UNIWERSYTET WARSZAWSKI

Wydział Psychologii

Warszawa, 27.10.2025

Prof. dr hab. Maciej Stolarski

Uniwersytet Warszawski

Wydział Psychologii

Recenzja rozprawy doktorskiej mgr. Tomasza Raka
pt. *„Weryfikacja jednowymiarowej struktury zmiennej latentnej w konfirmacyjnej analizie czynnikowej: analiza trafności wskaźników w badaniach symulacyjnych oraz empirycznych”*,
napisanej pod kierunkiem dr. hab. Jana Cieciucha, prof. ucz.

Praca doktorska mgr Tomasza Raka podejmuje ważny problem metodologiczny w obszarze psychometrii – ocenę przydatności klasycznych miar dopasowania w konfirmacyjnej analizie czynnikowej (CFA) do rozróżniania struktur jedno- i wielowymiarowych w sytuacji, gdy prawdziwa struktura jest jednowymiarowa. Tematyka rozprawy lokuje ją zatem na pograniczu psychologii różnic indywidualnych (pomiar konstruktu o charakterze różnicowym – cechy ilościowej), psychometrii i metodologii badań psychologicznych. Już w tym miejscu chciałbym podkreślić, że prace doktorskie w tym obszarze są nieporównywalnie rzadsze niż rozprawy, w których hipotezy i pytania badawcze dotyczą specyfiki zależności pomiędzy zmiennymi psychologicznymi czy też wpływu manipulacji eksperymentalnych na zachowania człowieka. Mamy tu zatem do czynienia z rozprawą o charakterze dość unikatowym, której napisanie świadczy o zainteresowaniu jej Autora zagadnieniami, które dla wielu badaczy-psychologów pozostają swoistą *terra incognita*, zaś dla innych, stanowią jedynie podstawę dla rozwoju narzędzi psychometrycznych, przez co są używane instrumentalnie, z minimalną dozą refleksji nad złożonością i – jak się okazuje – niejednoznacznością, lub wręcz niewystarczalnością powszechnie przyjmowanych założeń i metod.

Zadanie, które postawił przed sobą Doktorant, wydało mi się w pierwszej chwili dość proste (w sensie wykonalności, niekoniecznie łatwości przeprowadzenia analiz). Spodziewałem się bowiem, że przynajmniej kilka spośród „znanych i lubianych" miar dopasowania modelu w CFA pozwoli na dość jednoznaczne rozróżnienie pomiędzy miarami jedno- i nie jednoczynnikowymi. Tymczasem już po przeczytaniu streszczenia rozprawy zrozumiałem, że nic bardziej mylnego. Okazało się, że możliwości faktycznego rozróżnienia tych dwóch grup modeli są w istocie bardzo zaskakująco ograniczone, zaś osiągniecie tego celu z wykorzystaniem klasycznych miar dopasowania jest zasadniczo skazana na niepowodzenie. Z tym większym zainteresowaniem rozpocząłem lekturę zasadniczej części pracy.

Recenzowana dysertacja ma charakter monograficzny (choć w mojej ocenie zdecydowanie „zasługuje" – ze względu na jej jakość oraz przybliżoną poniżej wartość dla praktyki psychometrycznej – na publikację w formie artykułu; rzecz jasna po odpowiednim skróceniu). Praca zawiera standardowe elementy rozprawy doktorskiej: szczegółowy spis treści, streszczenia w języku polskim i angielskim, wprowadzenie teoretyczne, opis celów i problemu badawczego, dwa rozdziały empiryczne (badanie symulacyjne i badanie z danymi rzeczywistymi), dyskusję z podsumowaniem, rekomendacjami i ograniczeniami, obszerną bibliografię oraz suplement (tabele i załączniki). Układ ten jest przejrzysty i typowy dla prac empirycznych z zakresu psychologii, opartych o analizy o charakterze ilościowym. Objętość rozprawy (około 100 stron zasadniczego tekstu plus bibliografia i załączniki) mieści się w standardowych wymogach formułowanych względem monograficznych prac doktorskich. Język jest poprawny i spójny, nieliczne błędy mają zwykle charakter literówek i nie utrudniają w większym stopniu lektury tekstu (wyjątkiem, który wymagał ode mnie zatrzymania się podczas czytania pracy jest fragment ze s. 11: „różne aspekty cechy letenteji inólnie tworzą nadrzędne konstrukty"). Formatowanie tekstu jest zgodne ze standardami edytorskimi naukowych tekstów psychologicznych w języku polskim (na podstawie wytycznych APA 7). Nie wszystkie cytowane teksty, zapewne przez przeoczenie, znalazły się też w bibliografii końcowej (np. cytowany własny tekst Autora – Rak i Wrześniowski, 2024). Reasumując, struktura pracy, język i aspekty formalne spełniają normy stawiane rozprawom doktorskim.

Wartość naukowa i oryginalność pracy

Autor koncentruje się na uzupełnieniu dostrzeżonej przez siebie luki, związanej z ewidentnym brakiem obecności w literaturze jasnych wskazówek dotyczących możliwości wiarygodnego rozróżniania modelów jedno- i wielowymiarowych za pomocą klasycznych miar dopasowania. Zaproponowane przez Autora podejście jest nowatorskie, ponieważ dotychczas

stosowane w tym obszarze metody – porównania modeli w oparciu o wskaźniki dopasowania – były przez badaczy stosowane z założeniem, że taki rodzaj wnioskowania faktycznie pozwala na wykrycie „prawdziwej natury" mierzonej zmiennej. Tymczasem, jak dowodzi mgr Rak, w rzeczywistości dotychczas brakowało danych, które faktycznie potwierdzałyby, że tego rodzaju podejście prowadzi do poprawnych wniosków.

Autor formułuje trzy główne pytania badawcze (P1–P3) dotyczące rozróżniania modeli oraz progów decyzyjnych dla wskaźników. Połączenie rozbudowanej analizy symulacyjnej z solidną próbą empiryczną oraz uwzględnienie zarówno globalnych miar dopasowania (χ^2, CFI, RMSEA itp.), jak i lokalnych wskaźników modyfikacji (np. *nm%*) sprawia, że praca dostarcza rzetelnych danych, z których Doktorant wyciąga pogłębione i – w mojej ocenie – zasadne wnioski, istotne (choć niekoniecznie optymistyczne) dla praktyki badawczej i psychometrii. Niestety, uzyskane przezeń wyniki nie pozwalają na sformułowanie konkretnych rekomendacji (np. w postaci punktów odcięcia), które mogłyby pozwolić twórcom narzędzi psychometrycznych na uniknięcie „błądzenia we mgle", zwłaszcza w sytuacji, w której brak jasnych koncepcyjnych przesłanek dotyczących natury (struktury) badanej zmiennej, a empiryczne próby jej opisania w oparciu o powszechnie stosowane podejścia (np. porównania wskaźników χ^2/df) nie prowadzą do jednoznacznych konkluzji.

Recenzowana praca nie tylko w udany i rzetelny sposób systematyzuje istniejącą wiedzę (patrz obszerny przegląd miar dopasowania, który zdarzyło mi się już nawet wykorzystać jako źródło informacji przy wyborze wskaźników dopasowania podczas akurat prowadzonych przeze mnie analiz), lecz przede wszystkim dostarcza nowych wyników empirycznych i symulacyjnych, istotnych dla badań nad konstrukcją testów. W mojej ocenie dysertacja mgr. Raka wnosi istotny wkład do współczesnej psychometrii, dostarczając przekonujących danych, pozwalających na wybór metody analitycznej, którą twórcy narzędzi diagnostycznych mogą wykorzystać przy poszukiwaniu empirycznych argumentów za jedno- vs wieloczynnikową strukturą mierzonych przez nie konstruktów latentnych.

Poprawność metodologiczna i jakość analiz

Podejście badawcze zastosowane przez Doktoranta jest w mojej ocenie zasadne i adekwatne do postawionych celów. Badanie symulacyjne objęło aż 31 500 wygenerowanych zbiorów danych o znanej, jednoznacznie jednowymiarowej strukturze. Proces ten przeprowadzono za pomocą skryptu w R (udostępnionego przez Doktoranta wraz z wynikami w repozytorium OSF). Kontrolowano różnorodne parametry: liczbę itemów (10–50), liczbę

obserwacji (100–1000), średnią korelację między pozycjami (0,1–0,9), typ rozkładu (jednorodny lub normalny) oraz rodzaj skali (dyskretny: od 2- do 7-stopniowej). Tak szeroki zakres warunków wejściowych zwiększa wiarygodność wyników i pozwala określić, czy uzyskane rezultaty nie zależą np. od wielkości próby czy rozkładu. W tym miejscu chciałbym zwrócić uwagę, że na miejscu Doktoranta nie pozostawiałbym w symulacji modeli, które nie zostały rozpoznane jako jednowymiarowe w oparciu o kryteria 1-4 (Cattell, analiza równoległa, osypisko i MAP). Przyjmuję argumentację przytoczoną przez Autora, jednak w mojej ocenie za priorytetowe należałoby uznać upewnienie się, że wśród analizowanych dalej zbiorów danych nie znajdują się takie, co do których nie mamy pewności odnośnie ich jednoczynnikowej struktury. Podobne wątpliwości budzi we mnie przyjęcie często wykorzystywanych w literaturze punktów odcięcia, które – w moim, podkreślam, subiektywnym odczuciu, opartym głównie o (skądinąd dość liczne ale wciąż w żadnej mierze nie reprezentatywne) przeprowadzone przeze mnie analizy danych z badań własnych – wydają się mocno zróżnicowane jeśli chodzi o ich konserwatyzm. I tak na przykład, χ^2_p na poziomie 0,05 wydaje mi się dużo bardziej radykalny niż χ^2/df na poziomie 5,00, zwłaszcza zważywszy na średnią wartość wielkości próby w przeprowadzonych symulacjach. Nie jestem pewny jakie podejście w tym przypadku byłoby najsensowniejsze – rozumiem, że Autor bazuje na najczęściej cytowanych i przyjmowanych w literaturze punktach odcięcia, ale wydaje mi się, że zdecydowana większość analiz strukturalnych przeprowadzonych w odniesieniu do znanych mi testów dostarczała danych, które spełniłyby kryterium χ^2/df, ale już bardzo nieliczne (i raczej prowadzone na niewielkich próbach) uzyskałyby nieistotny wynik χ^2_p. Jednak zaznaczam – nie mam „twardych" danych na poparcie tej tezy; jest to moje subiektywne zdanie.

Do każdego zestawu danych budowano dwa modele CFA: poprawny (jeden czynnik) i błędny (2–5 czynników, przydział itemów losowy), co oddaje sytuację badacza mylnie przyjmującego nieprawidłową strukturę czynnikową. Jako kryterium oceny stosowano szeroką gamę miar dopasowania (χ^2, RMSEA, CFI, TLI, GFI, AGFI, RMR, SRMR, AIC, BIC) oraz indeksy modyfikacji (MI, NCP, ECP, m, nm itp.). Wyniki obu części badania (symulacji i danych empirycznych N≈5172) okazały się raczej spójne, co dodatkowo wzmacnia ich wiarygodność – można nawet powiedzieć, że w pewnym sensie zaproponowane przez Autora podejście odzwierciedlało ideę systematycznie modyfikowanych autoreplikacji.

Wnioski płynące z analiz sugerują, że klasyczne miary dopasowania **nie** rozróżniają wiarygodnie modeli jedno- vs wielowymiarowych na danych jednoznacznie jednowymiarowych (P1). Różnice w tych wskaźnikach między modelami były przeważnie

nieistotne statystycznie i przypadkowe, co oznacza, że same wartości CFI, RMSEA itp. mogą wprowadzać w błąd badacza, próbującego ustalić, czy mierzony jego narzędziem konstrukt jest w istocie jednowymiarowy. Natomiast jedynie wskaźniki oparte na błędach lokalnych specyfikacji – w tym zwłaszcza *nm%* – dawały relatywnie jasny sygnał, pozwalający na trafne rozróżnienie modeli jedno- i niejednowymiarowych (P2). Próby wyznaczenia uniwersalnych progów odcięcia (P3) zakończyły się niepowodzeniem – autor słusznie konkluduje, że brak jest stabilnych „granic", pozwalających automatycznie zakwalifikować model jako jedno- lub wielowymiarowy. Wnioski te zostały dobrze uzasadnione i zilustrowane na podstawie przeprowadzonych testów i wyników symulacji. Należy docenić rzetelność analizy – przykładowo, mgr Rak zadbał o to, by uzyskane wyniki nie były przypadkowym artefaktem symulacji (wielokrotne powtórzenia, sprawdzenie, czy matryce wyników są reprezentatywne).

Podsumowując, warto jeszcze raz podkreślić, że praca zawiera aktualny przegląd literatury (w tym odniesienia do najnowszych źródeł, np. Zheng i Bentler, 2024), starannie zaplanowaną część empiryczną i szczegółowo opisane, pogłębione analizy statystyczne. Wnioski są adekwatne do zadań badawczych i nie wykraczają poza uzyskane dane – autor otwarcie wskazuje na ograniczenia przeprowadzonych analiz (brak uniwersalnych progów, przeprowadzenie analiz tylko dla danych jednoznacznie jednowymiarowych, oparcie wyłącznie na metodzie *maximum likelihood*) i precyzyjnie interpretuje znaczenie wyników.

Zastosowania praktyczne wyników

Rezultaty pracy niosą konkretne implikacje praktyczne dla psychometrii i metodologii badań – niestety, podobnie zresztą jak w przypadku większość prac publikowanych w tym obszarze (np. Finch, 2020, Brauer i Ranger, 2021), są to w głównej mierze implikacje o charakterze „negatywnym", wskazujące na niewystarczającość stosowanych dotychczas podejść do podejmowania zasadnych empirycznie decyzji odnośnie struktury czynnikowej badanych zmiennych. Przede wszystkim wskazują one na konieczność ostrożności przy podejmowaniu decyzji o liczbie wymiarów w CFA wyłącznie na podstawie ogólnych miar dopasowania. W praktyce konstruowania testów psychologicznych może to oznaczać, że badacz mylnie założy nadmierną złożoność konstruktu (tj. jego wielowymiarową strukturę), co w naturalny sposób będzie miało negatywne konsekwencje dla interpretacji wyników i jakości samego narzędzia. Autor rozprawy słusznie podkreśla, że niewłaściwe wnioski co do struktury mogą prowadzić do istotnych błędów wnioskowania w oparciu o zebrane dane, i niewłaściwych konkluzji o charakterze teoretycznym. Mgr Rak formułuje konkretną rekomendację: w przypadkach wątpliwości co do wymiarowości warto uwzględniać także miary oparte

o wskaźniki modyfikacji (takie jak *nm%, EPC)*, które wydają się w tym kontekście znacznie lepszym wyborem, pozwalającym na rzeczywiste różnicowanie w kwestii struktury czynnikowej zmiennej latentnej. Taka podpowiedź ma sens w kontekście praktycznym, ponieważ pokazuje badaczom alternatywne, efektywniejsze podejście analityczne (zwłaszcza gdy klasyczne miary są nie dostarczają konkluzywnych wniosków). Warto też zauważyć, że praca wskazuje na ograniczenia popularnych praktyk (np. ślepe poleganie na progach wartości RMSEA/CFI) i sugeruje rozwój narzędzi oceny dopasowania modelu do danych (np. dalsza eksploracja indeksów *nm*%, ECP itp.). Raportowane w pracy wyniki mają praktyczne znaczenie zarówno dla badaczy zajmujących się tworzeniem nowych kwestionariuszy, jak i dla interpretacji badań empirycznych w psychologii – w tym ostrożniejszego, refleksyjnego korzystania z niektórych wyników CFA i rozwijania nowych kryteriów oceny struktury czynnikowej.

Uwagi krytyczne i sugestie odnośnie przyszłych badań

W recenzowanej pracy trudno było mi wskazać istotne braki merytoryczne czy poważne błędy metodologiczne. Poniższe uwagi mają raczej charakter „kosmetyczny”:

- Część teoretyczną można byłoby rozbudować o szersze omówienie wskaźników modyfikacji – choć został im poświęcony podrozdział, można by nieco lepiej zilustrować ich specyfikę, różnice pomiędzy nimi, i przykłady ich zastosowania / interpretacji.
- W pracy zastosowano wyłącznie metodę estymacji ML, tymczasem o ile mi wiadomo, ML dobrze sprawdza się dla skal Likerta tylko jeśli format odpowiedzi ma 5-7 opcji (punktów), ponieważ wtedy dopiero można je traktować jako w przybliżeniu ciągłe, zaś dla skal < 5 kategorii (2-4) znacznie lepiej sprawdza się WLSMV (*Weighted Least Squares Mean and Variance adjusted*); (zob. np. Li, 2016; Rhemtulla i in., 2012; Kline, 2023).
- Drobna uwaga redakcyjna: jako, że praca została napisana w języku polskim wskazane byłoby stosowanie w rycinach polskojęzycznych terminów (np. średnia zamiast *mean*, wynik surowy zamiast *raw result*), zaś tabele zawierające skróty byłyby łatwiejsze w odbiorze, gdyby każda z nich zawierała legendę rozwijającą owe skróty.

Poza powyższymi uwagami warto oczywiście rozważyć całe spektrum alternatywnych porównań i testów (np. między modelami o różnej liczbie czynników, czy też z uwzględnieniem

możliwości uzyskania lepszego dopasowania dla tzw. *bifactor models*); jednak praktycznie wszystkie, które przyszły mi do głowy, zostały opisane przez Autora w rozdziale 6.3. Szczególnie interesująca wydaje się idea zestawienia raportowanych w pracy analiz z sytuacją „odwrotną" w której badacz dokonuje błędnego uproszczenia struktury i dla danych wielowymiarowych zakłada strukturę jednoczynnikową.

Innym pomysłem, który mógłby okazać się ciekawym rozwiązaniem, pozwalającym dokonać integracji potencjalnie trafnych metod różnicowania struktury opisanych w pracy byłoby wykorzystanie metaheurystyk, takich jak ant colony optimization (ACO), czy algorytmy genetyczne. Co prawda metody te dopiero zaczynają być testowane w kontekście CFA. Metody takie, odpowiednio skonfigurowane, mogłyby pozwolić na automatyczne poszukiwanie optymalnej struktury. Na przykład, w takiej adaptacji ACO każda możliwa struktura modelu mogłaby być traktowana jako „ścieżka" lub „rozwiązanie", a algorytm iteracyjnie eksplorowałby różne kombinacje czynników i ładunków czynnikowych. Funkcja dopasowania (fitness) mogłaby być oparta na statystykach dopasowania modelu z analizy CFA (np. wskaźników nm%, średni MI, itp.). Podejście to byłoby bardzo obciążające obliczeniowo, ale w mojej ocenie wydaje się mieć potencjał dla rozwiązania problemów omawianych w pracy.

Wniosek i rekomendacja

Reasumując, chciałbym stanowczo podkreślić, iż Autor w mojej ocenie zrealizował interesujący i wartościowy poznawczo program badawczy, posiadający wyraźny potencjał aplikacyjny w praktyce konstruowania narzędzi psychometrycznych. Złożona rozprawa prezentuje wysoki poziom naukowy i formalny. W mojej ocenie świadczy ona o posiadaniu przez P. mgr. Tomasza Raka rozległej wiedzy w zakresie tematycznym, którego dotyczy praca, oraz umiejętności prowadzenia wysokiej jakości badań naukowych, a także prezentuje oryginalne rozwiązanie problemu naukowego. Tym samym uznaję, że przedłożona mi do recenzji dysertacja spełnia kryteria stawiane rozprawom doktorskim w ustawie z dn. 20 lipca 2018 r. – Prawo o szkolnictwie wyższym i nauce, z późniejszymi zmianami, i z pełnym przekonaniem **wnioskuję o dopuszczenie mgr. Tomasza Raka do dalszych etapów postępowania doktorskiego**.

Podpisany elektronicznie przez
Maciej Stolarski; Uniwersytet Warszawski
27.10.2025
17:22:58 +01'00'

Prof. dr hab. Maciej Stolarski