

Sławomir Pasikowski

Uniwersytet Łódzki

E-MAIL: slawomir.pasikowski@now.uni.lodz.pl ORCID: 0000-0002-0768-1596

**Niepewności psychometrycznej praktyki.
W nurcie odczytań książki Pawła Kleki pt.
Trafność pomiaru w praktyce psychometrycznej.
Badania empiryczne,
Poznań: Wydawnictwo Nauk Humanistycznych
i Społecznych Uniwersytetu im. Adama Mickiewicza, 2021**

STRESZCZENIE

W artykule przybliżona została publikacja Pawła Kleki pt. *Trafność pomiaru w praktyce psychometrycznej. Badania empiryczne* oraz podjęta dyskusja z kilkoma zaprezentowanymi w niej stanowiskami. Uczyniono to w kontekście nakreślonego wprzód zagadnienia niepewności towarzyszących praktyce pomiaru i redukowania ich kosztem podnoszenia niepewności wyników obserwacji.

SŁOWA KLUCZOWE: psychometria, praktyka pomiaru, niepewność

Upominanie się o utrzymywanie nie byle jakiej relacji między tym, co jest mierzone a systemem reprezentacji oraz kontekstem osobowym i sytuacyjnym pomiaru wynika z niepewności wzbudzanej praktyką mierzenia i dotyczy nieredukowalnej niepewności wyników pomiarowych. Zbyt często w pomiarze cech psychologicznych i jego konstrukcji przewagę zyskują nawyki, bezrefleksyjne decyzje oraz brak troski o uzasadnienie uznania liczby za trafną reprezentację mierzonej cechy. W dodatku te nawyki i decyzje nierzadko podlegają środowiskowej legitymizacji udzielanej na mocy powszechności jakiegoś postępowania. To ewidentny błąd systemowy nauki. Pospieszna dążność w praktykach nauk społecznych do precyzowania obserwacji i uściślenia jej wyników, aby w precyzji i ścisłości dorównać *Science*, skutkuje naiwnym wyobrażeniem, że w odtwarzaniu struktury mierzonego zjawiska wystarczy posłużyć się na przykład numeryczną skalą szacunkową, przy pomocy której respondent przekłada wartości cechy na liczby. Użycie takiej skali szacunkowej miałoby uzasadniać zastosowanie szerokiego reper-

tuaru zaawansowanych metod matematycznej obróbki danych w celu odtwarzania i modelowania latentnych cech, w tym formułowania uprawnionych wniosków o strukturze lub dynamice badanego zjawiska. Przy okazji takich oczekiwań pomija się, że wybór przez respondenta pozycji na skali szacunkowej uzależniony jest od dynamiki złożonego układu interakcyjnych czynników, jak własności osoby i warunków, w jakich osoba funkcjonuje. Mówiąc krócej, niepewność w praktyce badawczej znoszona bywa kosztem podnoszenia niepewności wyników obserwacji, co zdarza się na skutek pośpiesznego i pobieżnego matematyzowania badanych zagadnień oraz odwoływania się do powszechnych rozwiązań bez oceny ich relewantności, w tym ocen ryzyka generowania artefaktów.

O takich kwestiach szerzej i twórczo traktuje monografia Pawła Kleki *Trafność pomiaru w praktyce psychometrycznej. Badania empiryczne*. Inspirowany tą monografią, a w niej przywołaniem słów uznanego psychometry Frederica Lorda: „Liczby nie wiedzą, skąd pochodzą”, sparafrazowałbym, że wartość liczbowa, nawet ta uzyskana najdoskonalszą metodą, potrzebuje być zinterpretowana w nurcie teorii zjawiska i potwierdzonego układu translacji struktury cechy na system reprezentacji. Jednak zamiast tego spotkać się można, jak zauważa Autor, z występującą w środowiskach naukowych presją na budowanie narracji wolnej od niepewności, wątpliwości i niejednoznacznych wyników badań.

Paweł Kleka przedstawia przemyślane stanowisko w kwestii modelowania cech i modelowania pomiaru oraz prezentuje wady i zalety przyjmowanych rozwiązań w walidacji instrumentów pomiaru. Prezentuje przy tym wyniki realizacji własnych projektów badawczych. Pierwszy, będący symulacją numeryczną, dotyczył roli introspekcji i wielkości próby w kształtowaniu wyników testu. Drugi projekt, oparty na systematycznym przeglądzie literatury, poświęcony został praktyce definiowania i badania trafności narzędzi badawczych. Trzeci dotyczył ustalenia zbieżności rezultatów analiz badań walidacyjnych opartych na konstrukcie psychologicznym z takimi, które takiego konstruktów są pozbawione.

Monografia Pawła Kleki nie jest opracowaniem podręcznikowym, lecz rozprawą metodologiczną charakteryzującą się świeżym spojrzeniem i dojrzałością. Spieszę jednak podkreślić, że nie tylko zaprawiony metodologicznie czytelnik odnajdzie pożytek i przyjemność w lekturze tej książki.

Na marginesie cennych dla mnie odczytań chciałbym sformułować kilka uwag wobec wybranych fragmentów przywołanej tu monografii.

Po pierwsze, jej Autor zwraca uwagę na naiwne zawieranie metodom statystycznym w procesie odtwarzania struktury cechy, które to znajdują się w stan-

dardowym zestawie praktyka, takim jak: analiza czynnikowa, współczynniki zgodności wewnętrznej testu oraz analiza współzmienności implementowana w celu ustalania zbieżności rozkładu wyników testu z zewnętrznym kryterium. Oczywiście nie w metodach upatrywany jest problem, lecz w sposobach ich użytkowania oraz warunkach ustalania struktury cechy. Zarazem jednak spostrzegłem stanowisko zawężające pojęcie pomiaru do aktów dotyczących cech ilościowych, a w szerszych ujęciach upatrujące źródło nietrafności pomiarów. Uważam, że to raczej szcążkowe odwoływanie się do teorii odwzorowań oraz błędy w pobieraniu prób ze zbiorowości generalnych są źródłem większości uchybień metodycznych i fiaska praktyki psychometrycznej niż brak możliwości dobrania analogicznych struktur, a tym samym systemów algebraicznych, dla psychologicznych konstruktów. Wprawdzie to ostatnie wymaga angażujących procedur, w tym eksperymentowania uzasadniającego konstrukt, co też zdecydowanie rekomenduje Paweł Kleka. Jednak zawężanie pojęcia pomiaru budzi moją wątpliwość, tym bardziej, gdy pomyślę o pomiarze lingwistycznym oraz pomiarze rozmytym. Jeśli eksperymentowanie pozwala ustalać manifestacje latentnej cechy, to na podstawie uzyskanych wyników i wykorzystaniu teorii odwzorowań możliwe jest dobranie systemu reprezentacji o strukturze zbieżnej ze strukturą mezurandu. Następnie dysponując już zaledwie ustaleniem pojedynczej postaci manifestacji konstruktów, można wprowadzić liczbę będącą miarą na zbiorze tych manifestacji. Taką miarą może być prawdopodobieństwo, tym bardziej że konstrukty psychologiczne łatwo uznać za wielkości losowe. Najprostszym rozwiązaniem jest w tym zakresie rachowanie częstości względnej wg powszechnie znanego wzoru na prawdopodobieństwo definiowane jako frekwencyjne. Warunek przeliczalności, którego spełnienia Autor wydaje się odmawiać przynajmniej części z konstruktów, jest wówczas zachowany. Nawet w sytuacji poważnego ograniczenia wiedzy o zjawisku, jak to polegające na znajomości tylko jednej postaci manifestacji danego konstruktów, nadal można zachować warunek przeliczalności zbioru manifestacji i określać miarę, jaką jest prawdopodobieństwo, np. w schemacie Bernoulliego. To natomiast umożliwia modelowanie struktury cechy przynajmniej przy pomocy takich rozkładów wielkości losowej, jak zero-jedynkowy, dwumianowy czy równomierny skokowy. Poruszam tu kwestie podstawowe, więc przypuszczam, że źródłem moich wątpliwości jest raczej sposób odczytywania narracji monografii.

Drugi komentarz dotyczy zaprezentowanego w monografii jako trzeciego z kolei badania nad rolą konstruktów dla wyników pomiaru testem. W badaniu tym zastosowano test GSES J. Juczyńskiego do oceny przekonań na temat samoskuteczności w wersji oryginalnej (2001) i trzech wersjach zmodyfikowanych w celu redukcji sensu pozycji testowych GSES, a tym samym zniesienia

konstruktu. Sądzę, że ta procedura dostarcza argumentów raczej na rzecz tezy, że rozumienie testu przez jego autora lub użytkownika bywa różne od rozumienia respondentów. Trudniej jest ją uznać za procedurę obrazującą artefaktyczną zbieżność wersji, która to zbieżność zachodzi wskutek zastosowanych metod analizy danych. Otóż w tak zaprojektowanym badaniu najprawdopodobniej nie kontrolowano procesu interpretacji dokonywanej przez respondentów, więc nie można wykluczyć, że respondent i tak starał się nadawać sens bezsensownemu twierdzeniu. Ten argument zyskuje na mocy, gdy dodatkowo wziąć pod uwagę, iż respondenci mogli pochodzić z populacji studentów lub nauczycieli psychologii i innych nauk społecznych (Autor nie scharakteryzował bliżej populacji i pobranej z niej próby), a więc przystępujących do pomiaru z założeniem, że test coś jednak mierzy, mimo iż jego pozycje mogą być zupełnie niezrozumiałe. Nazwa „samoskuteczność” dookreślająca test mogłaby wzmacniać to oczekiwanie względem twierdzeń utworzonych zarówno z bezsensownych zdań, jak i wyrazów powstałych z losowo wygenerowanych liter. Tym samym trudniej uznać konkluzję, że „wariancja wyjaśniona przez rozwiązania czynnikowe, nie wyjaśnia niczego” i że wersja testu z bezsensownymi twierdzeniami „...niczego nie mierzy!”. To „niczego” odnosi się w istocie do procesu uruchamianego interakcją postaci pytania, skojarzeń, jakie ono wzbudziło u respondenta oraz jego oczekiwań wobec testu. Zastosowanie racjonalnych kryteriów redukcji sensu nie oznacza, że test zastosowany wobec *homo interpretans* (Michel, 2019) dostarczy danych nomologicznie nietrafnych i pozbawionych sensu w sposób zobiektywizowany. Sens tych danych w tak zaplanowanej procedurze z udziałem respondentów doszukujących się znaczeń w dowolnych strukturach jest raczej ukryty, nieodtworzalny albo przynajmniej niemożliwy do wykluczenia. Oczywiście, może też być tak, że o strukturze odpowiedzi decyduje styl ich udzielania w warunkach niepewności. Zresztą Paweł Kleka (2021, s. 99) zdaje się sugerować podobną interpretację, pisząc o czynnikach kontekstowych silniejszych od czynników postulowanych teorią mezurandu.

Osobno moją uwagę zwróciło rozumowanie, z którego wyprowadzono konkluzję, że wątpliwe jest oczekiwanie, iż „występowanie cech świadczy na rzecz konstruktu” psychologicznego (Kleka, 2021, s. 105–106). Rozumowanie to przyjęło następującą formułę: Jeżeli [konstrukty psychologiczne objawiają się w postaci cech] i [cechy te uzewnętrzniają się w zachowaniach] i [ludzie potrafią obserwować, rejestrować i komunikować swe zachowania], to [pytając ludzi o zachowania, można ustalić występowanie konstruktu psychologicznego].

Taka formuła faktycznie kwalifikuje wniosek jako zawodny, ale pod tym względem nie różni się ono od większości rozumowań występujących w naukach empirycznych. Zgoda również co do spostrzeżenia Autora, że występo-

wać mogą okoliczności ograniczające zachodzenie stanów rzeczy opisywanych przez przesłanki. Jednak jeśli poprawić pod względem formalnym strukturę tego rozumowania, przekształcając ją na przykład według postaci *modus tollendo tollens*, to tym trudniej przyznać słuszność Autorowi, że wniosek pozostaje w stosunku sprzeczności do przesłanek. Ponadto zawodność wnioskowania z następstwa o poprzedniku nie wyklucza użyteczności rozumowania, a tak zdaje się stawiać sprawę Autor, gdy pisze o relacji konstruktów i cechy. Nawet jeśli przyjąć, że w tym fragmencie monografii bardziej niż o formalną stronę rozumowania chodziło o niezachodzenie izomorficznych (szerzej: homomorficznych) odwzorowań struktury wielkości latentnej na strukturę wielkości obserwowalnej, to podkreślić warto, iż w poznaniu racjonalnym dopuszcza się idealizację oraz stosuje warunek *ceteris paribus*. Kluczowa rola przypada przybliżeniom, a nie doskonałym reprezentacjom. Zresztą pomiar nie jest w stanie obejść się bez idealizacji w zakresie na przykład tego, że instrument charakteryzuje się odpowiednią dokładnością oraz warunkiem *ceteris paribus*, gdy chodzi o niesystematyczne oddziaływania na wartość zaobserwowaną w akcie pomiaru. Teoria błędów pomiarowych wsparta jest na obu.

Swoją wypowiedź zakończę opinią. Otóż monografia Pawła Kleki wycelowana jest w deficyty świadomości metodologicznej, brak troski o uzasadnianie przyjmowanych rozwiązań, oraz w poleganie na środowiskowo legitymizowanych praktykach badawczych zamiast na refleksyjnym dobieraniu środków poznania i postępowaniu zgodnie z teorią aplikowanych metod. Podobną pod tym względem publikacją jest monografia Wiesława Szymczaka *Praktyka wnioskowania statystycznego* (2019), do której odniosłem się w innym miejscu (Pasikowski, 2020). Obie te monografie, wzięte razem, obnażają naiwność praktyk pomiaru i analizy danych w naukach społecznych oraz demystyfikują szereg wyobrażeń na temat metodologicznych standardów. Warto mieć te monografie w pamięci, a nie tylko na półce lub w zasobach cyfrowych.

BIBLIOGRAFIA

- Kleka, P. (2021). *Trafność pomiaru w praktyce psychometrycznej. Badania empiryczne*. Wydawnictwo Nauk Humanistycznych i Społecznych UAM.
- Michel, J. (2019). *Homo Interpretans: Towards a Transformation of Hermeneutics*. Rowman & Littlefield Publishing Group.
- Pasikowski, S. (2020). O myśleniu statystycznym w pedagogice. Na marginesie odczytań książki Wiesława Szymczaka pt. „Praktyka wnioskowania statystycznego”. *Rocznik Pedagogiczny*, 43, 252–261.
- Szymczak, W. (2019). *Praktyka wnioskowania statystycznego*. Wydawnictwo Uniwersytetu Łódzkiego.

SUMMARY

**The Uncertainties of Psychometric Practice.
In the Readings of Paweł Kleka's Book,
Measurement accuracy in psychometric practice. Empirical studies.
Poznań: Wydawnictwo Nauk Humanistycznych
i Społecznych UAM, 2021**

The article presents a publication by Paweł Kleka entitled *Measurement accuracy in psychometric practice. Empirical studies* and a discussion with several positions presented in it. This was done in the context of the issue of uncertainties accompanying the practice of measurement, outlined earlier, and reducing them at the expense of raising the uncertainty of observational results.

KEY WORDS: psychometrics, measurement practice, uncertainty