

Edynburg, 15.09.2022

dr hab. Oleg Gorbaniuk
Instytut Psychologii
Wydział Nauk Społecznych
Katolicki Uniwersytet Lubelski Jana Pawła II

**Opinia dotycząca rozprawy habilitacyjnej i dorobku naukowego dra Pawła Kleki,
opracowana w ramach przewodu habilitacyjnego procedowanego
przez Radę Dyscypliny Psychologia Uniwersytetu Adama Mickiewicza w Poznaniu**

1. Informacje ogólne

Dr Paweł Kleka jest aktualnie pracownikiem Wydziału Psychologii i Kognitywistyki Uniwersytetu Adama Mickiewicza w Poznaniu, w którym w 2001 roku uzyskał stopień magistra psychologii, a następnie w 2013 roku stopień doktora nauk humanistycznych.

2. Ocena głównego osiągnięcia naukowego

Jako główne osiągnięcie naukowe, stanowiące podstawę wniosku habilitacyjnego, dr Paweł Kleka wskazał monografię pt. „Trafność pomiaru w praktyce psychometrycznej. Badanie empiryczne” (124 strony). Problem trafności pomiaru przy użyciu narzędzi kwestionariuszowych, którego podjął się Habilitant, jest jednym z kluczowych, jeżeli nie najważniejszych w psychologii. Trafność pomiaru determinuje jakość kolejnych etapów procesu badawczego, który bazuje na zebranych danych. Zatem waga problemu badawczego jest adekwatna do rangi przewodu habilitacyjnego. Specyfika metodologiczna problemu sprawia jednak, że nie jest łatwo w tym temacie o wygenerowanie obiektywnie nowej wartości dodanej. Praca składa się z czterech rozdziałów, każdy z których realizuje inny cel, tworząc na poziomie celów logiczną a zarazem różnorodną/bogatą całość. W pierwszym rozdziale Habilitant przedstawił problematykę pomiaru, w szczególności krytykę klasycznego podejścia oraz analizę warunków, których spełnienie pozwala na pomiar ilościowy konstruktów psychologicznych. Kolejne trzy rozdziały - po sformułowaniu problemu badawczego - zawierają badania/analizy własne Habilitanta. Cechą wyróżniającą tych analiz jest odmienny charakter danych stanowiących ich podstawę. Są one odpowiednio efektem symulacji

komputerowej, metaanalizy publikacji oraz badań eksperymentalnych. Świadczy to z jednej strony o przemyślanej koncepcji pracy, z drugiej o kompetencjach analitycznych Autora.

Faktycznym celem badań przedstawionych w drugim rozdziale – przynajmniej tak, jak to odczytuję po zapoznaniu się z jego wynikami, było przetestowanie wrażliwości wskaźników trafności na trzy rodzaje deformacji danych w stosunku do danych początkowych na bazie sztucznie wygenerowanych danych. Przedstawione wyniki symulacji zwiększają u czytelnika (szczególnie mniej obeznanego z naturą matematyczną tej statystyki) świadomość ograniczeń współczynnika korelacji liniowej jako potencjalnej miary trafności. Zobowiązuje też do refleksji nad alternatywną interpretacją wysokości współczynnika korelacji i jego weryfikacji poprzez zastosowanie innych narzędzi statystycznych. Stanowi podstawę do generowania hipotez odnośnie możliwych systematycznych deformacji zebranych danych (w stosunku do rzeczywistego nasilenia zmiennej ukrytej) pod wpływem różnych czynników po stronie respondenta i warunków badań.

Problem wpływu samowiedzy / zdolności do introspekcji na jakość udzielanych odpowiedzi i trafność narzędzia sam w sobie jest ważny z punktu widzenia metodologicznego. Jednak zapowiedź jego zgłębienia w postaci tytułu rozdziału jest nieadekwatna w stosunku do treści, która opiera się na sztucznych danych, jedynie symulujących różne możliwe sytuacje, a sama symulacja bazuje na czysto hipotetycznych przewidywaniach, jak zdolność do introspekcji może zniekształcać zebrane dane. Należy przypuszczać, że może istnieć wiele zmiennych psychologicznych, które mogą powodować podobne zniekształcenia danych, jakie hipotetycznie były przypisane zdolności do introspekcji. Powstaje zatem problem trafności operacjonalizacji (jej specyficzności dla badanej zmiennej), czyli czegoś, co sam Autor niejednokrotnie krytykuje. Biorąc pod uwagę specyfikę użytych danych i wykonanych analiz bardziej adekwatne byłoby odwrócić kierunek rozumowania: przetestowano kilka możliwych deformacji zebranych danych w stosunku do prawdziwych, których wystąpienie może zmniejszać trafność narzędzia i na które tradycyjne wskaźniki trafności mogą być niewystarczająco wrażliwe. Jednym z wielu czynników, który może powodować takie deformacje, jest poziom samowiedzy lub zdolność do introspekcji. Tu przydałby się porządny a zarazem zwięzły przegląd literatury przedmiotu uzasadniający postulowany charakter deformacji danych przez samowiedzę/introspekcję. Przeprowadzenie własnych badań empirycznych (eksperymentalnych) i zebranie realnych danych poprzedzony systematycznym przeglądem literatury na temat deformującego wpływu czynników sprzężonych z metodami

samoopisowymi na wyniki badań psychologicznych byłoby znaczącym posunięciem uzupełniającym zaprezentowane symulacje. Podsumowując drugi rozdział można stwierdzić, że zawiera interesujące symulacje komputerowe pokazujące ograniczenia często stosowanego wskaźnika trafności, ale obietnica dostarczenia wiedzy na temat wpływu metod samoopisowych na wyniki badań psychologicznych wynikająca z tytułu rozdziału została niezrealizowana, zatem tytuł w mojej ocenie jest nieadekwatny do treści.

Trzeci rozdział pracy zawiera analizę i ocenę stanu wykorzystywanych wskaźników trafności na bazie metaanalizy recenzowanych publikacji w latach 1999-2018, dokonaną przez Habilitanta. Metaanaliza ta została poprzedzona interesującym przeglądem wskaźników trafności w aspekcie historycznym, który stanowi punkt odniesienia w ocenie wyników badań własnych. Pomysł zrobienia metaanalizy w podanym odcinku czasu jest godny pochwały, aczkolwiek jego realizacja budzi już zastrzeżenia natury metodologicznej. Analiza zebranych danych ma charakter opisowy i daje fragmentaryczny wgląd w analizowany odcinek czasu, ukazując dominację współczynnika korelacji i analizy czynnikowej jako głównych metod weryfikacji trafności narzędzi psychologicznych. Biorąc pod uwagę specyfikę analizowanego okresu, w którym komputery i programy statystyczne stały się szeroko dostępne i nastąpił skok kompetencyjny użytkowników, przydałaby się analiza dynamiki wzrostu/spadku wykorzystywania wskaźników trafności, co pozwoliłoby nie tylko na adekwatną ocenę przeszłości, lecz przewidzieć tendencję na najbliższe lata. Aby zwiększyć wiarygodność takich analiz należałoby sięgnąć po znacznie większą próbę publikacji (Habilitant uwzględnił tylko 101) zdecydowanie dbając jednocześnie ich reprezentatywność: w obecnej próbie 64% ukazało się w 2018 roku, co samo w sobie powinno wzbudzić wątpliwości Autora odnośnie reprezentatywności materiału; nie uwzględniono płatnych artykułów, które raczej reprezentują ponadprzeciętną jakość, a do których można łatwo dotrzeć np. zwracając się wprost do autorów. Z kolei w artykułach wykorzystujących CFA jako metodę oceny trafności pożądane byłoby wydzielić te, w których użyto AVE (*average variance extracted*).

Ogólnie w całej pracy Habilitant wyraża przekonanie, że dominujące aktualnie w badaniach psychologicznych wskaźniki trafności są mało wiarygodnymi jej miarami. Aby potwierdzić zasadność swojego krytycznego/sceptycznego nastawienia (wspieranego przez cytowanie metaanaliz literatury przedmiotu innych autorów) Habilitant zaprojektował badania opisane w rozdziale czwartym, które w znacznym stopniu sprzyjają potwierdzeniu takiego oczekiwania poprzez wybór metody badań najczęściej obniżającej jakość zbieranych danych:

(1) jako metodę zbierania danych wybrano ankietę internetową; (2) uzyskano bardzo niski stopień zwrotu wypełnionych kwestionariuszy (kluczową techniką, aby tego uniknąć, jest tu przypominanie); (3) kwestionariusz, którego treścią manipulowano, był jednowymiarowy (wielowymiarowy pozwoliłby na znacznie bogatsze analizy, szczególnie w kontekście wieloczynnikowej CFA); (4) pozycje kwestionariusza nie były wzajemnie odwrócone (wtedy np. błąd serii nie zawyżałby sztucznie współczynnika zgodności wewnętrznej); (5) respondenci byli zmuszani do udzielania odpowiedzi (nie było możliwości udzielenia odpowiedzi „nie wiem”, „trudno powiedzieć”, „nie dotyczy” - nie stanowiących przy tym bynajmniej środka skali odpowiedzi, których obecność znacznie obniża prawdopodobieństwo ustosunkowywania się do twierdzeń/pytań, co do których respondent nie ma zdania); (6) poprzez umieszczenie pytań metryczkowych na początku badania respondent, nie znając jeszcze treści pytań, był przynajmniej częściowo pozbawiany anonimowości, co wpływa na jego postawę wobec badań; (7) opis procedury badań nie wskazuje, aby zastosowana metoda badań (pojęcie szersze niż kwestionariusz/e) motywowała respondentów do rzetelnego wypełniania kwestionariuszy; (8) zabrakło definicji populacji, (9) liczebność próby nie była ustalona w oparciu o analizę mocy, (10) nie przedstawiono dowodów na spełnienie warunków II zasady randomizacji (szczególnie w kontekście przerwania wypełnienia kwestionariusza przez 1/4 respondentów), (11) analizowany jest efekt główny w układzie niezrównoważonym itd. Wypunktowałem ważniejsze słabe punkty i wady metodologiczne zrealizowanych przez Autora badań empirycznych. Oczywiście, takiej niskiej jakości procedury badań [2, 4-10] można nieraz spotkać w praktyce badawczej, ale trzeba mieć świadomość, że przez to wnioski z badań Habilitanta są niestety zawężone do badań o (bardzo) niskiej jakości metodologicznej. Z kolei Habilitant w swoich wnioskach/podsumowaniu do nich nie ogranicza się, traktując zrealizowane badania raczej jako standard w psychologii rozszerzając tym samym w sposób nieuzasadniony zakres generalizacji wniosków z badań własnych. Oczywiście problemy z walidacją występują również w badaniach lepiej zaprojektowanych, ale ich natężenie będzie wówczas prawdopodobnie niższe, a wówczas efekty manipulacji nie byłyby tak spektakularne. Potwierdzić tego typu przypuszczenia mogą jedynie badania. Warto odnotować, że w dyskusji wyników Habilitant nie przyjmuje postawy samokrytycznej i nie przedstawia w zasadzie ograniczeń badań własnych. Dyskusja tych ograniczeń, które naturalnie występują w każdym badaniach, mogłaby stanowić punkt wyjścia do sformułowania kolejnych pytań/hipotez badawczych, które pozwalałyby głębiej i szerzej zbadać problem trafności zwiększając zakres

generalizacji wniosków. W szczególności jednym z problemów, który należałoby zbadać, jest odpowiedź na pytanie, w jaki sposób jakość metody badań (operacjonalizacja, próba, procedura) wpływa na wskaźniki trafności i ich wartość informacyjną. W moim przekonaniu podjęty przez Habilitanta problem z uwagi na swój potencjał wymaga szerzej zaplanowanego programu badań empirycznych, szczególnie jeżeli wynik tych badań ma stanowić główne osiągnięcie Habilitanta i wносить znaczący wkład w psychologię/metodologię odpowiadający randze przewodu habilitacyjnego.

Bez względu na moją krytyczną ocenę badań przedstawionych w rozdziale czwartym trzeba zaznaczyć ponownie, że one zachowują swoją wartość naukową – w pewnym zakresie przedstawionych analiz – w odniesieniu do badań o niskiej jakości metodologicznej, w których autorzy mogą także stawiać sobie za cel walidację narzędzi psychologicznych. Badania eksperymentalne Habilitanta (z ww. zastrzeżeniem [10]) wyraźnie dowodzą, że przy metodzie badań o niskiej jakości możemy uzyskiwać przyzwoite klasyczne wskaźniki psychometryczne narzędzi, które to są pozbawione sensu. Prawdopodobnie zawdzięczamy to bezrefleksyjnemu zaznaczaniu tej samej odpowiedzi przez respondenta, który nie ma motywacji lub merytorycznego powodu (w przypadku bezsensownych pozycji) do ich różnicowania. Swoją drogą warto byłoby wykonać analizy umożliwiające wgląd w style zaznaczania, analizując profile odpowiedzi i porównując średnie wyniki w grupach eksperymentalnych, które stanowiłyby dowód, że respondenci „udzielają tych samych odpowiedzi” (s.97) bez względu na wymianę kluczowego pojęcia. Podobieństwo współczynników korelacji nie oznacza identyczności średnich (ale też nie wyklucza).

Recenzowana praca habilitacyjna objętościowo nie jest obszerna i z pewnością nie jest przeciążona przeglądem literatury przedmiotu ukazującego aktualny stan wiedzy w zakresie omawianych zagadnień. Osobiście w kilku miejscach, gdzie Autor wprost deklaruje, że nie zamierza rozwijać zagadnienia lub dokonywać przeglądu literatury, oczekiwałem takiego przeglądu. W różnych miejscach pracy (wprowadzeniu teoretycznym, interpretacji wyników badań a szczególnie w dyskusji/podsumowaniu wyników) możemy ponadto spotkać obszerne fragmenty, w których Habilitant dzieli się obserwacjami (w tym na bazie własnych badań) lub doświadczeniem ze swojej praktyki analitycznej/badawczej, z którymi osobiście często się zgadzam. Brakuje tu jednak odwołań do literatury, gdzie tego typu prawidłowości oparte na wynikach badań są dobrze opisane i usystematyzowane. Przykładem może być szereg opisywanych efektów związanych z procesem gromadzenia danych, które są typowe dla

różnych ankiet (czyli technik pośredniej komunikacji z respondentem), a które stanowią przedmiot systematycznych badań od co najmniej 40 lat i są bardzo dobrze opisane w literaturze przedmiotu poświęconej metodologii badań ankietowych. Habilitant trafnie wypunktowuje prawidłowości na przykładzie swoich badań, nie robi jednak odniesienia do literatury, przez co czytelnik nie jest w stanie ocenić, na ile są to obserwacje odkrywcze, a na ile są potwierdzeniem znanych od dawna prawidłowości. Może to sugerować niewystarczające rozeznanie w literaturze, która jest bardzo bogata na tematy poruszane, ale nie rozwijane przez Autora i w konsekwencji nie pozycjonowane w tejże literaturze. W moim przekonaniu Habilitant posiada w tym zakresie znacznie większy potencjał.

Ostatnia część pracy habilitacyjnej (zakończenie), zawiera szereg dojrzałych refleksji, będących z jednej strony pochodną problematyki analizowanej w rozdziale pierwszym, z drugiej zaś strony są one kontynuacją wniosków wynikających z uzyskanych wyników badań przez Habilitanta. W szczególności Autor proponuje niektóre rozwiązania zidentyfikowanych problemów związanych z pomiarem i badaniem trafności narzędzi psychologicznych. Te rozważania, podobnie jak treści zamieszczone we wcześniejszych rozdziałach, niewątpliwie wzbudzą refleksję nad cudzą i własną praktyką psychometryczną i dydaktyczną zainteresowanego czytelnika, zmuszając do bardziej krytycznej i powściągliwej postawy.

3. Ocena pozostałego dorobku naukowego

Dorobek towarzyszący Habilitanta obejmuje 43 publikacje naukowe, w tym 16 publikacji w czasopismach z listy JCR, 23 publikacje w czasopismach z listy B i C MNiSW oraz cztery rozdziały w pracach zbiorowych. Sumaryczny Impact Factor, liczba uzyskanych punktów wg MNiSW, indeks h są na przyzwoitym poziomie. W zdecydowanej większości przypadków Habilitant występuje w roli drugiego lub kolejnego autora (36 z 43), z kolei w roli pierwszego autora występuje w 5 artykułach opublikowanych w czasopismach i 2 rozdziałach w pracach zbiorowych. W publikacjach wieloautorskich, gdzie Habilitant nie występuje w roli pierwszego autora, tylko w ok. 15% przypadków były zamieszczone dane odnośnie wkładu każdego autora. W przypadku Habilitanta typową rolą było dbanie o stronę metodologiczną badań lub samej publikacji, w szczególności wykonanie analiz statystycznych i napisanie prawdopodobnie odpowiednich części artykułu. Trudno określić, na ile trafna jest generalizacja tej roli na pozostałe publikacje, w których nie określono wkładu autorów. Tematyka publikacji współautorskich jest bardzo szeroka, ale w znacznej części przypadków obejmowały one

opracowanie nowych testów lub adaptacja istniejących, gdzie rola Habilitanta, jak można przypuszczać, była kluczowa. Również publikacje, w których Habilitant jest pierwszym autorem, dotyczą w 4-5 z 7 przypadków problematyki metodologii badań i psychometrii. Ranga tych czasopism jest zróżnicowana, dwa najwyżej punktowane mają 70 i 140 pkt (wg punktacji MNiSW). Podsumowując, w ramach dorobku towarzyszącego Habilitant z jednej strony zdobył duże doświadczenie metodologiczne i psychometryczne, z drugiej strony dzięki jego kompetencjom w tym obszarze teksty skutecznie przeszły proces recenzji i ukazały się w wysoko punktowanych czasopismach reprezentujących różne dyscypliny naukowe.

W okresie po doktoracie Habilitant był autorem lub współautorem 29 prezentacji (w połowie przypadków jako pierwszy autor) na konferencjach odbywających w Polsce i 2 – na zagranicznych (wyłącznie w roli drugiego autora). Habilitant był także członkiem komitetu organizacyjnego 4 konferencji krajowych. O ile aktywność konferencyjną na rynku krajowym można uznać za wystarczającą, o tyle w przestrzeni międzynarodowej, szczególnie w kontekście tematyki pracy habilitacyjnej, aktywność jest niewystarczająca – głos Habilitanta nie jest obecny w tej formie na poziomie międzynarodowym.

W okresie po doktoracie nie występował w roli kierownika grantów, ale był członkiem zespołu badawczego realizującego m.in. granty NCN. Habilitant nie określił swojej roli od strony formalnej, ale od strony merytorycznej wykonane prace dotyczyły analiz statystycznych i psychometrycznych.

Habilitant był lub jest członkiem dwóch stowarzyszeń (krajowego i międzynarodowego) po 2013 roku. W tym okresie kilkanaście razy występował w roli recenzenta artykułów skierowanych czasopism polskich i zagranicznych. Kilkakrotnie wystąpił w roli eksperta, wykonując ekspertyzy z zakresu metodologii badań dla podmiotów zewnętrznych.

W okresie po doktoracie Habilitant systematycznie doskonalił swoje kompetencje poprzez zaoczny udział w warsztatach organizowanych przez czołowe zagraniczne uniwersytety (ukończył 12 kursów z zakresu m.in. metodologii, statystyki, e-learningu).

4. Dorobek dydaktyczny i organizacyjny

Habilitant ma wieloletnie doświadczenie w prowadzeniu różnych zajęć o charakterze metodologicznym, w tym z zakresu statystyki i psychometrii w UAM oraz Uniwersytecie Zielonogórskim dla studentów studiów magisterskich i doktoranckich. Od 2015 roku występuje

w roli promotora prac dyplomowych (psychologia i kognitywistyka), promując w tym okresie 12 prac licencjackich i 10 magisterskich.

W ramach UAM przez lata Habilitant pełnił funkcje pełnomocnika ds. e-learningu. Ogólnie dorobek organizacyjny Habilitanta w świetle dostarczonych przez niego materiałów jest dość skromny.

5. Konkluzja końcowa

Oceniając dorobek i aktywność naukową dr Pawła Kleki na podstawie kryteriów zawartych w art. 219 ust. 1 pkt. 2 Ustawy z dnia 20 lipca 2018 r. należy stwierdzić, że Habilitant posiada znaczącą część osiągnięć, które wymienia ustawa. Posiada on duże kompetencje w zakresie metodologii badań, w tym w zakresie analiz statystycznych i psychometrii. W wieloautorskich przedsięwzięciach Habilitant najczęściej zadowala się jednak drugoplanową rolą. Analiza dorobku badawczego i dydaktycznego wskazuje na wybór przez Habilitanta raczej ścieżki wąsko specjalizacyjnej w nauce. Posiadane kompetencje zostały udokumentowane/potwierdzone w postaci licznych publikacji i są przekazywane kolejnym pokoleniom studentów. W mniejszym stopniu te kompetencje zostały ujawnione w pracy habilitacyjnej w zakresie zrealizowanych przez Habilitanta badań własnych, w których można było pokusić się o bogatszy/szerszy program badawczy, aby w pełni wykorzystać posiadany potencjał. Oceniając główne osiągnięcie naukowe, mimo zgłoszonych uwag, uważam wciąż, że stanowi ono wystarczający ustawowo wymagany wkład Habilitanta w psychologię jako naukę. Wnoszę na tej podstawie o nadanie dr Pawłowi Klece stopnia naukowego doktora habilitowanego nauk społecznych w zakresie psychologii.

