–1,53***	0,29	–0,82***	0,11	2,310	*
mot_z←dom	0,03ni	0,02	0,23***	0,04	4,532	***
lek←dom	0,16***	0,02	–0,32***	0,03	–4,363	***
mot_z←szk	1,52***	0,31	1,15***	0,16	–1,072	ni.
mot_w←mot_z	0,07ni	0,11	0,29ni	0,35	–0,982	ni.
smooc←dom	–0,01ni	0,01	0,06***	0,02	3,949	***
samooc←lęk	–0,82***	0,02	–0,80***	0,02	0,836	ni.
mot_w←szk	2,33***	0,54	1,67**	0,52	–0,869	ni.
mot_w←dom	0,06ni	0,02	0,28***	0,10	2,181	*
szk↔dom	–0,07*	0,03	–0,04*	0,02	0,661	ni.
mat←lęk	–77,75***	9,22	–56,76***	5,92	1,915	ni.
mat←mot_z	41,66**	16,26	146,70ni	179,20	0,584	ni.
mat←mot_w	138,34ni	136,61	200,81ni	288,41	0,195	ni.
mat←szk	–551,74ni	408,73	–491,55ni	617,46	0,081	ni.
mat←samooc	37,66***	4,38	38,06***	3,66	0,071	ni.
mat←dom	2,40ni	–1,53***	–41,76ni	104,80	–0,419	ni.

gdzie: ni – nie istotne; * – $p < 0,05$; ** – $p < 0,01$; *** – $p < 0,001$; CR – Critical Ratios

w 2003 roku ujawniły znaczący wpływ *domu* tylko na poziom *lęku matematycznego*. W roku 2012 *dom* wywierał ponadto istotny wpływ na *motywację zewnętrzną* i *wewnętrzną* oraz *samoocenę umiejętności matematycznych*. Im lepiej wykształceni rodzice, większy kapitał kulturowy i dostatek, tym wyższy poziom motywacji, samooceny i niższy poziom lęku mają ich dzieci. Podobnie, jak w przypadku szkoły, bez względu na rok prowadzenia badań, nie potwierdził się bezpośredni wpływ domu na *osiągnięcia matematyczne* uczniów. Bezpośredni wpływ na *osiągnięcia matematyczne* ma *lęk matematyczny* i *samoocena umiejętności matematycznych,* a w badaniach w 2003 roku dodatkowo *motywacja zewnętrzna*.

Porównanie współczynników ścieżkowych modelu w 2003 i 2012 roku wskazuje na istotne zmiany w mocy predykcyjnej domu (tab. 8). Znacząco wzrósł

jego wpływ na *motywację zewnętrzną* i *wewnętrzną*, *lęk matematyczny* oraz *samoocenę umiejętności matematycznych*. Zmniejszył się też wpływ szkoły na *lęk matematyczny* uczniów. Pozostałe związki między zmiennymi nie różnią się istotnie.

W tabeli 9 zestawiono wartości efektów całkowitych (bezpośrednich i pośrednich) dla wszystkich zmiennych endogenicznych uwzględnionych w modelu.

Tabela 9. Standaryzowane efekty całkowite

Zmienne endogeniczne	Rok badań	Zmienne objaśniające					
		dom	szkoła	lęk	samooc	mot_z	mot_w
lęk	2003	–0,42	–0,62	–	–	–	–
	2012	–0,34	–0,59	–	–	–	–
samooc	2003	0,33	0,52	–0,84	–	–	–
	2012	0,33	0,47	–0,80	–	–	–
mot_z	2003	0,08	0,68	–	–	–	–
	2012	0,25	0,87	–	–	–	–
mot_w	2003	0,16	0,96	–	–	0,06	–
	2012	0,22	0,94	–	–	0,27	–
matematyka	2003	0,52	0,04	–0,74	0,25	0,31	0,96
	2012	0,55	0,10	–0,69	0,30	0,67	1,64

gdzie „–" oznacza nie występowanie ścieżki w modelu

W przyjętym modelu najważniejszymi predyktorami osiągnięć matematycznych uczniów, bez względu na rok prowadzenia badań, jest *motywacja wewnętrzna* i *lęk matematyczny* (tab. 9). Wyższy poziom motywacji wewnętrznej wpływa na wyższe *osiągnięcia matematyczne*. W przypadku lęku sytuacja jest odwrotna. Wyższemu poziomowi *lęku* towarzyszą niższe *osiągnięcia matematyczne*. W roku 2012 wzrósł całkowity wpływ *motywacji zewnętrznej* na osiągnięcia matematyczne. Porównanie oddziaływań domu i szkoły na wyniki testu z matematyki wskazuje na wyraźnie wyższy wpływ domu. Należy jednak podkreślić, że wpływ domu i szkoły obejmuje ograniczoną grupę zmiennych kontrolowanych.

Na uwagę zasługuje też wpływ szkoły na pozapoznawcze osiągnięcia uczniów (tab. 9). W każdym przypadku jest on wyraźnie wyższy niż wpływ środowiska rodzinnego uczniów.

Wyznaczone wartości R^2 pokazują, że przyjęty model wyjaśnia 72% wariancji *osiągnięć matematycznych* uczniów w roku 2003 i 95% wariancji w roku 2012.

Podsumowanie

Badania PISA diagnozują osiągnięcia uczniów 15-letnich oraz liczne zmienne związane z procesem kształcenia. W latach 2003 i 2012 wśród polskich 15-latków odnotowano wyraźną poprawę wyników uzyskanych w testach umiejętności matematycznych. W związku z tym podjęto próbę ustalenia, czy zmianom tym towarzyszą zmiany w kontrolowanych w PISA uwarunkowaniach procesu kształcenia i w jakim kierunku one zachodzą.

W opinii uczniów w środowisku szkolnym zaszły niekorzystne zmiany. Poprawiły się nieznacznie *relacje uczniów z nauczycielami*, ale pogorszyła *postawa uczniów wobec szkoły*. Zmalało też, choć w niewielkim stopniu, ich *poczucie przynależności do szkoły*. W badaniach w 2003 i 2012 roku *lęk matematyczny* i *samoocena umiejętności matematycznych* pozostają na niezmienionym poziomie. Obniżyła się *motywacja zewnętrzna* i *wewnętrzna uczniów*, ale zmiany te są małe (tab. 7). Największe, korzystne zmiany zaszły w domach rodzinnych uczestników badań. Wzrósł odsetek rodziców posiadających wyższe wykształcenie, a zmalało bezrobocie. Poprawił się *dobytek rodzin*. Uczniowie mają dużo lepszy dostęp do komputerów, Internetu oraz programów edukacyjnych.

Przeprowadzone analizy pokazują, że na umiejętności osiągnięcia matematyczne uczniów największy wpływ mają niezmiennie ich *wewnętrzna motywacja do uczenia się matematyki*, a następnie poziom *lęku matematycznego* oraz *samoocena umiejętności matematycznych*. Każda z tych zmiennych pozostaje pod większym wpływem szkoły niż środowiska rodzinnego (tab. 9).

Należy jednak wyraźnie podkreślić, że związki przyczynowo-skutkowe zostały przyjęte na podstawie przesłanek teoretycznych, a zastosowanie SEM pozwala jedynie stwierdzić, że nie pozostają one w sprzeczności z analizowanymi danymi, co nie jest równoważne potwierdzeniu tych związków.

Diagnozy z bliska i diagnozy z daleka pozwalają spojrzeć na osiągnięcia uczniów z różnych perspektyw: klasy, szkoły, regionów, Polski czy usytuowania na tle innych krajów. Przydatność obu perspektyw w znacznej mierze zależy od celu, do realizacji którego sięga się po dostarczane przez nie informacje. Są one źródłem komplementarnej wiedzy na temat efektów procesu kształcenia i obie są niezbędne.

Bibliografia

Bradshaw J. 2002. *Child poverty and child outcomes*, „Children & Society", 16, s. 131–140.

Branigan A.R., McCallum K.J., Freese J. 2013. *Variation in the heritability of educational attainment: An international meta-analysis*, Northwestern University Institute for Policy Research, Evanston.

Ciżkowicz B. 2013a. *Wybrane uwarunkowania osiągnięć polskich i fińskich 15-latków w badaniach PISA 2009*, [w:] *Polska edukacja w świetle diagnoz prowadzonych z różnych perspektyw badawczych*, red. B. Niemierko, M.K. Szmigel, Grupa Tomami, Kraków, s. 73–83.

Ciżkowicz B. 2013b. *Obraz polskiej szkoły w badaniach 15-latków*. „Teraźniejszość – Człowiek – Edukacja", nr 4, s. 117–130.

Ciżkowicz B. 2014. *O potrzebie ewaluacji efektów kształcenia w szkołach wyższych: stan, problemy, perspektywy*, „Ruch Pedagogiczny", nr 2, s. 123–134.

Connolly P. 2006. *The effects of social class and ethnicity on gender differences in GCSE attainment: A secondary analysis of the Youth Cohort Study of England and Wales 1997–2001*, „British Educational Research Journal", 32(1), s. 3–21.

Deci E., Ryan R. 1985. *Intrinsic motivation and self-determination in human behavior*, Plenum Press, New York.

Eisenberger R., Selbst M. 1994. *Dose reward increase or decrease creativity?*, „Journal of Personality & Social Psychology", nr 6, s. 1116–1127.

Eisenberger R., Cameron J. 1996. *Detrimental effects of reward*, „American Psychologist", nr 11, s. 1153–1166.

Eisenberger R., Rhoades L., Cameron J. 1999. *Does Pay for Performance Increase or Decrease Perceived Self-Determination and Intrinsic Motivation?*, „Journal of Personality & Social Psychology", nr 5, s. 1026–1040.

Feinstein L., Duckworth K., Sabates R. 2008, *Education and the family: Passing success across the generations*, Routledge, London.

Fleming Ch., Catalano R., Haggerty K. 2010. *Relationships between level and change in family, school, and peer factors during two periods of adolescence and problem behavior at age 19*, „Jurnal of Youth and Adolescence", nr 39, s. 670–682.

Gil-Flores J., Padilla-Carmona M., Suarez-Ortega M. 2011. *Influence of gender, educational attainment and family environment on the educational aspirations of secondary school students*, „Educational Review", 63, 3, s. 345–363.

Greenwald R., Hedges L., Laine R. 1996. *The effect of school resources on achievement*, „Review of Educational Research", nr 66 (3), s. 361–396.

Hembree R. 1990. *The nature, effects, and relief of mathematics anxiety*, „Journal for Research in Mathematics Education", 21, s. 33–46.

Izdebska H. 1993. *Rodzina i jej funkcja wychowawcza*, [w:] *Encyklopedia Pedagogiki*, red. W. Pomykało, Fundacja Innowacja, Warszawa.

Jakubowski M. 2013. *Polska w świetle międzynarodowych badań umiejętności uczniów*, [w:] *Polska edukacja w świetle diagnoz prowadzonych z różnych perspektyw badawczych*, red. B. Niemierko, M.K. Szmigel, Grupa Tomami, Kraków, s. 99–100.

Kline R. 2011. *Principles and practice of structural equation modeling*, The Guilford Press, New York.

Krinzinger H., Kaufmann L., Willmes K. 2009. *Math anxiety and math ability in early primary school years*, „Journal of Psychoeducational Assessment", nr 27, s. 206–225.

Kyunghee L. 2011. *Impacts of the duration of head start enrollment on children's academic outcomes: Moderation effects of family risk factors and earlier outcomes*, „Journal of Community Psychology", nr 39, s. 698–716.

Lee V., Simith J., Croninger R. 1997. *How high school organization influences the equitable distribution of learning in mathematics and science*, „Sociology of Education", nr 70, s. 128–150.

Liu Y, Wu A., Zumbo B. 2006. *The relation between outside of school factors and mathematics achievement: A cross-country study among the U.S. and five top-performing asian countries*, „Journal of Educational Research & Policy Studies", nr 6, s. 1–35.

MEN 2012. *Program Międzynarodowej Oceny Umiejętności Uczniów OECD PISA. Wyniki badania 2012 w Polsce*, dostępny na: http://www.ibe.edu.pl/images/prasa/PISA-2012-raport_krajowy.pdf (otwarty 10 stycznia 2015).

Niemierko B. 2007. *Kształcenie szkolne. Podręcznik skutecznej dydaktyki*, Wydawnictwo Akademickie i Profesjonalne, Warszawa.

Niemierko B. 2013. *Diagnoza edukacyjna duża i mała*, [w:] *Polska edukacja w świetle diagnoz prowadzonych z różnych perspektyw badawczych*, red. B. Niemierko, M.K. Szmigel, Grupa Tomami, Kraków, s. 23–39.

Niemierko B. 2014. *Dwie dekady dojrzewania diagnostyki edukacyjnej w Polsce*, „Ruch Pedagogiczny", nr 2, s. 5–26.

OECD 2003. Dostępny na: http://pisa2003.acer.edu.au/downloads.php (otwarty 10 stycznia 2015).

OECD 2012a. *Raport krajowy. Wyniki badania PISA 2012 w Polsce*, dostępny na: http://www.ibe.edu.pl/images/prasa/PISA-2012-raport_krajowy.pdf (otwarty 10 stycznia 2015).

OECD 2012b. Dostępny na: http://pisa2012.acer.edu.au/downloads.php (otwarty 10 stycznia 2015).

Okoń W. 2007. *Nowy słownik pedagogiczny*, Wydawnictwo Akademickie Żak, Warszawa.

PISA 2003a. Dostępny na: http://pisa2003.acer.edu.au/interactive.php (otwarty 10 stycznia 2015).

PISA 2003b. Dostępny na: http://pisa2003.acer.edu.au/downloads/StQ_Questionnaire2003.pdf (otwarty 5 lutego 2015).

PISA 2012a. Dostępny na: http://pisa2012.acer.edu.au/ (otwarty 10 stycznia 2015).

PISA 2012b. Dostępny na: http://pisa2012.acer.edu.au/interactive_results.php (otwarty 10 stycznia 2015).

Richardson F.C., Woolfolk R.L. 1980. *Mathematics anxiety*, [w:] *Test anxiety: Theory, research and application*, red. I.G. Sarason, Erlbaum, Hillsdale, NJ, s. 271–288.

Shonkoff J., Phillips D. 2000. *Family resources*, [w:] *From neurons to neighborhoods*, red. J. Shonkoff, D. Phillips, National Academy Press, Washington.

Smith J.R., Brooks-Gunn J., Klebanov P.K. 1997. *Consequences of living in poverty for young children's cognitive and verbal ability and early school achievement*, [w:] *Consequences of growing up poor*, red. G.J. Duncan, J. Brooks-Gunn, Russell Sage Foundation, New York.

Strelau J. 2000. *Psychologia. Podręcznik akademicki*, t. 3, Gdańskie Wydawnictwo Psychologiczne, Sopot.

Strelitz J., Lister R. 2008. *Why money matters: family income, poverty and children's live*, Save the Children, London.

Szyling G. 2013. *Zmiana skali oceniania jako element zmiany myślenia o jakości osiągnięć uczniów*, [w:] *Polska edukacja w świetle diagnoz prowadzonych z różnych perspektyw badawczych*, red. B. Niemierko, M.K. Szmigel, Grupa Tomami, Kraków, s. 237–238.

Wang J., Lin E. 2005. *Comparative studies on USA and Chinese mathematics learning and the implications for standards-based mathematics teaching reform*, „Educational Researcher", nr 34 (5), s. 3–13.

Źródło: Eurostat, UOE (otwarty 10 stycznia 2015).

KWARTALNIK PEDAGOGICZNY
2015 NUMER 4(238)

Zofia Lisiecka
Centralna Komisja Egzaminacyjna w Warszawie

Model eksternistyczny jako przykład *uczenia się przez całe życie*

Summary

EXTRAMURAL GENERAL EDUCATION EXAMS AS AN EXAMPLE OF *LIFELONG LEARNING*

For many years the national extramural exams were placed on the margins of the Polish educational system. In 2007 the Central Examination Board was entrust with the assessment of skills of adult applicants who wanted to have their comprehensive knowledge officially confirmed on the primary, lower secondary or upper secondary school level. Since 2012 the applicants have also been able to take the extramural exams in general and practical knowledge on the level of a basic vocational school. This resolution is consistent with the guidelines of the European Parliament from the year 2008 about European Qualification Framework for Lifelong Learning and, accepted in 2012, the OECD strategy: Better Skills, Better Jobs, Better Lives.

In the Polish educational system, the applicants for extramural exams are adults who do not attend any school but learn by themselves and decide individually when they are ready to take the exams. Every year a few thousand of mostly young people make this decision and sit the exams, most often on the upper secondary school level.

The aim of this article is an attempt to create a collective portrait of all these people who left the regular school system, for different reasons, and try to obtain their education.

Since 2008 the author of the article has been doing extensive research on this population, not only concentrating on the statistics index but also carrying out a detailed qualitative study.

Key words: national extramural exams, lifelong learning, validation of non-formal and informal learning, self-directed learning.

red. Paulina Marchlik

Egzaminy eksternistyczne – wymuszona transformacja

Egzaminy eksternistyczne są ostatnim elementem, który w roku 2007 dopełnił budowany w Polsce od 1999 roku zewnętrzny system egzaminacyjny. Katalizatorem dla nieplanowanego[1] wcześniej poszerzenia obszaru objętego egzaminami zewnętrznymi stał się raport Najwyższej Izby Kontroli (*Informacja o wynikach kontroli*... 2006) o przeprowadzaniu w szkołach oraz centrach kształcenia ustawicznego organizowanych przez kuratoria oświaty egzaminów eksternistycznych. Z uwagi na liczbę osób pragnących zdobyć kwalifikacje zawodowe bądź uzupełnić wykształcenie ogólne, ujawnione w nim nieprawidłowości, niektóre o charakterze korupcyjnym, a nawet przestępczym, stanowiły wówczas problem o dużym znaczeniu społecznym i gospodarczym. W analizowanym przez NIK okresie (lata 2002–2004) do egzaminów eksternistycznych przystąpiło bowiem łącznie ponad 300 tys. osób. W tym samym czasie powszechna stała się opinia, iż zdanie egzaminu w trybie eksternistycznym w większym stopniu zależy od przychylności szkolnej komisji niż od umiejętności zdającego. Może o tym świadczyć zarówno odsetek osób, które na egzaminach uzyskiwały oceny pozytywne (najwyższy wskaźnik procentowy, tj. 98,1% odnotowano w szkołach policealnych, najniższy w szkołach podstawowych – 75%[2]), jak i nagminna praktyka organizowania w placówkach, które egzaminy przeprowadzały, kursów przygotowujących do tych egzaminów, z zajęciami realizowanymi przez późniejszych egzaminatorów. Tym samym eksternistyczny sposób zdawania egzaminów w owym czasie z jednej strony stał się praktyką masową, bo łatwo dostępną, z drugiej zaś deprecjonował osiągnięte przez zdających rezultaty, w opinii społecznej skażone nagannymi, a nawet nielegalnymi praktykami egzaminatorów i egzaminowanych, takimi jak kupowanie przychylności komisji, ujawnianie pytań egzaminacyjnych, stwarzanie pozorowanej sytuacji egzaminacyjnej, ściąganie, itp.

Wprowadzone w 2007 roku unormowania prawne nakazały badanie osiągnięć osób pragnących potwierdzić wykształcenie ogólne w zakresie szkoły podstawowej, gimnazjum oraz liceum ogólnokształcącego dla dorosłych przez system egzaminowania zewnętrznego (Dz.U. nr 179, poz. 1273). Od roku 2012 eksterni mogą także przystępować do identycznych w całym kraju egzaminów z zakresu kształcenia ogólnego na poziomie zasadniczej szkoły zawodowej, a także potwier-

[1] Por. dokumenty programowe związane z reformą edukacji w 1999 roku. Brak w nich jakichkolwiek wzmianek o zewnętrznych egzaminach eksternistycznych.

[2] Dane MEN dotyczące lat szkolnych 2002/2003 i 2003/2004. Cyt. za: *Informacja o wynikach kontroli*... 2006: 11.

dzać kwalifikacje zawodowe, zdobyte nie tylko w wyniku kształcenia szkolnego czy ustawicznego, ale też nieformalnie (Dz.U. 2012 poz. 188). Rozwiązanie to jest zgodne ze strategicznymi celami Unii Europejskiej, wyznaczanymi przez jej najważniejsze instytucje – Parlament Europejski, Radę Unii Europejskiej i Komisję Europejską, będące w ciągu ostatniej dekady inicjatorami opracowania licznych strategii rozwoju społecznego, w tym edukacji, takich jak m.in. proces boloński, proces kopenhaski, strategia lizbońska czy *Europa 2020*.

Strategia *uczenia się przez całe życie* w praktyce

Szerokie podejście do edukacji, które obok edukacji formalnej uwzględnia również inne formy edukacji i uczenia się, spotykane w rozwoju człowieka od najwcześniejszych lat do późnej starości, promuje przyjęty przez polski rząd w 2013 roku strategiczny dokument *Perspektywa uczenia się przez całe życie*. Dokument ten wytycza kierunki działania na rzecz: 1) uczenia się w różnych kontekstach – formalnym, pozaformalnym i nieformalnym, w różnych formach i miejscach (ang. *Lifewide learning*), przy równym traktowaniu wszystkich form uczenia się, 2) uczenia się na wszystkich etapach życia, począwszy od najmłodszych lat do późnej starości (ang. *Lifelong learning*, *LLL*), przedkładając „uczenie się" nad „nauczanie", 3) identyfikacji, oceny i potwierdzania efektów uczenia się niezależnie od tego, gdzie, jak i kiedy ono zachodziło – w odniesieniu do europejskich oraz krajowych ram kwalifikacji (*Perspektywa*... 2013: 3). Za cel wszystkich tych działań można by przyjąć strategię przyświecającą Organizacji Współpracy Gospodarczej i Rozwoju (OECD): *Better Skills, Better Jobs, Better Lives* („wyższe umiejętności, lepsza praca, lepsze życie").

Koncepcja obecnego systemu egzaminów eksternistycznych pozwala na pełną realizację powyższych strategii Unii Europejskiej i rządu polskiego. Eksterni bowiem to osoby dorosłe (powyżej 18. roku życia), które postawiły na samodzielność w wyborze drogi do celu, jakim jest poddanie własnych umiejętności procesowi walidacji, polegającemu na rozpoznawaniu, ocenianiu, określaniu wartości i uznawaniu efektów uczenia się (Narwojsz i in. 2008: 8–9). W myśl rozporządzenia MEN z 2012 r. kwestią wolnego wyboru eksternów jest sposób przygotowywania się do egzaminów. Drogami, które obierają, są: a) uczenie się formalne, czyli udział w procesie kształcenia lub szkolenia, który prowadzi do uzyskania kwalifikacji ujętych w krajowym systemie kwalifikacji (w przypadku eksterna ten sposób uczenia się jest jego doświadczeniem minionym), b) uczenie

się pozaformalne, które może prowadzić do uzyskania kwalifikacji istniejących poza krajowym systemem kwalifikacji, uznawanych dobrowolnie, bądź c) uczenie się nieformalne, postrzegane jako proces niesformalizowany, zachodzący przez całe życie, albo jako proces zamierzony (planowane, samodzielne uczenie się) i niezamierzony (uczenie się mimowolne) (Narwojsz i in. 2008: 6).

Samokształcenie sposobem wyjścia z grupy *wykluczonych*

Mimo że ekstern to z natury ktoś lokujący się poza głównym nurtem edukacyjnym (łac. *externus* – „pozostający na zewnątrz"), każda z osób przystępujących do egzaminów eksternistycznych ma za sobą okres szkolny. Dlaczego więc przerwali naukę w szkole? Dlaczego, mówiąc o tej grupie, często używa się określenia *wykluczeni*? Czy to stygmatyzujące nacechowanie w ich przypadku okazuje się trafne? Problematyka wykluczania edukacyjnego jest często uwzględniana w analizach przyczyn wykluczania społecznego, na co wskazują prace takich autorów, jak: Zbigniew Kwieciński, Maria Jarosz czy Ireneusz Białecki (zob.: Kwieciński 2002; Jarosz 2008; Białecki 2010). Ten ostatni w raporcie *Wykluczenie edukacyjne* wymienia główne problemy, które przyczyniają się do powstawania lub pogłębiania się nierówności w dostępie do kształcenia, utożsamiane przez niego z nierównościami szans życiowych, mogącymi skutkować wykluczeniem społecznym. Zdaniem badacza należą do nich m.in.: nierówna dostępność do szkół średnich, a następnie kształcenia i dokształcania dorosłych, selekcja szkolna, czy też porzucanie nauki przed zakończeniem obowiązkowego kształcenia. Tym samym nierówności edukacyjne odnoszą się do dwóch obszarów: nierówności warunków (różnice w sytuacji osób np. pochodzących z miast i ze wsi; różnice związane ze środowiskiem rodzinnym, wykształceniem rodziców, systemem wartości, sytuacją życiową) oraz nierówności wyników/rezultatów (różnice w wynikach nauki, na poziomie testów kompetencji, w proporcjach osób studiujących, najczęściej w odniesieniu do pochodzenia z miasta/ze wsi oraz z rodzin zamożnych/ubogich).

Rezultaty powyższych analiz potwierdzają przeprowadzone w 2008 roku przez Wydział Badań i Ewaluacji Centralnej Komisji Egzaminacyjnej badania pełnej populacji eksternów przystępujących w nowym wówczas trybie do zewnętrznych egzaminów z zakresu liceum ogólnokształcącego dla dorosłych (Lisiecka 2008). Miały one na celu nakreślenie zbiorowego *portretu eksterna*, w tym pokazanie głównych przyczyn wykluczenia (się) tych osób z systemu edukacji formalnej.

Jak podali sami zainteresowani, ich decyzje o przerwaniu nauki w szkole spowodowane były bądź niepowodzeniami wynikającymi z doświadczeń szkolnych, w tym przede wszystkim niezadowalającymi efektami kształcenia, skutkującymi kłopotami wychowawczymi, absencją, powtarzaniem kolejnych klas, bezradnością intelektualną (zob.: Ciżkowicz 2007, 2009), itp. a w efekcie finalnym skreśleniem z listy uczniów lub porzuceniem szkoły, bądź z młodzieńczą kontestacją szkolnej rzeczywistości i niezgodą na uczestniczenie w tym aspekcie życia społecznego, bądź też z różnorodnymi zdarzeniami związanymi z życiem osobistym respondentów (np. konieczność podjęcia pracy zarobkowej, założenie rodziny, opieka nad dzieckiem lub innym krewnym, wyjazd za granicę, w tym jako osoba małoletnia, z powodu emigracji zarobkowej rodziców, długotrwała choroba, pogłębiający się nałóg, niedostosowanie społeczne, odbywanie kary więzienia, sytuacja materialna, w tym brak środków na dalsze kształcenie, itp.), bądź też wreszcie ze świadomie obieraną drogą pozwalającą danej osobie godzić uczenie się z pracą zarobkową (Lisiecka 2008: 13).

Eksterni to jednak specyficzna grupa *wykluczonych*. Nie wykazują bowiem postaw przypisywanych tej grupie społecznej przez badaczy – często destrukcyjnych, skutkujących poczuciem niemożności, apatii, zwątpienia, powodujących *ucieczkę przegranych*: w zachowania aspołeczne bądź niezgodne z prawem, nałogi, depresję, z targnięciem się na życie włącznie (zob. Jarosz 2004). To ludzie, którzy przechodząc przez etap *wykluczenia* (*się*), w następstwie swych decyzji przezwyciężają trudności, „wychodzą z cienia", odnajdują motywację do działania. Wyniki wzmiankowanych własnych badań wskazały na różnorodność motywacji respondentów do podjęcia trudu uczenia się na nowo (Lisiecka 2008: 16–18)[3]. Większość z nich (62%) zadeklarowała, iż kierowali się motywacją wewnętrzną, wynikającą z potrzeby doskonalenia się, „stawania się mądrzejszym" lub nawet „lepszym", poznawania nowych rzeczy oraz z chęci sprawdzenia się „czy jeszcze potrafię, czy podołam, czy jestem coś wart...". Mniejszy odsetek badanych (27%) wskazywał na przewagę bodźców zewnętrznych, postrzeganych zarówno w kategorii nagrody, jak i kary. W tym przypadku ważniejsze od własnego samodoskonalenia są oczekiwania na benefitia lub poczucie realnego zagrożenia statusu jednostki[4]. Najczęściej badani wskazywali przykłady związane

[3] Wnioski te potwierdzają także nieopublikowane wyniki badań jakościowych (wywiadów), prowadzonych przez autorkę w latach 2009–2015.

[4] Jak podaje W. Szewczuk (1985), motywacja zewnętrzna, jako tendencja do podejmowania i kontynuowania działań ze względu na pewne konsekwencje, do których one prowadzą (uzyskanie nagrody lub innej uchwytnej korzyści), jest związana z poczuciem, że przyczyna

z możliwością uzyskania pracy czy też obietnicę lepszej pracy vs. zagrożenie utratą pracy; uzyskanie wsparcia socjalnego od rodziny lub instytucji vs. pozostawienie samemu sobie; uzyskanie przywileju uczenia się w więzieniu vs. brak przywilejów, itp. Pozostali respondenci (11%) wykazywali, iż na podejmowanie decyzji o uczeniu się w celu potwierdzenia kwalifikacji w trybie eksternistycznym miała wpływ zarówno motywacja wewnętrzna, jak i zewnętrzna.

Egzamin eksternistyczny jako sposób walidacji rezultatów uczenia się pozaszkolnego

Jakie możliwości potwierdzenia rezultatów nieformalnego lub pozaformalnego *uczenia się przez całe życie* stwarza obecnie polski system edukacji? Zewnętrzne egzaminy eksternistyczne organizowane są dwa razy do roku – w przypadku przedmiotów ogólnokształcących, oraz ustawicznie – w przypadku nowych egzaminów potwierdzających kwalifikacje w zawodzie. Populacja zdających nie jest tak liczna, jak w czasie, gdy egzaminy organizowały kuratoria (przypomnijmy – wg wskazań NIK było to ok. 300 tys. osób rocznie). Dotychczas nie przekroczyła ona 10 tys. zdających egzaminy z przedmiotów ogólnokształcących z zakresu szkoły podstawowej, gimnazjum, liceum ogólnokształcącego, zasadniczej szkoły zawodowej dla dorosłych oraz potwierdzające kwalifikacje zawodowe w jednym roku.

Na uzyskanie świadectwa ukończenia szkoły (np. liceum ogólnokształcącego) zdający mają trzy lata (sześć sesji egzaminacyjnych; w sytuacjach indywidualnych czas ten może zostać wydłużony o kolejne dwie sesje). W przypadku potwierdzenia kwalifikacji w zakresie LO konieczne jest zdanie egzaminów z 11 przedmiotów, dla których wymagania są zbieżne z obowiązującymi podstawami programowymi i standardami wymagań egzaminacyjnych[5]. Zdający indywidualnie budują ścieżkę egzaminacyjną, decydując, które przedmioty zdają w danej sesji. Najczęściej wybierają oni od 1 do 5 przedmiotów; średnio tylko

i kontrola działania ma charakter zewnętrzny i nie jest zależna od jednostki. Polega ona na wzbudzeniu potrzeb przez stosowanie kar i nagród, informowaniu o możliwościach zawartych w różnego rodzaju sytuacjach i manipulowaniu tymi możliwościami.

[5] Od roku 2016 zadania egzaminacyjne z zakresu LO będą odnosić się do nowej podstawy programowej – zob. *Informator o egzaminie eksternistycznym przeprowadzanym od sesji jesiennej 2016 z zakresu liceum ogólnokształcącego* Centralnej Komisji Egzaminacyjnej.

10% zdających w latach 2010–2015 decydowało się na zdawanie wszystkich egzaminów w jednej sesji[6].

Odsetek zdanych egzaminów ogółem oscyluje w przedziale od maks. 100% do min. 51%[7]. Zdawalność poszczególnych przedmiotów oraz pozytywny wynik potwierdzenia różnych kwalifikacji w zawodzie są w kolejnych sesjach bardzo zróżnicowane. Na przykład, uśredniony wskaźnik zdawalności egzaminów eksternistycznych w latach 2010–2015 dla największej populacji zdających egzaminy w zakresie liceum ogólnokształcącego dla dorosłych waha się od 97% zdanych egzaminów z podstaw przedsiębiorczości oraz wiedzy o społeczeństwie do 38% z matematyki i 46% z fizyki[8]. Egzamin z przedmiotów ogólnokształcących uważa się za zdany w przypadku, gdy zdający otrzyma minimum 30% punktów możliwych do uzyskania w teście egzaminacyjnym. Aby zdać egzamin potwierdzający kwalifikację w danym zawodzie, należy uzyskać co najmniej 50% punktów możliwych do uzyskania z części pisemnej i co najmniej 75% punktów z części praktycznej.

Egzamin potwierdzający kwalifikację w zawodzie oceniany jest w punktach. Oceniając egzaminy eksternistyczne z przedmiotów ogólnokształcących, stosuje się skalę szkolną – od oceny niedostatecznej po celującą. Analiza danych Centralnej Komisji Egzaminacyjnej dowodzi, iż średnia uzyskiwanych ocen nie jest najwyższa – dla sesji egzaminacyjnych w latach 2010–2015 wynosi ona 2,48[9], przy czym z przedmiotów matematyczno-przyrodniczych, takich jak matematyka, fizyka z astronomią i chemia, spada poniżej oceny dopuszczającej. Z kolei współczynnik zdawalności oraz średnia ocen świadczą o tym, iż eksterni nieźle orientują się w podstawach przedsiębiorczości oraz wiedzy o społeczeństwie (średnia odpowiednio: 3,4 oraz 3,18)[10]. Zdarzają się indywidualne przypadki i ocen celujących, i bardzo wysokich wyników eksterna ze wszystkich zdawanych przedmiotów.

Eksterni to osoby od 18. do ponad 60. roku życia. Wśród zdających egzaminy z przedmiotów ogólnokształcących dominują jednakże ludzie młodzi, od 18 do 35 lat, stanowiąc średnio 78% populacji[11]. Zdający mają więc zazwyczaj niewielką przerwę w edukacji. Odnotowano jednakże przypadki, gdy pauza

[6] Na podstawie danych Centralnej Komisji Egzaminacyjnej za lata 2010–2015.

[7] Dotyczy egzaminów z zakresu gimnazjum dla dorosłych przeprowadzanych w sesji zimowej 2008 roku.

[8] Na podstawie danych Centralnej Komisji Egzaminacyjnej za lata 2010–2015.

[9] Tamże.

[10] Tamże.

[11] Na podstawie danych Centralnej Komisji Egzaminacyjnej za lata 2007–2015.

w uczeniu się wynosiła kilkadziesiąt lat (np. zdający egzaminy 65-latek ukończył ośmioklasową szkołę podstawową 50 lat wcześniej).

Od samouczenia się do samorealizacji

Egzamin eksternistyczny to w polskim systemie oświaty jedyny sposób walidacji i potwierdzenia kwalifikacji nabytych w drodze nieformalnej lub pozaformalnej edukacji z zakresu kształcenia ogólnego na poziomie II, III i IV etapu edukacyjnego (od szkoły podstawowej po liceum ogólnokształcące). W przypadku kwalifikacji zawodowych, można je potwierdzić albo przystępując do egzaminów zewnętrznych, albo zdając egzamin czeladniczy lub mistrzowski w cechu rzemieślniczym. Zdający chętnie podkreślają, iż decyzja o powrocie na ścieżkę edukacyjną, choć często trudna w realizacji, dała im poczucie satysfakcji i umocniła w przekonaniu, że warto się trudzić. Jak twierdzą sami zainteresowani, „jak ktoś chce się uczyć (...), egzaminy eksternistyczne to chyba najlepsza forma z możliwych. (...) Osoby, które przez nie przebrną, mogą być naprawdę z siebie dumne. To ciężka praca nad sobą, która potęguje poczucie odniesionego sukcesu”[12]. Czterdzieści dwa procent badanych zdających egzaminy eksternistyczne z zakresu LO zadeklarowało, iż zdobycie wykształcenia średniego to jedynie droga do celu, gdyż 82% z nich chce zdawać maturę, a 76% – studiować (Lisiecka 2008: 27). Tym samym można z przekonaniem stwierdzić, iż polski model uczenia się pozaformalnego bądź nieformalnego, a przede wszystkim samouczenia się (*self-directed learning*)[13], którego rezultaty potwierdza się poprzez przystąpienie do egzaminów eksternistycznych, jest nie tylko drożny, ale też staje się wartością samą w sobie.

Bibliografia

Bacia E. 2012. *Jak potwierdzać kompetencje zdobyte na drodze pozaformalnej i nieformalnej?*, „Trendy”, nr 4, s. 64–68.

Badanie stosowanych procedur potwierdzania efektów uczenia się osiągniętych na drodze edukacji pozaformalnej i uczenia się nieformalnego. 2012, IBE, Warszawa.

[12] Wypowiedź osoby o nicku *gość* z 11 lutego 2015 roku. Zob.: http://f.kafeteria.pl/temat/f1/egzaminy-eksternistyczne-lo-p_4546203/276 [dostęp 17.03.2015].

[13] Za J. Półturzyckim (2002) można zdefiniować proces samokształcenia (samouczenia się) jako *proces uczenia się prowadzony świadomie, z możliwością wykorzystania różnych form pomocy innych osób lub instytucji, (…) którego cele, treść, formy, źródła i metody dobiera i ustala osoba ucząca się*.

Better Skills, Better Jobs, Better Lives: A Strategic Approach to Skills Policies. 2012. OECD Publishing, https://skills.oecd.org/documents/OECDSkills StrategyFINALENG.pdf

Informacja o wynikach kontroli organizacji egzaminów eksternistycznych w systemie oświaty. 2006, Najwyższa Izba Kontroli, Warszawa, http://www.nik.gov.pl/kontrole/wyniki-kontroli/nik/pobierz,px_2005025x. pdf, typ.k. pdf

Białecki I. 2010. *Wykluczenie edukacyjne*, [w:] *Ubóstwo i wykluczenie społeczne w Polsce*, Raport krajowy Polskiej Koalicji Social Watch i Polskiego Komitetu European Anti-Poverty Network, Warszawa.

Chłoń-Domińczak A., Kotowska I.E., Magda I., red. 2014. *Uwarunkowania decyzji edukacyjnych. Wyniki pierwszej rundy badania panelowego gospodarstw domowych*, IBE, Warszawa.

Ciechański J. 2004. *Procesy „boloński" i „kopenhaski" w europejskiej strategii kształcenia i szkolenia*, [w:] *Polityka zagraniczna i wewnętrzna państwa w procesie integracji europejskiej*, red. E. Haliżak, Brandt, Bydgoszcz, s. 327–354.

Ciżkowicz B. 2007. *Bezradność intelektualna a styl wyjaśniania zdarzeń*, [w:] *Czego obawiają się ludzie? Współczesne zagrożenia społeczne – diagnoza i przeciwdziałanie*, red. M. Libiszowska-Żółtkowska, Wydawnictwa Uniwersytetu Warszawskiego, Warszawa, s. 180–187.

Ciżkowicz B. 2009. *Wyuczona bezradność młodzieży*, Wydawnictwo Uniwersytetu Kazimierza Wielkiego, Bydgoszcz.

Jarosz M. 2004. *Samobójstwo. Ucieczka przegranych*, Wydawnictwo Naukowe PWN, Warszawa.

Jarosz M. red. 2008. *Wykluczeni. Wymiar społeczny, materialny i etniczny*, ISP PAN, Warszawa.

Kozarzewski P. 2008. *Wykluczanie edukacyjne*, [w:] *Wykluczeni*, red. M. Jarosz, ISP PAN, Warszawa.

Kraśniewski A. 2009. *Proces Boloński – to już 10 lat*, Fundacja Rozwoju Systemu Edukacji, Warszawa.

Kwieciński Z. 2002. *Wykluczanie*, Wydawnictwo Naukowe Uniwersytetu Mikołaja Kopernika w Toruniu, Toruń.

Lisiecka Z. 2008. *Portret eksterna*, Centralna Komisja Egzaminacyjna, Warszawa.

Lisiecka Z. 2015. *Pogmatwane drogi eksternów do „wyższych umiejętności, lepszej pracy i lepszego życia"*, [w:] *Zastosowania diagnozy edukacyjnej*, red. B. Niemirko i K.M. Szmigiel, Bydgoszcz.

Muras M. 2006. *Wykluczanie i integracja społeczna a rozwój zrównoważony*, [w:] *Wykluczanie i integracja społeczna w Polsce. Ujęcie wskaźnikowe*, UNDP i MPiPS, Warszawa.

Narwojsz H., Krawczyński C., Symela K., Zwiefka K. 2008. *Walidacja kompetencji nieformalnych uzyskanych poprzez doświadczenie w pracy. Metodyka – Procedury – Dokumentacja*, Warmińsko-Mazurski Zakład Doskonalenia Zawodowego, Olsztyn.

Perspektywa uczenia się przez całe życie. 2013, http://men.gov.pl/wp-content/uploads/2013/09/perspektywy_uczenia_sie_przez_cale_zycie.pdf

Półturzycki J. 2002. *Dydaktyka dla nauczycieli*, Wyd. Novum, s. 209–210.

Rozporządzenie Ministra Edukacji Narodowej z dnia 18 września 2007 r. w sprawie egzaminów eksternistycznych, Dz.U. nr 179, poz. 1273.

Rozporządzenie Ministra Edukacji Narodowej z dnia 11 stycznia 2012 r. w sprawie egzaminów eksternistycznych, Dz.U. poz. 188.

Sawiński Z. 2008. *Zmiany systemowe a nierówności w dostępie do wykształcenia*, [w:] *Zmiany stratyfikacji społecznej w Polsce*, red. H. Domański, IFiS PAN, Warszawa.

Sprawozdanie z przebiegu egzaminu potwierdzającego kwalifikacje w zawodzie w 2013 roku (nowa formuła egzaminu). 2014, Centralna Komisja Egzaminacyjna, Warszawa.

Sprawozdanie z osiągnięć zdających egzamin potwierdzający kwalifikacje w zawodzie w roku szkolnym 2013/2014. 2014, Centralna Komisja Egzaminacyjna, Warszawa.

Osiągnięcia zdających egzaminy eksternistyczne w roku szkolnym 2012/2013. 2013, Centralna Komisja Egzaminacyjna, Warszawa.

Sprawozdania z egzaminów eksternistycznych w sesji jesiennej 2013 oraz *2014*. 2013; 2015 oraz *Sprawozdania z egzaminów eksternistycznych w sesji zimowej 2014* oraz *2015*. 2014; 2015. Centralna Komisja Egzaminacyjna, Warszawa.

Szewczuk W. red. 1985. *Słownik psychologiczny*, Wiedza Powszechna, Warszawa.

www.cke.edu.pl/egzaminy eksternistyczne

KWARTALNIK PEDAGOGICZNY
2015 NUMER 4(238)

RECENZJE

American Educational Research Association, American Psychological Association, National Council on Measurement in Education, *Standards for educational and psychological testing*, American Educational Research Association, Washington 2014, 240 s.

W drugiej połowie ubiegłego roku Amerykańskie Towarzystwo Badań Edukacyjnych (*American Educational Research Association* – AERA) opublikowało kolejną edycję *Standardów edukacyjnego i psychologicznego testowania* (AERA i in. 2014). Wersję tę przygotował Wspólny Komitet ds. Standardów Pomiaru Edukacyjnego i Psychologicznego (*Joint Committee on the Standards for Educational and Psychological Testing*) trzech towarzystw naukowych: Amerykańskiego Towarzystwa Psychologicznego (*American Psychological Association* – APA), Narodowej Rady do spraw Pomiarów Edukacyjnych (*National Council on Measurement in Education* – NCME) oraz AERA. W składzie Komitetu, powołanego w 2008 roku, znaleźli się: Barbara S. Plake (jako współprzewodnicząca), Lauress L. Wise (współprzewodniczący), Linda L. Cook, Fritz Drasgow, Brian T. Gong, Laura S. Hamilton, Jo-Ida Hansen, Joan L. Herman, Michael T. Kane, Michael J. Kolen, Antonio E. Puente, Paul R. Sackett, Nancy T. Tippins, Walter D. Way oraz Frank C. Worrell.

Nowa edycja *Standardów* zawiera 249 standardów profesjonalizmu w zakresie pomiaru edukacyjnego i psychologicznego, uporządkowanych w trzech częściach. Część pierwsza – *Podstawy* – obejmuje 68 standardów związanych z problematyką trafności (1), rzetelności (2) oraz stronniczości (3) testów. W części *Operacje* zamieszczono 119 standardów dotyczących projektowania i konstrukcji testu (4), wyników, skal i norm (5), przeprowadzania testu, punkto-

wania zadań, raportów i interpretacji (6), dokumentacji testu (7), a także praw i obowiązków jego użytkowników (8–9). W *Zastosowaniach testów* znalazły się standardy rekomendowane w dziedzinach testów psychologicznych (10), zawodowych (11), edukacyjnych (12) oraz ewaluacji (13). W każdym rozdziale dwóch pierwszych części wskazano rekomendację nadrzędną, rozwiniętą w standardach jej podporządkowanych, z których każdy opatrzono wielozdaniowym komentarzem, i poprzedzoną kilkustronicowym wprowadzeniem teoretycznym. W części trzeciej zastosowano ten sam schemat kompozycyjny – z wyjątkiem nadrzędnych standardów, których nie wyodrębniono. Teoretyczny wymiar *Standardów* dopełniono obszernym słownikiem oraz indeksem zastosowanych terminów pomiarowych, ulokowanymi w końcowych partiach publikacji.

Wśród wspomnianych dziewięciu rekomendacji nadrzędnych znalazły się między innymi standardy formułowania każdej zamierzonej interpretacji wyników testu, przedstawiania – w odniesieniu do proponowanego zastosowania testu – odpowiednich dowodów jego trafności, rzetelności, bezstronności (*fairness*), minimalizowania – w trakcie procesu konstruowania i użytkowania testu – wariancji irrelewantnej względem konstruktu, udostępniania dokumentacji powstałej podczas prac konstruktorskich, precyzyjnego określania procedur administrowania testu, punktowania zadań, tworzenia raportów i interpretowania wyników, poszanowania określonych praw przystępujących do testu oraz przejmowania odpowiedzialności za właściwe użytkowanie testu przez jego użytkowników. Tradycyjnie rekomendacjom powyższym, podobnie jak towarzyszącym im standardom, nadano formę zdań powinnościowych, na przykład nadrzędna formuła podrozdziału *Standardy rzetelności/precyzji* otrzymała postać: „Odpowiednie dowody rzetelności/precyzji powinny być udostępnione na potrzeby interpretacji związanej z każdym zamierzonym użyciem wyniku" (standard 2.0), zaś jeden z 20 uszczegóławiających ją standardów: „Standardowy błąd pomiaru [...] powinien być podany w jednostkach każdego komunikowanego wyniku" (2.13).

Zasygnalizowana powyżej priorytetyzacja standardów oraz inwersja kompozycyjna, polegająca między innymi na umiejscowieniu w *Podstawach* standardów powiązanych z problematyką bezstronności testu, w poprzedniej – również strukturalnie trójdzielnej – edycji zamieszczonych w jej czterorozdziałowej części drugiej, zatytułowanej: *Bezstronność testowania*, stanowią dwie z trzech istotnych modyfikacji wprowadzonych w piątym już wydaniu *Standardów*. Trzecią współtworzą substancjalne oraz stylistyczne zmiany w tekście przejętym z edycji wcześniejszej. Te pierwsze dotyczą przede wszystkim wspomnianej już kwestii bezstronności praktyk pomiarowych. Obejmują także problemy związane z roz-

wojem technologii informacyjnej oraz złożoności towarzyszące użytkowaniu testów i ich wyników przez podmioty zainteresowane zewnętrzną ewaluacją edukacji. Do ujednolicenia terminologii oraz stylu prowadzą pozostałe zmiany w tekście *Standardów*.

Adresatami w ten sposób zaktualizowanego kodeksu profesjonalizmu w dziedzinie pomiaru edukacyjnego i psychologicznego pozostają nie tylko twórcy, lecz także użytkownicy testów. *Standardy* stanowią bowiem zbiór kryteriów oceny wartości testu, przydatny podczas refleksji i dyskusji nad stanem praktyki pomiarowej. W zamyśle pełnić mogą również funkcję dydaktyczną, nie wyłącznie w amerykańskim kręgu kulturowym. O roli, jaką rzeczywiście odgrywają, wspominają między innymi Plake i Wise (2015), podkreślając, iż jak dotąd *Standardom* nie nadano mocy prawnej, i wskazując potencjalne nadużycia związane z ich użytkowaniem, na przykład przedkładanie litery *Standardów* nad sąd oparty na wiedzy.

Bibliografia

Plake B.S., Wise L.L. 2014. *What is the role and importance of the revised AERA, APA, NCME Standards for educational and psychological testing?*, „Educational Measurement: Issues and Practice" (Wydanie specjalne: The AERA/APA/NCME *Standards for Educational and Psychological Testing*), t. 33, nr 4, s. 4–12.

Piotr Maciej Skorupiński

Noty o autorach

Ciżkowicz Barbara – doktor habilitowany, profesor Instytutu Psychologii Uniwersytetu Kazimierza Wielkiego w Bydgoszczy. Zainteresowania badawcze: metodologia badań społecznych, wielozmiennowe analizy ilościowe, psychometria. Wybrane publikacje: *Wyuczona bezradność młodzieży* (2009); *Obraz polskiej szkoły w badaniach 15-latków*, „*Teraźniejszość – Człowiek – Edukacja*" 2013, nr 4 (64), s. 117–130; *O potrzebie ewaluacji efektów kształcenia w zreformowanych szkołach wyższych: stan, problemy, perspektywy*, „Ruch Pedagogiczny" 2014, 2, s. 123–134. Adres do korespondencji: Instytut Psychologii UKW, 85-867 Bydgoszcz, ul. Staffa 1. E-mail: cizbar@ukw.edu.pl

Daszkiewicz Michał – doktor, adiunkt na Wydziale Nauk Społecznych Uniwersytetu Gdańskiego. Pracuje w Instytucie Pedagogiki w Zakładzie Badań nad Dzieciństwem i Szkołą. Zainteresowania badawcze: językowe zdarzenia edukacyjne, werbalizacja wiedzy, fosylizacja języka obcego i ojczystego, dwujęzyczność, wypowiedź ukierunkowana. Wybrane publikacje: *On the gap between understanding and applying second-language expressions and its dependence on textual features*, [w:] *PASE Papers 2007: Studies in Language and Methodology of Teaching Foreign Languages*, red. J. Arabski, D. Gabryś-Barker, A. Łyda; *Common Sense of L2 Instruction in Early Education. Its Salience and Transferability*, „Problemy Wczesnej Edukacji", nr 2 (14); *Theoretical Discussion on How Interdisciplinary Knowledge Serves Educational Practice*, „Journal of Social Sciences", Elbląg: EUH-E, s. 119–139. Adres do korespondencji: ul. Bażyńskiego 4, 80-952 Gdańsk.
E-mail: pedmd@univ.gda.pl

Filipiak Ewa Maria – profesor zwyczajny, pracownik naukowy na Uniwersytecie Kazimierza Wielkiego w Bydgoszczy, na Wydziale Pedagogiki i Psychologii; kierownik Katedry Dydaktyki i Studiów nad Kulturą Edukacji. Zainteresowania naukowe: edukacja jako czynnik rozwoju jednostki, zwłaszcza problem rozwijania zdolności uczenia się oraz dydaktyczne aplikacji socjokulturowe koncepcji Lwa S. Wygotskiego i Jerome'a Brunera. Autorka licznych publikacji z zakresu dydaktyki i pedagogiki wczesnoszkolnej. Pomysłodawca i realizator projektów badawczych zorientowanych na podnoszenie jakości edukacji. Kierownik projektu MNiSW w latach 2006–2009: „Rozwój zdolności uczenia się uczniów szkoły podstawowej i gimnazjum". Aktualnie kierownik projektu „Akademickie Centrum Kreatywności". Ważniejsze publikacje – cykl wydawniczy: *Z Wygotskim i Brunerem w tle: słownik pojęć kluczowych* (2011); *Z Wygotskim i Brunerem w tle. Rozwijanie zdolności uczenia* (2012). Adres do korespondencji: Instytut Pedagogiki, ul. Chodkiewicza 30, 85-064 Bydgoszcz. E-mail: ewafil@ukw.edu.pl

Groenwald Maria – doktor habilitowany, profesor Uniwersytetu Gdańskiego. Pracuje w Zakładzie Badań nad Dzieciństwem i Szkołą, Instytutu Pedagogiki na Wydziale Nauk Społecznych UG. Jej zainteresowania badawcze koncentrują wokół diagnostyki edukacyjnej, pedagogiki szkolnej,

etycznych aspektów edukacji oraz pedeutologii. Autorka szeregu opracowań związanych z tą tematyką, m.in. książki: *Etyczne aspekty egzaminów szkolnych* (2011). Adres do korespondencji: Instytut Pedagogiki WNS Uniwersytetu Gdańskiego, ul. Bażyńskiego 4, 80-952 Gdańsk. E-mail: maria.groenwald@ug.edu.pl

Humenny Grzegorz – doktor, pracuje w Instytucie Badań Edukacyjnych, w Pracowni Edukacyjnej Wartości Dodanej, asystent, specjalista ds. badań i analiz. Zajmuje się socjologią edukacji, badaniem sieci społecznych, poczuciem samotności, wykorzystuje wielowymiarową i wielopoziomową analizę danych. Obecnie analizuje trafność wskaźników EWD dla szkół podstawowych. Wybrane publikacje: *Trafność metody edukacyjnej wartości dodanej dla gimnazjów* (2013), *Kontekstowy model efektywności nauczania po pierwszym etapie edukacyjnym* (2014), *Peer Rejection and Perceived Quality of Relations With Schoolmates Among Children With ADHD* (2014), *Wielowymiarowa struktura latentna w perspektywie analizy czynnikowej* (w druku), *Poza ścisłą jedno- i wielowymiarowość. Struktura czynnikowa skali samotności de Jong Gierveld wśród dzieci* (w druku). E-mail: g.humenny@ibe.edu.pl

Jasińska-Maciążek Aleksandra – doktor, pracuje w Instytucie Badań Edukacyjnych, w Pracowni Edukacyjnej Wartości Dodanej, asystent, specjalista ds. badań i analiz pedagogicznych. Uczestniczyła w projektach badawczych poświęconych m.in. uwarunkowaniom wyników nauczania. Jej zainteresowania naukowe koncentrują się wokół pomiaru dydaktycznego, ewaluacji instytucji edukacyjnych, metod oceny efektywności nauczania oraz modelowania statystycznego. Wybrane publikacje: *Kontekstowy model efektywności nauczania po pierwszym etapie edukacyjnym* (2014), *Testy osiągnięć szkolnych TOS3: przykład narzędzia skonstruowanego z wykorzystaniem modelu Rascha* (2014), *Trafność metody edukacyjnej wartości dodanej dla gimnazjów* (2013), *Międzyszkolne zróżnicowanie wyników nauczania po pierwszym etapie kształcenia* (2013). Adres do korespondencji: IBE, ul. Górczewska 8, 01-180 Warszawa. E-mail a.jasinska@ibe.edu.pl

Jaskuła Sylwia – doktor pedagogiki, adiunkt w Instytucie Społeczno-Humanistycznym w Państwowej Wyższej Szkole Informatyki i Przedsiębiorczości w Łomży, wykładowca Akademii Pedagogiki Specjalnej, Collegium Civitas oraz UKSW w Warszawie. Publikuje wiele nt. ewaluacji, transformacji i społecznych wyzwań edukacji, teorii przestrzeni międzykulturowej, społeczeństwa sieci oraz kompetencji informacyjno-komunikacyjnych w kontekście globalnych i lokalnych zmian tożsamości kulturowej. Animatorka badań ewaluacyjnych w edukacji oraz porównawczych analiz społeczeństwa informacyjnego w warunkach przemian rynkowych i kulturowych. Jest autorką programów szkoleniowych animujących uspołeczniony model ewaluacji, w tym ewaluacji międzykulturowej. Ważniejsze publikacje: *Odrębność i współgranie. Polskie relacje diagnostyki i ewaluacji edukacyjnej* (2014), *Pedagogika ewaluacji* (2013), *Metodologia procesu ewaluacyjnego* (2013), *Ewaluacja rozwojowa jako wyzwanie współczesnego systemu edukacji* (2012), *Mutualizm ewaluacji i diagnozy edukacyjnej* (2012), *Ewaluacja versus akredytacja – dwa światy w jednej przestrzeni* (2011), *O potrzebie re-ewaluacji w społeczeństwie nadmiaru informacji* (2010). Adres do korespondencji: Instytut Społeczno-Humanistyczny PWSIiP, ul. Akademicka 14, 18-400 Łomża. E-mail: sylwia.jaskula@poczta.onet.pl

Korporowicz Leszek – doktor habilitowany, socjolog kultury, pracownik naukowy Instytutu Studiów Międzykulturowych UJ; współzałożyciel, a obecnie prezes honorowy Polskiego Towarzystwa Ewaluacyjnego. Jest ekspertem wielu urzędów centralnych, współautorem rządowych

projektów ewaluacyjnych. W latach 2004–2006 był członkiem Państwowej Komisji Akredytacyjnej, pełnił kierownicze funkcje akademickie w uczelniach Warszawy i Krakowa. Autor wielu prac z dziedziny socjologii kultury, studiów międzykulturowych, w tym: komunikacji międzykulturowej, społecznych i metodologicznych podstaw badań ewaluacyjnych, szczególnie ewaluacji rozwojowej oraz edukacyjnych i badawczych projektów animujących humanistyczny kierunek analiz kultury współczesnej, jak Jagiellońskie Studia Kulturowe. Autor takich prac, jak: *Socjologia kulturowa* (2011), *Osobowość i komunikacja w społeczeństwie transformacji* (1997), *Ewaluacja w edukacji* (red. 1996). E-mail: leszek.korporowicz@poczta.onet.pl

Lisiecka Zofia – ekspert ds. egzaminów humanistycznych w Zespole Egzaminów Eksternistycznych Centralnej Komisji Egzaminacyjnej w Warszawie. Polonistka, ukończyła także studia podyplomowe z zarządzania, ewaluacji dydaktycznej oraz ewaluacji projektów unijnych. Wcześniej nauczycielka w szkole podstawowej, gimnazjum i liceum ogólnokształcącym, nauczyciel akademicki. Doświadczony trener-szkoleniowiec, recenzent i ewaluator projektów edukacyjnych. Autorka lub współautorka kilkudziesięciu publikacji, m.in.: *O potrzebie badań nad efektem zwrotnym egzaminu maturalnego z języka polskiego*, (współaut.), „Ruch Pedagogiczny" 2014, 2, s. 83–92, *W drodze do sukcesu – wspólnie i odrębnie. Teoretyczny model wspierania uczniów w Szkole Startu do Kariery*, (współaut.), „Meritum" 2013, 9, *Badania panelowe w szkołach podstawowych i gimnazjach a rozwój wskaźnika EWD. Doniesienie badawcze*, [w:] *Teraźniejszość i przyszłość oceniania szkolnego*, red. B. Niemierko, M.K. Szmigiel, XVI Krakowa Konferencja Diagnostyki Edukacyjnej, PTDE, Toruń 2010, s. 425–437. E-mail: zlisiecka@gmail.com

Niemierko Bolesław – profesor zwyczajny, obecnie pracuje na Uniwersytecie Humanistyczno-społecznym SWPS, w Zakładzie Wspomagania Rozwoju w Sopocie. Zainteresowania badawcze: dydaktyka, pomiar dydaktyczny, diagnostyka edukacyjna. Najważniejsze publikacje: *Testy osiągnięć szkolnych. Podstawowe pojęcia i techniki obliczeniowe* (1975), *Pomiar sprawdzający w dydaktyce. Teoria i zastosowania* (1990), *Pomiar wyników kształcenia* (1999), *Ocenianie szkolne bez tajemnic* (2002), *Kształcenie szkolne. Podręcznik skutecznej dydaktyki* (2007), *Diagnostyka edukacyjna. Podręcznik akademicki* (2009). Adres do korespondencji: Uniwersytet SWPS, Wydział Zamiejscowy w Sopocie, Zakład Wspomagania Rozwoju, 81-745 Sopot, ul. Polna 16-20.
E-mail: pedbn@ug.edu.pl

Skorupiński Piotr Maciej – doktor, nauczyciel III Liceum Ogólnokształcącego im. gen. J. Sowińskiego w Warszawie. Zainteresowania naukowe: teoria i praktyka pomiaru edukacyjnego i psychologicznego. Autor artykułów poświęconych trafności pomiaru, m.in.: *Geneza pojęcia trafności teoretycznej* (2003), *Walidacja konstruktu* (2005), *Modele trafności pomiaru* (2013), *Trafność użytkowania wyników egzaminu* (2013) i *Recepcja koncepcji trafności w diagnostyce edukacyjnej* (2014). E-mail: kadmeita@wp.pl

Skórska Paulina – pracuje jako statystyk w Zespole Analiz Osiągnięć Uczniów w Instytucie Badań Edukacyjnych w Warszawie. Absolwentka socjologii i psychologii Uniwersytetu Jagiellońskiego. Współpracuje od kilku lat z Centrum Ewaluacji i Analiz Polityk Publicznych Uniwersytetu Jagiellońskiego, jako koordynator i ekspert w projektach z obszaru polityk publicznych ukierunkowanych na rozwój regionalny, planowanie strategiczne oraz rozwój systemu oświaty. Współpracowała z Ministerstwem Edukacji Narodowej oraz Ośrodkiem Rozwoju Edukacji jako

ekspert do spraw metodologii badań społecznych i statystycznej analizy danych. Zainteresowania naukowe skupiają się na metodologii badawczej nauk społecznych, technikach statystycznej analizy danych oraz zagadnieniach socjologii edukacji. Adres do korespondencji: IBE, ul. Górczewska 8, 01-180 Warszawa. E-mail: p.skorska@ibe.edu.pl

Szyling Grażyna – doktor nauk humanistycznych w dziedzinie pedagogiki, adiunkt w Zakładzie Badań nad Dzieciństwem i Szkołą Instytutu Pedagogiki Uniwersytetu Gdańskiego. Zainteresowania badawcze: diagnostyka edukacyjna i ewaluacja oświatowa, a zwłaszcza ocenianie szkolne, wpisane w konteksty dydaktyczne, ideologiczne i społeczne. Wybrane publikacje: *Nauczycielskie praktyki oceniania poza standardami* (2011). Adres do korespondencji: Instytut Pedagogiki Wydziału Nauk Społecznych Uniwersytetu Gdańskiego, ul. Bażyńskiego 4, 80-952 Gdańsk. E-mail: pedgs@univ.gda.pl

Trempała Janusz – profesor zwyczajny, pracuje w Instytucie Psychologii Uniwersytetu Kazimierza Wielkiego w Bydgoszczy. Jego zainteresowania badawcze koncentrują się na psychologii rozwoju człowieka. Wybrane publikacje: *Psychologia rozwoju człowieka* (red. 2011). Adres do korespondencji: Instytut Psychologii UKW, 85-867 Bydgoszcz, ul. Staffa 1.
E-mail: tremjan@ukw.edu.pl

Informacja o czasopiśmie

„Kwartalnik Pedagogiczny" to ogólnopolskie czasopismo naukowe ukazujące się od 1956 roku, publikujące artykuły teoretyczne i badawcze z zakresu szeroko pojętej pedagogiki i jej subdyscyplin oraz innych nauk społecznych.

„Kwartalnik Pedagogiczny" jest czasopismem punktowanym, umieszczonym w bazie ERIH (European Reference Index for the Humanities), C.E.E.O.L. (Central and Eastern European Online Library) oraz CEJSH (The Central European Journal of Social Sciences and Humanities). Zgodnie z ostatnim wykazem czasopism naukowych MNiSW za publikację w KP przysługuje Autorowi 10 punktów.

Rzetelność naukowa

Redakcja dokłada wszelkich starań, aby eliminować wszelkie przejawy nierzetelności naukowej w procesie publikacyjnym, takie jak *ghostwriting* oraz *guest autorship*.

Z *ghostwriting* mamy do czynienia wówczas, gdy ktoś wniósł istotny wkład w powstanie publikacji, bez ujawnienia swojego udziału jako jeden z autorów lub bez wymieniania jego roli w podziękowaniach zamieszczonych w publikacji. Redakcja prosi więc o podawanie informacji na temat Współautorów badań, których wyniki referowane są w artykule, jak również określenie własnego wkładu w uzyskane rezultaty badawcze.

Z *guest authorship* mamy do czynienia wówczas, gdy udział Autora jest znikomy lub w ogóle nie miał miejsca, a mimo to jest on podany jako Autor bądź Współautor publikacji.

W związku z tym Redakcja wdrożyła procedury zapobiegające przypadkom nierzetelności naukowej i wymaga od Autorów: ujawnienia wkładu poszczególnych autorów w powstanie publikacji oraz podawania źródeł finansowania publikacji, a także informacji o wkładzie instytucji naukowo-bawawczych, stowarzyszeń i innych podmiotów. Wszelkie przejawy nierzetelności naukowej będą dokumentowane i ujawniane.

Procedura recenzyjna

Procedura recenzyjna „Kwartalnika Pedagogicznego" jest zgodna z wytycznymi Ministerstwa Nauki i Szkolnictwa Wyższego. Redakcja nadaje nadesłanej pracy numer, identyfikujący ją podczas całego procesu recenzyjnego, i dokonuje wstępnej oceny nadesłanych materiałów pod kątem zgodności z profilem naukowym pisma, wymogów formalnych oraz poziomu merytorycznego. Teksty wstępnie zakwalifikowane do druku przekazywane są dwóm recenzentom zewnętrznym do szczegółowej recenzji zgodnie z procedurą *double-blind reviews* (podwójna ślepa recenzja), co oznacza, że zarówno Autor, jak i Recenzent pozostają dla siebie anonimowi. W przypadku tekstów powstałych w języku obcym co najmniej jeden z recenzentów jest afiliowany w instytucji zagranicznej innej niż narodowość Autora pracy. Recenzenci są samodzielnymi pracownikami naukowymi i nie są członkami Redakcji ani Rady Naukowej pisma, nie są zatrudnieni w jednostce naukowej, z której pochodzi Autor, ani w jednostce wydającej czasopismo. Recenzja ma formę pisemną i jest udostępniana Autorowi. Recenzent może: odrzucić tekst, zakwalifikować bez zastrzeżeń, zakwalifikować warunkowo (co wiąże się z dokonaniem przez Autora poprawek zasugerowanych w recenzji). Lista Recenzentów publikowana jest raz do roku. Ostateczną kwalifikację do druku podejmuje redaktor naczelny.

Informacje dla Autorów

Prosimy Autorów, zamierzających opublikować swój tekst w „Kwartalniku Pedagogicznym", o przestrzeganie następujących zasad przy przygotowaniu maszynopisu:

- Maksymalna objętość całego tekstu (z bibliografią i streszczeniem) powinna wynosić 1 arkusz wydawniczy, to jest 40 000 znaków typograficznych (ze spacjami) lub ok. 22 strony znormalizowane (30 wierszy na stronę, około 60 miejsc znakowych w wierszu).

  Prosimy nie zamieszczać w tekście kolorowych diagramów, wykresów i ilustracji. Akceptujemy tylko tabele.

  Teksty należy wysłać pocztą elektroniczną jako załącznik na adres:

  kwartalnik.pedagogiczny@uw.edu.pl

- W tekście należy podać pełne imiona osób wymienionych po raz pierwszy.
- Tabele i ilustracje, ponumerowane i opatrzone informacją, do którego miejsca w tekście się odnoszą, powinny być dołączone na końcu maszynopisu, na oddzielnych stronach.
- Diagramy, wykresy i ilustracje muszą być dodatkowo zapisane jako oddzielne pliki z rozszerzeniem tiff lub jpeg.
- Obowiązuje następujący standard przypisów:

  Przypisy bibliograficzne umieszczone są bezpośrednio w tekście i zawierają tylko nazwisko autora, rok wydania dzieła i numer strony, np.: (Jaeger 2000: 5).

  Gdy przytaczamy dwie książki tego samego autora wydane w tym samym roku, stosujemy zapis: (Kozakiewicz 2001a; Kozakiewicz 2001b).

  Jeśli cytujemy kilku autorów, przyjmujemy porządek chronologiczny, a nie alfabetyczny: (Kwaśnica 2001: 5; Cannatella 2006: 78).

  Jeśli przytaczana praca ma dwóch autorów, umieszczamy oba nazwiska; w przypadku większej liczby autorów, podajemy pierwsze nazwisko i „i in.", np.: (Dehnel i in. 2001).

  Jeśli cytujemy dokumenty, podajemy początek zapisu bibliograficznego, np.: (Dz.U. 2006, nr 97, poz. 674), (Ankieta socjologiczna...).

- Taka forma przypisów wymaga odpowiedniego zacytowania w bibliografii, zamieszczonej po tekście artykułu, np.:

  Jaeger W. 2000. *Paideia. Formowanie człowieka greckiego*, tłum. M. Plezia, H. Bednarek, Fundacja Aletheia, Warszawa.

  Reut M. 2005. *Narracja, tożsamość, etyczność*, [w:] *Filozofia a pedagogika. Studia i szkice*, red. P. Dehnel, P. Gutowski, Wydawnictwo Naukowe Dolnośląskiej Szkoły Wyższej Edukacji TWP we Wrocławiu, Wrocław, s. 66–79.

  Kozakiewicz M. 1993. *Wychowanie do demokracji w okresie wielkiej zmiany w Polsce*, „Kwartalnik Pedagogiczny", nr 2, s. 21–28.

  Ustawa z 26.01.1982, Karta Nauczyciela; tekst ujednolicony: „Dziennik Ustaw" 2006 nr 97, poz. 674 z poprawkami.

  Ankieta socjologiczna na temat świadomości historycznej młodzieży, przeprowadzona na zlecenie OBOP w latach 1991–92, 827 ankiet, BUW Rps nr akces. 3174.

Opis bibliograficzny strony WWW musi zawierać nazwisko autora i tytuł pracy, tryb dostępu, protokół dostępu (miejsce, ścieżka, nazwa), a także datę dostępu, np.:

Teresa Hejnicka-Bezwińska, *Teraźniejszość w dialogu z edukacyjną przeszłością (po roku 1989)*, dostępny na: http://www.pedagog.uw.edu.pl/THE/Hejnicka.pdf (otwarty 25 stycznia 2007).

- W bibliografii należy podać nazwiska tłumaczy cytowanych prac obcojęzycznych.
- Numerowane i umieszczone na dole strony są tylko przypisy odautorskie.
- Do tekstu należy dołączyć krótkie streszczenie w języku polskim i angielskim (20 wierszy łącznie z tytułem artykułu) oraz słowa klucze w języku angielskim, a także na osobnej stronie krótką notę o autorze. Jeśli występuje więcej niż jeden autor tekstu, należy nadesłać noty biograficzne wszystkich autorów oraz pełny adres korespondencyjny i adres poczty elektronicznej autora, który będzie prowadził korespondencję dotyczącą publikacji.
- Nadesłane teksty zgodne z wymaganiami redakcyjnymi są opiniowane anonimowo (bez identyfikacji autora tekstu) przez trzech recenzentów.

Przed nadesłaniem tekstu do publikacji prosimy o wypełnienie METRYCZKI, która znajduje się na stronie http://www.kp.edu.pl/ –> Dla Autorów, i załączenie jej jako odrębnego pliku.

Maszynopis przysłany do Redakcji jest traktowany przez Redakcję i Wydawcę jako utwór oryginalny, do którego Autor dysponuje pełnią praw. Jeśli Autor utworu publikował go przed nadesłaniem do „Kwartalnika Pedagogicznego", musi opatrzyć utwór (swój tekst) stosownym przypisem, zawierającym informację o miejscu i czasie publikacji.

Maszynopis przysłanego do Redakcji tłumaczenia jest traktowany przez Redakcję i Wydawcę jako tłumaczenie (utwór), do dokonania którego Autor dysponował pełnią praw.

Redakcja zastrzega sobie prawo do wprowadzenia poprawek i skrótów. Redakcja nie zwraca tekstów nadesłanych, które nie ukazały się w „Kwartalniku Pedagogicznym".

INFORMACJA O PRENUMERACIE

Ceny „Kwartalnika Pedagogicznego" w roku 2015:
Prenumerata roczna (4 numery) – 80,00 zł,
Prenumerata półroczna (2 numery) – 40,00 zł,
Opłata za pojedynczy numer – 20,00 zł.

Zamówienia na pojedyncze egzemplarze prosimy kierować na adres:
dz.handlowy@uw.edu.pl

Zamówienia na prenumeratę „Kwartalnika Pedagogicznego":

RUCH S.A.
Zamówienia na prenumeratę w wersji papierowej i na e-wydania można składać bezpośrednio na stronie www.prenumerata.ruch.com.pl
Ewentualne pytania prosimy kierować na adres e-mail: prenumerata@ruch.com.pl
lub kontaktując się z Centrum Obsługi Klienta „RUCH" pod numerami: 22 693 70 00
lub 801 800 803 – czynne w dni robocze w godzinach 7^{00} – 17^{00}.
Koszt połączenia wg taryfy operatora.

Zamówienia na prenumeratę przyjmują również:
KOLPORTER SA, http://dp.kolporter.com.pl/prenumerata
GARMOND PRESS SA, www.garmondpress.pl/garmond/prenumerata
Subscription orders for all magazines published in Poland available through the local press distributors or directly through the Foreign Trade Enterprise **ARS POLONA SA, ul. Obrońców 25, 00-933 Warszawa, Poland**. BANK HANDLOWY SA, Oddział w Warszawie, PL 02 1030 1016 0000 0000 0089 5001 or **IPS, ul. Noakowskiego 10 lok. 38, 00-664 Warszawa, Poland**, tel. (48 22) 625 16 63, e-mail: ma@ips.com.pl

Indeks 363146 PL ISSN 0023-5938