



Krzysztof Rubacha

Standardy metodologiczne w zbieraniu i analizie danych. Pracownia Narzędzi Badawczych Komitetu Nauk Pedagogicznych PAN

DOI: <http://dx.doi.org/10.12775/PBE.2013.003>

Jedną z ważniejszych miar wiarygodności wyjaśnień naukowych jest trafność wewnętrzna schematu badania. We wszystkich schematach, czyli ilościowych i jakościowych, jednym ze wskaźników trafności wewnętrznej jest trafność narzędzi badawczych. W strategii ilościowej jej szacowanie poprzedza zbieranie danych. W strategii jakościowej można trafność narzędzi zweryfikować w trakcie zbierania danych oraz w trakcie ich analizy, posługując się różnymi rodzajami triangulacji. Biorąc pod uwagę czasochłonność badań nad trafnością narzędzi, związaną z wymaganiami metodologicznymi, można zauważyć, że trafność wewnętrzna badania rośnie wraz z posługiwaniem się przez badaczy zweryfikowanymi w innych badaniach narzędziami zbierania danych. Próby doraźnego budowania autorskich narzędzi pomiaru – z uwagi na brak czasu oraz kompetencje psychometryczne ich twórców – przynoszą wyniki, których trafność nie jest zweryfikowana. Można sobie jednak wyobrazić, że dysponujemy bazą wystandaryzowanych i zweryfikowanych w wielu badaniach narzędzi służących do pomiaru konkretnych zmiennych. Korzystanie z tak przygotowanych narzędzi zdecydowanie podnosi trafność wewnętrzną badania, a także dostarcza danych do ich kolejnych modyfikacji. W rezultacie dyscyplina systemowo wytwarzająca profesjonalne narzędzia badań podnosi własne standardy metodologiczne. W pedagogicznych badaniach nad praktykami edukacyjnymi

obserwujemy duże rozproszenie narzędzi badawczych, które nie były, na ogół, poddawane standaryzacji. Czytając raporty z badań, można zauważyć, że badacze zaczynają od budowania własnego narzędzia, w zdecydowanej większości doraźnego, którego nie poddają empirycznej weryfikacji. Nie szukają nawet wcześniejszych operacjonalizacji badanych zmiennych, tłumacząc to brakiem odpowiednich źródeł. I najczęściej nie mają innego wyjścia, ponieważ nie dysponujemy zorganizowaną bazą i dystrybucją empirycznie zweryfikowanych narzędzi badawczych. Rozumiejąc konsekwencje tego stanu rzeczy (setki badań wykonanych niewiarygodnymi, w sensie metodologicznym, narzędziami), Komitet Nauk Pedagogicznych PAN w czerwcu 2012 r., odpowiadając na inicjatywę autora tego tekstu, powołał do życia Pracownię Narzędzi Badawczych (PNB KNP PAN), obejmując ją merytorycznym patronatem. Misją Pracowni jest:

- opracowywanie, autoryzowanie i przygotowywanie do wydania podręczników narzędzi badawczych w badaniach nad edukacją,
- tworzenie bazy zweryfikowanych empirycznie narzędzi badań,
- budowanie standardów metodologicznych i etycznych konstruowania oraz użytkowania narzędzi badań,
- tworzenie środowiska badaczy zaangażowanych w realizację standardów metodologicznych w badaniach nad edukacją.

Opracowane narzędzia będą publikowane w „Przeglądzie Badań Edukacyjnych”, w dziale „Metody zbierania i analizy danych w badaniach edukacyjnych”, wraz z informacją o autoryzacji Pracowni Narzędzi Badawczych KNP PAN. Zakupienie danego numeru czasopisma będzie jednoczesnym nabyciem prawa do posługiwania się narzędziem w praktyce badawczej. Zgłoszone do publikacji podręczniki będą recenzowane przez członków Rady Naukowej Pracowni oraz specjalistów przez nich wskazanych. Prawa autorskie twórców narzędzi będą respektowane przez Wydawnictwo Naukowe UMK i chronione na podstawie umowy licencyjnej.

Pracownia jest otwarta na empiryczną weryfikację narzędzi badań stworzonych zarówno w strategii ilościowej, jak i jakościowej. Kierując się standardami metodologicznymi i psychometrycznymi, przyjmujemy, że zgłaszane do publikacji podręczniki będą zbudowane według struktury pozwalającej poznać wszystkie właściwości narzędzia, dostępne ze względu na jego charakter. Poniżej znajduje się ogólna analiza kryteriów konstruowania narzędzi badawczych. Jest ona opracowana dla testów, których wymagania maksymalizują te kryteria. W przypadku mniej wymagających narzędzi definicje tych kryteriów będą ulegały złagodzeniu.

Podstawy teoretyczne

Każde narzędzie stanowi operacjonalizację pojęcia umiejscowionego w konkretnej teorii. Jej zasięg, język, rodzaj twierdzeń, rodowód paradygmatyczny wyznaczają możliwości problematyzowania rzeczywistości edukacyjnej oraz interpretacji wyników badań. Dlatego w pierwszej części podręcznika należy zamieścić analizy teoretyczne leżące u podstaw definiowania operacjonalizowanej zmiennej. Szczególnie ważne jest pokazanie zakresu wyjaśniania zmiennej i rodzaju owego wyjaśnienia, które można uzyskać przez przełożenie zmiennej z języka teoretycznego na empiryczny. Sam proces przekładania zmiennej na wskaźniki powinien być wyraźnie wyodrębniony za pomocą zdefiniowania zakresu doświadczenia badanych, do których narzędzie będzie się odwoływać. Zmienne złożone strukturalnie powinny mieć zdefiniowane i zoperacjonalizowane wszystkie komponenty oraz ich wzajemne zależności. W przypadku testów typu papier–ołówki jest to najczęściej tworzenie itemów, zawierających poszczególne wskaźniki. W przypadkach narzędzi opartych na pomiarze behawioralnym jest to z kolei tworzenie sytuacji testujących określoną aktywność badanych osób. W przypadku natomiast narzędzi jakościowych, np. wywiadów skoncentrowanych na teorii subiektywnej czy na materiale, jest to układ pytań i zadań, z jakimi będzie konfrontowana osoba badana.

Ostatecznie ta część podręcznika powinna doprowadzić do prezentacji zoperacjonalizowanej postaci zmiennej badania, np. wersji eksperymentalnej kwestionariusza testu. Biorąc pod uwagę, że dalsze losy weryfikacji empirycznej mogą wyeliminować część pozycji testu, warto konstruować ich więcej, tak by te, które pozostaną, np. po analizie mocy dyskryminacyjnej, proporcjonalnie reprezentowały i wyczerpywały elementy strukturalne zmiennej.

Analiza mocy dyskryminacyjnej

Moc dyskryminacyjna pozycji narzędzia badawczego to stopień, w jakim każda pozycja różnicuje badaną próbkę losową. Pozycje nieróżnicujące w stopniu dostatecznym, w zasadzie mierzą cechę stałą zamiast zmiennej i muszą być usunięte z puli eksperymentalnej testu. W takim przypadku konieczne jest sprawdzenie, czy pozycje pozostałe po tym etapie weryfikacji narzędzia pokrywają się z wyjściowym konstruktem teoretycznym. Do analizy mocy dyskryminacyjnej najczęściej używa się abaków ze współczynnikami korelacji punktowo czteropolowej – tablica Jürgensena (Rubacha 2011, s. 386), dwuseryjnej, punktowo dwuseryjnej – tablice Flanagan (Rubacha 2011, s. 382) Dla każdego

ze współczynników istnieje umowna dolna wartość jego akceptacji. Może zachodzić również korelacja pomiędzy wysokością współczynnika a liczbą N , co trzeba brać pod uwagę, dokonując obliczeń. Ostatecznie ten etap analiz powinien wstępnie zweryfikować pulę pozycji w wersji eksperymentalnej testu (zob. Brzeziński 2006, s. 507).

Rzetelność narzędzia

Twierdzenie teorii Gulliksen (1950, s. 28) opiera się na tym, że wynik zaobserwowany jest sumą wyniku prawdziwego i błędu pomiaru. Wynik prawdziwy jest ujmowany jako idealny wynik osoby badanej w idealnych warunkach, wśród których jest idealne narzędzie badawcze. Takie badanie jednak nie istnieje, a konstruktorzy testu, przez kolejne próby równoległe, mogą minimalizować wariancję błędu. I właśnie rzetelność jest miarą dokładności, z jaką narzędzie dokonuje pomiaru. Do obliczeń wykorzystuje się współczynniki mocy dyskryminacyjnej. Estymowanie rzetelności jest wykonywane wobec pozycji testu, wyniku ogólnego lub wyniku w częściach testu. Najczęściej stosuje się metody korelujące pozycje testowe z wynikiem ogólnym, np. wzór Spearmana-Browna z poprawką Gaylorda. Często także szacowanie rzetelności opiera się na analizie statystycznej pozycji wskaźników (pozycji testowych). Są to analizy przy pomocy współczynników alfa Cronbacha, Hoyta, Kudera-Richardsona, Brzezińskiego (por. Brzeziński 2006, s. 467). Rzadziej natomiast, bo jest to metoda dość zawodna, stosuje się współczynnik W-Kendalla jako metodę obliczania zgodności sędziów kompetentnych, nadających rangi poszczególnym pozycjom narzędzia. Wymagania dotyczące wysokości współczynnika rzetelności są ostre. Jako minimalny wskaźnik dopuszcza się 0,75, lecz nie wtedy, gdy diagnoza wymaga wyprowadzenia wyników w skalach. W takich przypadkach bierze się pod uwagę wartości powyżej 0,9.

Trafność narzędzia

Trafność jest odpowiedzią na pytanie, czy narzędzie mierzy to, co ma mierzyć: w sensie diagnostycznym i w sensie teoretycznym. Trafność diagnostyczna jest szacowana najczęściej przy pomocy kryterium zewnętrznego, które – jednak – musi mieć potwierdzoną trafność. Wskaźnikiem trafności diagnostycznej jest wartość współczynnika korelacji pomiędzy narzędziem i kryterium. Trafność teoretyczna z kolei jest weryfikowana przy pomocy analizy skupień (por. Noworol 1989) i/lub analizy czynnikowej. Drogi są dwie. Pierwsza jest realizowana

wtedy, gdy na etapie analizy podstaw teoretycznych dysponujemy konstruktorem wymagającym weryfikacji. W takim przypadku posiadamy punkt odniesienia do interpretacji powstałych w trakcie analizy czynników lub skupień. Sprawdzamy wówczas, czy pozycje o wysokich ładunkach czynnikowych odtwarzają jakąś konkretną strukturę teoretyczną, tę, która występuje w roli konstruktów hipotetycznego. Często te konstrukty – w przypadku założonych zmiennych – nie są homogeniczne, ale zbudowane na kilku poziomach o zróżnicowanej heterogeniczności. W wyniku analizy można znaleźć wskaźniki niższego rzędu – homogeniczne, które tworząc czynnik, budują kolejny poziom konstruktów. W drugim przypadku, gdy nie dysponujemy wyraźnie zdefiniowanymi konstruktami, wykonujemy analizę czynnikową, by te konstrukty zbudować jako konstrukty empiryczne. Bardzo często wyodrębnione czynniki tworzą skale testu. Nie zawsze jednak zmienna ma złożoną strukturę teoretyczną, a przynajmniej nie tak, by na podstawie analizy czynnikowej tworzyć skale testu. W takich sytuacjach zweryfikowany układ czynników po prostu potwierdza trafność teoretyczną narzędzia. Obok tych dwóch rodzajów trafności możemy testować także trafność prognostyczną, korelując ze sobą dwa pomiary oddalone w czasie.

Trafność jest kluczowa dla wartości narzędzia, ponieważ upewnia jego użytkowników, że mierzy wyjściowy konstrukt. Jest zresztą bezpośrednim kryterium trafności wewnętrznej całego badania. Nietrafne narzędzie przekreśla wartość wniosków, które z badania wyprowadzamy, a nieznaną trafność sprawia, że nie potrafimy oszacować, czy wnioski mają wartości wyjaśniającą. Warto też pamiętać, że trafne narzędzie jest narzędziem rzetelnym, a rzetelne narzędzie nie musi być trafne. Możemy więc bardzo dokładnie zmierzyć konstrukt, którego nie chcieliśmy zmierzyć.

Normalizacja narzędzia

W sytuacji, gdy wynik pomiaru chcemy interpretować w kategoriach ilościowych, musimy odnieść go do kryterium empirycznego, którym jest rozkład z populacji generalnej. Poszukujemy więc wyniku typowego w danej populacji. Narzędziem interpretacji wyniku surowego jest właśnie normalizacja. Do analiz potrzebne są duże próbki losowe, zazwyczaj $N > 300$, których wyniki zostaną przekształcone na skale interwałowe: stenową, centylową, tetronową. Dzięki tym skalom możemy zbudować tymczasowe przedziały dla wyników surowych interpretowanych jako wysokie (np. 7–10 sten), przeciętne (5–6 sten) i niskie (1–4 sten). Jeśli nie mamy warunków do takich obliczeń, przydatną miarą okazuje się średnia i odchylenie standardowe, które pozwalają na stworzenie rów-

niez trzech przedziałów wyników. Normalizowanie narzędzia jest konieczne, gdy ma ono charakter diagnostyczny lub gdy pomiar jest sensowny tylko wtedy, jeżeli jesteśmy w stanie określić poziomy zmiennej.

Administracja narzędzia

W tej części podręcznika powinny pojawić się informacje o zweryfikowanych sposobach zbierania danych, o czasie potrzebnym badanym na wypełnienia narzędzia, o instrukcji, jaką muszą otrzymać badani, o zachowaniu się badaczy podczas zbierania danych. Warto także, kierując się wcześniej obliczonymi normami, określić wyraźnie, jakie populacje mogą być badane za pomocą testu, a na jakich narzędzie nie było weryfikowane. I wreszcie, wracając do podstaw teoretycznych, należy napisać o interpretacji wyników w poszczególnych przedziałach. Podręcznik powinien kończyć się bibliografią, zawierającą wszystkie publikacje, w których były wcześniej prezentowane wyniki badań prowadzonych przy pomocy weryfikowanego narzędzia. Podręcznik powinien też zawierać arkusz testu oraz klucz.

Pokazane kryteria konstruowania podręczników są koniecznym minimum dla testów i innych narzędzi pomiarów ilościowych. Każdy jednak badacz indywidualnie decyduje o zakresie weryfikacji i użytych metodach analiz, respektujących uwarunkowania statystyczne i teoretyczne. W przypadku pomiarów jakościowych kluczowym wydaje się kryterium trafności, którą można zweryfikować przy pomocy triangulacji badaczy, metod, źródeł, teorii. Weryfikacja narzędzi badawczych w strategii jakościowej dotyczy, oczywiście, narzędzi, operacjonalizujących konkretne ramy pojęciowe badań, opartych na powtarzalnym algorytmie zbierania i analizowania danych (por. Stemplewska-Żakowicz 2009, s. 179–197).

Bibliografia

- Angosina M. (2010), *Badania etnograficzne i obserwacyjne*, WN PWN, Warszawa.
- Brzeziński J. (red.) (1988), *Problemy teorii, rzetelności, konstrukcji i analizy wyników testów psychologicznych*, PTP, Warszawa.
- Brzeziński J. (2006), *Metodologia badań psychologicznych*, WN PWN, Warszawa.
- Coen L., Manion L., Morrison K. (2005), *Research methods in education*, RoutledgeFalmer, London–New York.

- Ferguson G. A., Takane Y. (2003), *Analiza statystyczna w psychologii i pedagogice*, WN PWN, Warszawa.
- Gulliksen H. (1950), *Theory of Mental Tests*, Wiley, New York.
- Noworol Cz. (1984), *Analiza skupień w badaniach empirycznych. Rozmyte modele hierarchiczne*, PWN, Warszawa.
- Ripley T. (2010), *Analiza konwersacji, dyskursu, dokumentów*, WN PWN, Warszawa.
- Rubacha K. (2011), *Metodologia badań nad edukacją*, Oficyna Wydawnicza Łośgraf, Warszawa.
- Stemplewska-Żakowicz K. (2009), *Analiza dyskursu jako metoda analizy danych z wywiadu*, w: K. Stemplewska-Żakowicz, K. Krejtz (red.), *Wywiad psychologiczny 1. Wywiad jako postępowanie badawcze*, PTP, Warszawa.

Methodological standards in data collection and analysis Research Tools Workshop of The Committee on Pedagogical Sciences of Polish Academy of Sciences (PAN)

Summary

The article refers methodological problems of construction of a research tools in a social sciences. The author especially concerns to the discrimination analysis, reliability, accuracy and the normalization of a research tools.