
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 pomyślę 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; Koscicova 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 pracują¹.

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 się², nie jest określony stan, lecz jego zmiana – niezastnienie 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 wprowadzenie pomiaru 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 różne, 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 różnych 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. Popołniamy tu w sposób mniej lub bardziej świadomy wiele błędów związanych z wpływem czasu, w którym dane zjawisko badamy³. 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 dynamiczne⁵.

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 różnym 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 wielozmiennych 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. Nurmi (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

naprawiono⁶, 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 szczególnie gólnym przypadkiem sytuacji poprzednio omówionej, tj. uśredniania wyników pomiaru. Polega na traktowaniu różnic wewnątrz-osobniczych, 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 wewnątrz-osobniczych, 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 predykcyjna 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.

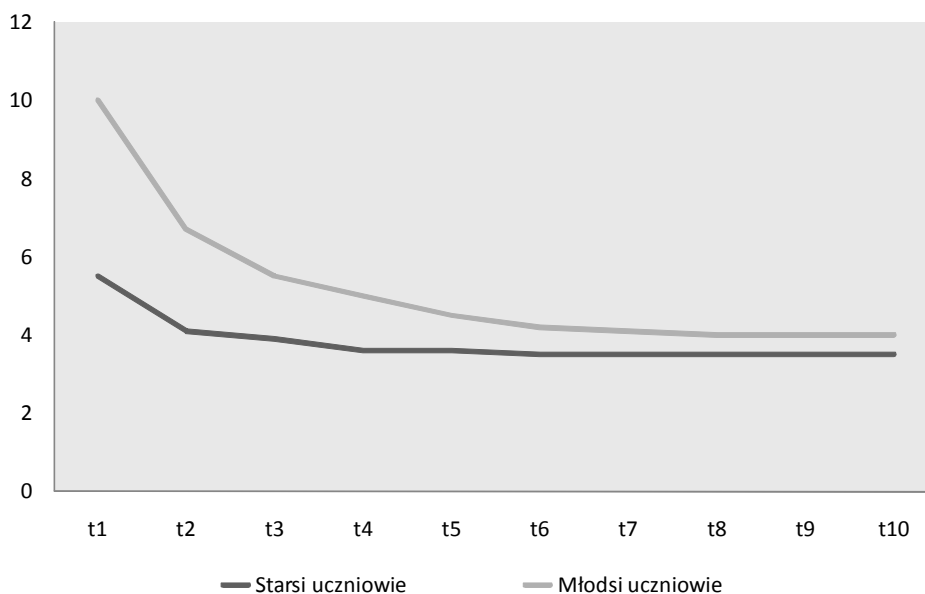
Trafność oceny różnic między średnimi grupowymi. Z problemem tym często mamy do czynienia w badaniach laboratoryjnych lub tzw. studiach treningowych (*training studies*), w których badani uczniowie w różnym wieku życia lub w różnych warunkach pomiaru (np. różnych programów edukacyjnych), wykonują w kolejnych 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 pomiaru⁷.

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-ciomą 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 młodszymi⁸. 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 (Trempeła 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 wprowadzie 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 (Tremplała, Ciechuch, w przyg.).

Bibliografia

- Aunola K., Leskinen E., Onatsu-Arviolommi 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. Pulkinen, 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. Connolly, 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 slows 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. Weintraub, Cambridge University Press, New York, s. 181–214.
- Sternberg R.J., Roediger III H.L., Halpieren 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., Ciecuch J. (w przyg.). *Analiza zmian w zachowaniu i rozwoju człowieka. O błędach w badaniu dynamiczności*.