



**WYDZIAŁ NAUK
O WYCHOWANIU**

Uniwersytet Łódzki

dr hab. Monika Wróbel, prof. UŁ

Instytut Psychologii

Uniwersytet Łódzki

Recenzja rozprawy doktorskiej mgr Zofii Mazur-Sochy pt. "Achievements in instrumental performance as a result of goal realization: The role of affect and implementation intentions"

napisanej pod kierunkiem prof. dr hab. Marioli Laguny i dr Emilii Mielniczuk

Przedłożona do recenzji rozprawa doktorska prezentuje cykl badań, których celem była systematyczna analiza relacji między afektem, jakiego doświadczają osoby uczące się gry na instrumencie, a czasem poświęcanym na tę naukę. Program badawczy został bardzo dobrze zaplanowany, a jego poszczególne etapy można uznać za wzorowy przykład tego, jak zaplanować kolejne kroki postępowania badawczego (czy – nawiązując o tematu pracy – dochodzić krok po kroku do celu). Pierwszy etap tego programu stanowiły bowiem systematyczny przegląd literatury i opracowanie narzędzi do pomiaru kluczowych zmiennych. Celem tych działań było przygotowanie stabilnego gruntu pod badania właściwe. Ten systematyczny, przemyślany program badawczy stanowi w mojej ocenie bardzo mocną stronę rozprawy. Realizację tego programu także oceniam wysoko, choć Autorce nie udało się uniknąć pewnych potknięć po drodze. Poniżej szczegółowo omawiam zarówno mocne, jak i słabsze strony rozprawy.

Ocena merytoryczna

Praca ma formę zbioru cyklu artykułów, z których część Autorka już opublikowała, część jest aktualnie w recenzji, a część w przygotowaniu. Artykuły są ze sobą logicznie powiązane, a całość

dotatkowo dobrze spinają sekcje *General Introduction* i *General Discussion*, które – odpowiednio – przedstawiają założenia i cele całego projektu oraz podsumowują najważniejsze wnioski z przeprowadzonych badań, ich ograniczenia i implikacje. Cykl stanowi bardzo dobrze uporządkowaną całość – Autorka prezentuje poszczególne kroki postępowania badawczego, a każdy z nich dobrze uzasadnia i wiąże z kolejnym. Kroki te obejmują: (1) przegląd literatury, dzięki któremu Doktorantka osadza swoje poszukiwania w szerszym kontekście teoretycznym; (2) opracowanie narzędzi badawczych do pomiaru kluczowych dla projektu zmiennych; (3) badania mające na celu odpowiedź na pytanie o związek między afektem a realizacją celu, jakim jest nauka gry na instrumencie oraz (4) badania testujące efektywność interwencji MCII skoncentrowanej na regulacji afektu związanego z ćwiczeniem gry na instrumencie. Pierwsze dwa kroki mają zatem charakter wstępny, natomiast krok trzeci i czwarty to badania właściwe testujące zależności pomiędzy zmiennymi w modelu. Jak wspomniałam, ten systematyczny, uporządkowany charakter postępowania badawczego oceniam bardzo pozytywnie. Jednocześnie jednak mam pewną wątpliwość związaną z realizacją poszczególnych etapów. Mianowicie, praca napisana jest w sposób, który sugeruje, że etapy te były realizowane sekwencyjnie, a zatem powodzenie w realizacji jednego powodowało przejście do następnego. Taka strategia pośrednio sugeruje – a przynajmniej rodzi we mnie jako Czytelniczce oczekiwanie – że badania wstępne i badania właściwe będą osobnymi badaniami. Tymczasem krok 2 i 3 zrealizowano za pomocą danych zebranych na tych samych próbach. Autorka posłużyła się tymi danymi po to, by stworzyć narzędzie do badania afektu związanego z nauką gry na instrumencie (rozdział 3) oraz narzędzie do badania realizacji celu w postaci nauki gry na instrumencie (rozdział 4), a następnie przeanalizować relacje między zmiennymi zmierzonymi za pomocą tych narzędzi (rozdział 5). Tym samym dwa badania posłużyły do przygotowania trzech publikacji, co nie tylko rodzi wątpliwość, że wyniki są specyficzne dla tych prób i znacznie ogranicza możliwość ich generalizacji, ale też może zostać ocenione jako „salami publishing”. Zdecydowanie lepszą strategią, byłaby realizacja kroku 3 na innej próbie czy próbach, do czego namawiam Autorkę.

Analiza literatury, którą przeprowadziła Doktorantka, jest imponująca nie tylko dlatego, że opiera się na bardzo wymagającej strategii systematycznego przeglądu, ale także dlatego, że pozwoliła na precyzyjną identyfikację luk w dotychczasowych badaniach w interesującym Autorkę obszarze. Za szczególnie cenne i dobrze porządkujące dotychczasową wiedzę uznaję rozróżnienie

trzech rodzajów afektu (związanego z nauką gry, związanego z publicznym występowaniem oraz niezależnego od kontekstu) oraz dwóch wskaźników związanych z nauką gry na instrumencie (tj. czasu poświęcanego na naukę oraz zaangażowania w naukę). Jednocześnie nie jest dla mnie do końca jasne, dlaczego – mimo że wnioski z przeglądu wskazywały na to, że najstabilniej zbadaną jest relacja między nauką gry na instrumencie a afektem niezależnym od kontekstu (*context-free affect*) – Autorka zupełnie porzuciła ten wątek we własnych badaniach, a skoncentrowała się na afekcie powiązanym z nauką (*practice-related affect*), w przypadku którego dane były stosunkowo najbardziej spójne. Oczywiście nie ma nic niepoprawnego w takiej strategii, jednak zabrakło uzasadnienia takiego postępowania. Rozumiem, że na tego rodzaju uzasadnienie mogło nie być miejsca w samych publikacjach, które – choć są składowymi projektu – muszą stanowić na tyle niezależne całości, by można je było czytać niezależnie od reszty artykułów. Niemniej, ze względu na wagę tego uzasadnienia dla zrozumienia założeń całego projektu, można je było przedstawić np. w *General Introduction* lub *General Discussion*.

Oba autorskie narzędzia, czyli Instrumental Practice-Related Affect (IPAM) oraz Instrumental Practice Goal Realization Inventory (IPGRI), zostały przygotowane z ogromną uważnością. Na szczególne podkreślenie zasługuje fakt, że własności psychometryczne obu kwestionariuszy zostały przetestowane w dwóch badaniach wykorzystujących zróżnicowane próby i podejścia metodologiczne. Moją wątpliwość budzi jednak zasadność tworzenia narzędzia do pomiaru afektu związanego z nauką gry na instrumencie. Chociaż sam pomysł wyjścia poza model dwuwymiarowy i sięgnięcia po model afektu Russella (1980) uważam za niezwykle wartościowy (pomysł ten zresztą, jak pokazują wyniki przedstawione w rozdziale 5, okazał się trafiony), nie do końca przekonuje mnie idea tworzenia nowego zestawu itemów do pomiaru afektu rozumianego zgodnie z przyjętym modelem, ale z założenia specyficznego dla nauki gry na instrumencie. O ile w przypadku IPGRI treść itemów faktycznie jest specyficzna dla mierzonego konstrukt (tj. nauki gry na instrumencie), o tyle 16 itemów wchodzących w skład IPAM dotyczy afektu w ogóle i wydaje się, że równie dobrze nadawałaby się do pomiaru afektu np. wzbudzonego podczas oglądania filmu czy nauki do egzaminu z matematyki. W zasadzie jedyne, co jest dla IPAM specyficzne, to instrukcja i poprzedzające pomiar afektu pytania dotyczące ćwiczonego utworu muzycznego. Tym samym, można było sięgnąć na przykład po Affect Grid (Russell i in., 1989), dodając do niego te specyficzne elementy. Możliwe, że

pominięcie Affect Grid wynikało z tego, że Doktorantce zależało na pomiarze każdego w wymiarów afektu na skali więcej niż jedno-itemowej (co ma oczywiście uzasadnienie), ale tej argumentacji w pracy nie znalazłam. Słowem, mimo iż w literaturze można spotkać podobne rozwiązania (np. Multi-Affect Indicator, Warr, 2007), nie jest dla mnie do końca jasne, jakie jest ich teoretyczne uzasadnienie. Wprawdzie na str. 69 Doktorantka, za Vouskoski i Eerolą (2011) podaje, że afekt odczuwany w kontekście związanym z doświadczeniami estetycznymi jest unikalny, ale na tym poziomie ogólności ten argument nie wyjaśnia, dlaczego: (1) nie można było np. zaadaptować istniejącego narzędzia w nowym kontekście; (2) na czym polega unikalność afektu związanego z nauką gry na instrumencie; (3) dlaczego ta unikalność wymaga unikalnego zestawu itemów.

W artykule dotyczącym walidacji IPAM zabrakło też komentarza na temat relacji pomiędzy modelem kołowym Russella (1980) a dwuwymiarowym modelem Watsona i in. (1998), poza krótkim stwierdzeniem: *these constructs are related but distinct from each other* (str. 80). Komentarz ten jest istotny nie tylko ze względu na popularność rozwiązania dwuwymiarowego, ale też ze względu na to, że weryfikacji trafności podskal IPAM Doktorantka dokonała, analizując ich relacje z afektem rozumianym jako cecha i mierzonym za pomocą PANAS (Watson i in., 1998). Szczególnie istotne w tym kontekście wydają się dwa wątki, które pominięto, a które pozwoliłyby lepiej zrozumieć oczekiwania omówione na str. 76. Po pierwsze, nie jest jasne, dlaczego – jak wynika z opisu oczekiwanego wzorca korelacji – Doktorantka nie przyjęła założenia Watsona i in. (1998; Watson, 2000) o ortogonalności afektu pozytywnego i negatywnego. Po drugie, nie do końca rozumiem, dlaczego analizy trafności dokonano za pomocą skali PANAS w wersji do pomiaru afektu-cechy, a nie w wersji z instrukcją obejmującą podobny przedział czasowy jak IPAM. W tym kontekście zastanawia mnie również weryfikacja stabilności czasowej IPAM, który – jak zrozumiałam – nie mierzy afektu jako cechy, a afekt wzbudzony podczas ćwiczenia konkretnego utworu. To zaś sugeruje, że zmiana ćwiczzonego utworu może powodować zmianę afektu, stąd nie do końca jest dla mnie jasne, o czym w tym wypadku ma świadczyć czasowa stabilność pomiaru. Chcę przy tym podkreślić, że nie sugeruję tu, że przyjęte rozwiązania były błędne – moje wątpliwości dotyczą raczej braku uzasadnienia tych decyzji.

Pozostałe pytania i hipotezy badawcze są właściwie uzasadnione. Na podkreślenie zasługuje nie tylko to, że hipotezy są logiczne wyprowadzone z przesłanek teoretycznych, ale także to, że za

każdym razem, kiedy nie udało się sformułować oczekiwań kierunkowych, Autorka jasno wyjaśnia, co było tego powodem. Dzięki temu testowane modele badawcze są bardzo dobrze ugruntowane w teorii. W tym kontekście tylko lekko mylący jest Rysunek 1 ze str. 125, na którym jednakowo przedstawiono zarówno oczekiwane zależności, jak i te, co do których nie formułowano kierunkowych przewidywań.

Testowanie modeli przeprowadzono z wykorzystaniem szerokiego wachlarza metod, co potwierdza metodologiczną i statystyczną biegłość Doktorantki. Mocną stroną badań jest także wykorzystywanie powtarzanych pomiarów. Doświadczenie w zbieraniu danych w planach podłużnych okazało się szczególnie cenne w przypadku badania opisanego w rozdziale 6, gdzie odroczony pomiar pozwolił Autorce na przetestowanie skuteczności zastosowanej interwencji. Kwestią dyskusyjną związaną z testowaniem modeli jest jednak ustalanie wielkości próby na podstawie „reguły kciuka”. Na przykład, tworząc IPAM i IPGRI Autorka przyjęła, że liczba obserwacji powinna być przynajmniej 5 razy taka, jak liczba itemów poddawanych analizie czynnikowej. Korzystanie z tego rodzaju reguł, mimo iż opisywane w literaturze, jest opatrzone błędem (por. MacCallum i in., 1999; Wolf i in., 2013), stąd lepszym podejściem byłoby uzasadnienie wielkości próby wynikami analizy mocy (np. z wykorzystaniem metod symulacji, por. Muthén i Muthén, 2002; Stephenson i Holbert, 2003). Podobny problem zresztą dotyczy nie tylko badań służących opracowaniu IPAM i IPGRI, ale też pozostałych analiz. Jak wspomniałam, na danych opisanych w rozdziałach 3 i 4 prowadzone były analizy relacji pomiędzy afektem a realizacją celu nauki gry na instrumencie (rozdział 5), co wymagało ponownego oszacowania, jakiej wielkości efekt pozwalały wykryć zebrane próby. W ostatnim badaniu (rozdział 6) również zabrakło analizy mocy, a w dyskusji po raz kolejny powołano się na „regułę kciuka”, zgodnie z którą $N = 100$ wystarczyło do przeprowadzenia SEM (str. 136). Rozumiem, że Doktorantka mogła nie oszacować wielkości prób a priori, jednak w tym wypadku nadal wskazane byłoby przeprowadzenie *sensitivity power analysis*.

Dyskusja wyników, zarówno na poziomie poszczególnych publikacji, jak i *General Discussion*, wskazuje na bardzo refleksyjne podejście Doktorantki do uzyskanych wyników. Szczególnie cenne wydaje mi się umiejętne balansowanie między proponowaniem alternatywnych interpretacji dla uzyskanych danych a kurczowym trzymaniem się teoretycznych założeń, dzięki czemu Dyskusja wychodzi poza to, co opisano we wprowadzeniu, jednocześnie nie proponując zbyt daleko idących

spekulacji. Problematiczne jednak wydaje mi się traktowanie wyników „marginalnie istotnych statystycznie” prawie na równi z tymi, które były istotne. Warto przy tym podkreślić, że już samo poleganie na takich wynikach, zwłaszcza przy próbach, o których nie wiadomo, jaki efekt pozwalają zidentyfikować, jest dyskusyjne, ponieważ stwarza ryzyko interpretowania wyników nieistotnych jako fałszywie pozytywnych. Namawiałabym zatem Autorkę, aby zrezygnować z ich raportowania (por. Olsson-Collentine i in., 2019), a tym bardziej dyskutowania tych wyników w sposób sugerujący jakoby opisywane zależności rzeczywiście zostały potwierdzone.

Podsumowując, czytając rozprawę, zidentyfikowałam kilka usterek zarówno w jej warstwie teoretycznej, jak i empirycznej, jednak usterki te nie przesłaniają mojej całościowej pozytywnej oceny pracy. Wierzę też, że ze względu na status niektórych artykułów (w recenzji lub w przygotowaniu), niektóre z wskazanych niedociągnięć uda się wyeliminować, zanim wyniki zostaną opublikowane.

Ocena formalna

Praca została przygotowana z bardzo dużą starannością. Została napisana komunikatywnym i zwięzłym językiem oraz zgodnie ze standardami APA. Struktura każdego z rozdziałów, jak i prezentacja wyników (w tekście i na wykresach/w tabelach) jest czytelna i logiczna, dzięki czemu podążanie za wywodem Doktorantki nie nastęrcza trudności. Jedyny moment, w którym się zgubiłam, to dość dowolne operowanie nazwami podskal IPAM – raz Doktorantka używa nazw HAPA, LAPA, HANA i LANA, a raz – odpowiednio – *enthusiasm*, *comfort*, *anxiety* i *depression*. Oczywiście takie postępowanie jest uzasadnione i ma wsparcie w modelu Russella (1980), jednak dopiero w *General Discussion* pojawia się informacja, że Autorka będzie używała tych nazw zamiennie. Czytając wcześniejsze rozdziały, nie do końca rozumiałam, kiedy i dlaczego stosowane są jedne nazwy, a kiedy drugie, czego najdobitniejszym przykładem jest rozdział 5, gdzie w tekście i większości tabel używane są skróty HAPA, LAPA, HANA i LANA, a w Tabeli 2 nagle pojawia się alternatywna nomenklatura. Integrowanie tych alternatywnych systemów nazewnictwa leży po stronie czytelnika, co oczywiście jest możliwe, niemniej sprawia problemy w odbiorze. Problem jest oczywiście łatwy do rozwiązania poprzez zamieszczenie informacji o tym, że systemy te są stosowane zamiennie w każdej z publikacji (zamiast dopiero w *General Discussion*, do której zresztą czytelnicy artykułów i tak nie sięgną).

Zresztą, poza tym drobnym mankamentem, dbałość Autorki o to, by jej przekaz był jasny dla odbiorcy jest ponadprzeciętna, a pracę czyta się z łatwością i przyjemnością.

Niewielkiego dopracowania wymagają jednak fragmenty, których nie publikowano (szczególnie rozdział 6 i *General Discussion*), ponieważ zawierają drobne błędy stylistyczne i gramatyczne (np. „Following, we tested...”). Choć błędy te nie utrudniają odbioru tekstu oraz nie wpływają na moją ocenę pracy, niepublikowane fragmenty rozprawy zyskałyby dzięki korekcie językowej. Uwaga ta ma znaczenie szczególnie w odniesieniu do rozdziału 6, który docelowo ma być publikowany w formie artykułu.

Wniosek

Podsumowując, pomimo wypunktowanych wyżej uwag krytycznych, wysoko oceniam przedłożoną do recenzji rozprawę doktorską. Rozprawą tą mgr Zofia Mazur-Socha udowodniła, że jej kompetencje odpowiadają kompetencjom, jakie stawiane są Kandydatom i Kandydatkom do stopnia doktora. Autorka pokazała, że potrafi zaoferować oryginalne rozwiązanie problemu naukowego, a jej wiedza i warsztat badawczy umożliwiają samodzielne prowadzenie pracy naukowej. Nie mam najmniejszych wątpliwości, że recenzowana praca spełnia kryteria określone w artykule 13.1 ustawy z dnia 14 marca 2003 r. o stopniach i tytule naukowym oraz stopniach i tytule w zakresie sztuki, i z pełnym przekonaniem wnioskuję o dopuszczenie Kandydatki do dalszych etapów postępowania doktorskiego.

Monika Lwóbel