

Prof. dr hab. Aleksandra Łuszczynska  
Wydział Psychologii we Wrocławiu  
SWPS Uniwersytet Humanistycznospołeczny

Wrocław, 30 grudnia 2020

**Recenzja**  
**rozprawy doktorskiej mgra Macieja Behnke**  
**pt. „Psychofizjologiczny model wyzwania i zagrożenia w E-sporcie”**

W skład recenzowanej rozprawy doktorskiej wchodzi trzy opublikowane manuskrypty, w których mgr Behnke jest pierwszym autorem, a także autorem korespondencyjnym. Choć do ocenianej rozprawy nie załączono deklaracji o wkładzie poszczególnych autorów, rola pierwszego i korespondencyjnego autora nie pozostawia wątpliwości, że wkład mgra Behnke w przygotowanie badań i odpowiednich tekstów jest kluczowy. Artykuły zostały opublikowane w czasopismach z 2-3 kwartyła (International Journal of Psychophysiology, Psychology of Sport and Exercise) i 1 kwartyła (Emotion) dla odpowiednich dyscyplin wyróżnianych przez Web of Science. Trzy publikacje stanowią zbiór spójny tematycznie i mogą być ocenione jako wyróżniające osiągnięcia młodego naukowca.

**Uwagi ogólne**

Tytuł sugeruje, iż rozprawa zaproponuje jeden model, ujmujący wyzwania i zagrożenia w E-sporcie. Zgodnie ze streszczeniem pracy, uważam iż zaprezentowane w rozprawie badania przyczyniły się do pozyskania nowej wiedzy o psychofizjologicznych aspektach E-sportu, przede wszystkim w kontekście wyzwania, zagrożenia i rozbieżności funkcji emocji), jednakże wbrew tytułowi zabrakło w niej sugerowanego w tytule jednego ‘modelu’ (teoretycznego lub empirycznego), spajającego uzyskane wyniki. Być może trafniejsze byłoby bardziej ‘miękkie’ wskazanie, np. iż badania prowadzono w kierunku stworzenia i weryfikacji modelu.

Streszczenie rozprawy zawiera opis i interpretację kluczowych wyników, ale korzystniej byłoby je przygotować w większej zgodności ze standardami raportów z oryginalnych badań ilościowych (Applebaum i in., 2018, American Psychologist). W szczególności, mam tu na myśli informacje o metodologii badań, które powinny pojawić się w streszczeniu, takie jak rodzaj uczestników (dorośli ludzie z populacji ogólnej?), wielkość próby, narzędzie/procedury, czy wielkości wykrytych efektów. Te informacje są kluczowe dla czytelnika jakiegokolwiek streszczenia, by mógł (np. z bazy wiedzy zawierającej same streszczenia) pozyskać wstępną wiedzę o jakości prowadzonych badań i możliwościach generalizacji wyników.

**Badanie 1 (Publikacja 1)**

Przedstawiony systematyczny przegląd badań oparty jest o solidne podstawy teoretyczne i stanowi wartościowe podsumowanie i uogólnienie kumulujących się badań w obszarze reakcji naczyniowo sercowej na zagrożenie/wyzwanie. Badanie ma szereg silnych stron, takich jak dobrze przeprowadzona meta-analiza

(biorąca pod uwagę heterogeniczność badań oryginalnych i potencjalny tzw. publication bias) czy adekwatnie precyzyjnie sformułowany obszar badań. Systematyczny przegląd jest oparty o standardy PRISMA, wyniki i ogólne konkluzje są zaprezentowane w spójny i przekonujący sposób, dobrze policzone i zinterpretowane efekty ważne.

Jako recenzent mam liczne drobne uwagi dotyczące zaprezentowanego badania. Na przykład, nie jest jasne jakie było źródło doboru słów kluczowych użytych w systematycznym przeglądzie: można byłoby wykorzystać/wskazać źródło uzasadniające, że ich dobór jest wystarczająco szeroki (np. użycie jedynie sformułowania 'performance' do wyszukiwania tekstów dot. performance, nie uwzględnienie np. 'perform\*'). Kryteria dotyczące wyłączenia mogłyby być opisane dokładniej (np. czy wyłączano badania korelacyjne przekrojowe? Czy wyłączano teksty w językach innych niż angielski). Niestety nie wszystkie etapy wyszukiwania danych, ich ekstrakcji i kodowania były wykonane przez co najmniej dwie niezależnie pracujące osoby, co zwiększa tzw. risk of bias. W tych częściach, gdzie w proces pozyskiwania danych i ich analizy zaangażowane były 2 osoby, warto byłoby wykazać poziom zgodności (np. policzyć współczynnik kappa czy ICC) oraz wskazać jak rozwiązywano sytuacje braku zgodności między dwoma badaczami. W raporcie i badaniach oryginalnych brakuje informacji o płci osób badanych czy źródła próby (o części wiadomo że byli studentami, inni byli z populacji ogólnej? Tabela 1), co utrudnia wnioski dotyczące możliwości generalizacji wyników. Zamiast opierać się (wyłącznie) o I-kwadrat w ocenie heterogeniczności warto wziąć pod uwagę np. Tau, ze względu na ograniczenia współczynnika I-kwadrat (por. np. Borenstein i in., 2017, Research Synthesis Methods). Kolejnym ograniczeniem jest brak oceny jakości włączonych badań (np. przy wykorzystaniu STROBE lub GRADE lub jakiegokolwiek innego instrumentu do oceny jakości badań oryginalnych, a przez to o ustalenia risk of bias w badaniach oryginalnych). Generalnie, najlepsze przeglądy i meta-analizy winny spełniać kryteria AMSTAR-2. Warto było również dokonać prerejestracji badania (np. w repozytorium PROSPERO).

Niezależnie od powyższych uwag, uznaję, iż przegląd zawarty w zbiorze publikacji jest wartościowym poszerzeniem wiedzy psychologicznej i ogólnie oceniam go pozytywnie.

## **Badanie 2 (Publikacja 2)**

Badanie i przygotowana publikacja mają wiele bardzo silnych stron, do których należy schemat badania (2 wersje manipulacji z pomiarem pre-post), staranny pomiar zmiennych wynikowych, bardzo precyzyjny opis pomiaru i manipulacji, pozwalający na replikację, wysoka trafność ekologiczna procedury, nowatorstwo badania (podobnych badań w kontekście psychologii sportu czy społecznej lub edukacyjnej jest naprawdę niewiele), kontrola wcześniejszego doświadczenia E-sportowców (w modelu ścieżkowym), wykorzystanie dobrze znanego modelu teoretycznego z psychologii społecznej. Opis badanej procedury jest bardzo dobry, bardzo pozytywnie należy też ocenić pomiar reakcji kardiowaskularnej. Schemat badania jest również bardzo dobrze opisany. Na pochwałę zasługuje przeprowadzony tzw. manipulation check. Ogólnie, oceniam badanie jako bardzo wartościowe i nowatorskie, stanowiące znaczący wkład w rozwój wiedzy i dyscypliny.

Jak każde badanie (oraz raport z badania, czyli Publikacja 2), oceniany tekst ma pewne słabości. Część (np. kwestie teoretyczne) można uznać za czysto dyskusyjne, inne (np. braki danych) odnoszą się

braku spełnienia wszystkich tzw. złotych standardów w raportowaniu wyników z badań ilościowych (por. np. Applebaum i in., 2018, American Psychologist).

W psychologii sportu (i społecznej czy edukacyjnej) wiodącym paradygmatem jest określenie celów i ich wartości motywującej (ego/mastery vs task/performance oriented achievement goals oraz approach vs avoidance performance orientation). Osoby z silną orientacją task/performance mają (teoretycznie) słabą reakcję na wyniki innych osób i motywowane tym, jak daleko im do wykonania według własnych, wewnętrznych standardów, które stanowią kluczowy punkt odniesienia. Osoby z silną orientacją typu ego/mastery dążą przede wszystkim do wygranej i oceniają sukces/porażkę odnosząc się do wykonania innych (Dweck i Leggett 1988; Nicholls, 1989, Roberts i in., 2007, Elliot, 1997, Lochbaum i in.). Approach-avoidance (lub kontrast między approach a avoidance wewnątrz osoby badanej) są również dobrze udokumentowanym nurtem zależności. Podejście wyzwanie vs zagrożenie w kontekście do emocji i odpowiedzi kardiovaskularnych w sporcie doczekało się również modelu teoretycznego (Jones i in. 2009, the Theory of challenge and threat states in athletes; <https://doi.org/10.1080/17509840902829331>), która opisuje uwzględnianie w Badaniu 2 zjawiska. Badacz z obszaru psychologii sportu może odczuwać rozczarowanie pewnym brakiem w Publikacji 2, który dotyczy braku odniesienia do tych – w moim przekonaniu – najważniejszych nurtów badań i teorii w psychologii sportu, a także związanych z nimi ograniczeniami badania własnego. Jest to również zaskakujące, ponieważ jednym z mierzonych wskaźników jest różnica w ewaluacji własnych zasobów i kompetencji innych zawodników (określone w artykule jako challenge-threat evaluation; podrozdział 2.2.3 w publikacji).

Jedno z założeń badania głosiło to, iż wygrana finansowa (15-60 USD) będzie wzmacniała motywację do udziału w badaniu. Czy wzmacniała również motywację do osiągnięcia celu (wygranej?), reakcji kardiovaskularnej, i czy podobnie u wszystkich uczestników?

Pewną bardziej generalną słabością schematu badania jest brak grupy kontrolnej (grupy bez manipulacji zagrożeniem/wyzwaniem). Brak takiej grupy uniemożliwia ocenę efektu czasu (Time effect; de facto bycie w sytuacji drugiego meczu w turnieju bez podawanej dodatkowej informacji o wynikach innych zawodników).

Metoda badania nie zawiera informacji o płci i wykształceniu osób badanych (poza tym brakiem, opis badanej próby jest bardzo staranny). Nie zakładano testowania żadnych dodatkowych zmiennych ubocznych (np. w celu sprawdzenia czy wyniki są tzw. robust po kontroli zakłóceń?).

W analizach wstępnych brakuje jednak tzw. randomization check – czy randomizacja (np. w zakresie poziomu zawodowstwa w E-sporcie czy threat-challenge self-evaluation) była efektywna? Nie jest również jasne jakie były procedury randomizacji (jak ją wykonano).

Nie jest jasne, dlaczego wprowadzono założenie przy estymacji wielkości badanej próby, sugerujące jedynie 2 predyktory w analizie. Czy oznacza to, iż założono estymację efektu jedynie przy kontroli jednej dodatkowej zmiennej poza manipulacją? W modelu przedstawionym na rycinie 2 widać zdecydowanie więcej zmiennych, a zatem więcej parametrów. Biorąc pod uwagę liczbę parametrów w modelu na rycinie 2 można sądzić iż  $N = 82$  będzie skrajnie niskie, i że model może nie mieć odpowiedniej mocy do wykrycia średnich efektów. Czytając o metodzie MLMVA oraz referowany w tekście Publikacji 2 tekst Bentlera (2007) nie znalazłam informacji, iż jest to adekwatna metoda estymacji dla prób z  $N < 100$ . Raczej, przy

próbie z  $N < 200$ , Bentler (2007) sugeruje testowanie co najmniej jednego dodatkowego modelu, alternatywnego do hipotetycznego.

Testując hipotezy analizowano zaproponowany model teoretyczny: niestety nie jest on dokładnie opisany w tekście, czyli nie wiadomo gdzie były w nim ścieżki regresji?? Jakie efekty dokładnie założono?. Poziomy bazowe 2 wskaźników psychofizjologicznych powinny być skorelowane, a w pomiarze 2 powinny być uwzględnione kowariancje między nimi (jest to teoretycznie i empirycznie uzasadnione). Prawdopodobnie, wskaźniki fizjologiczne T1 powinny być też skorelowane z doświadczeniem i ewaluacją w pomiarze 1? Ogólnie, nie jest jasne czy takie kowariancje/korelacje były uwzględnione.

W toku analiz z modelu hipotetycznego usunięto część ścieżek (te, które w modelu hipotetycznym były nieistotne). Ta strategia, polegająca na modyfikowaniu modelu hipotetycznego (założonego na podstawie teorii i wcześniejszych badań) poprzez usuwanie części ścieżek jest strategią eksploracyjną, którą można uznać za słabą strategię analityczną (biorąc pod uwagę wielkość i charakter próby oraz założony model hipotetyczny) i zagrożoną małą replikowalnością (w porównaniu z innymi próbami populacyjnymi). Szkoda również, że nie przeprowadzono tzw. sensitivity analysis, polegającej np. na powtórzeniu analizy z kowariantami (zmiennymi zakłócającymi) i prześledzeniu podobieństw we wzorze wyników.

Jak w każdym badaniu, warto byłoby wskazać czy były jakiegokolwiek braki danych i jak z nimi postępowano, a także opisać nieco dokładnie tzw. flow of participants, czyli ile osób rekrutowano, ile ostatecznie zgodziło się wziąć udział w badaniu, ile uczestniczyło we wszystkich pomiarach. Można by również dokonać prerejestracji (lub nawet post-rejestracji badania), co przyda się w kolejnych meta-analizach uwzględniających repozytoria rejestracyjne.

### Badanie 3/Publikacja 3

Badanie 3 ma bardzo wysokiej jakości wprowadzenie teoretyczno-empiryczne i niezwykle ciekawe oraz ważne hipotezy. Osiągnięcie nowatorstwa w tym obszarze psychologii (emocje – ocena poznawcza – reakcja kardiowaskularna) jest trudne, w obliczu bardzo dużej liczby badań, ale w przedstawionym w publikacji badaniu udało się to osiągnąć. Schemat badania jest bardzo przekonujący, porównujący różne rodzaje manipulacji i zmienne wynikowe mierzone ‘obiektywnie’, w ekologicznie trafnym kontekście. Silne strony to m.in. staranne przeprowadzenie tzw. manipulation check.

Ponownie, z obowiązku recenzenckiego przedstawiono poniżej pulę uwag, które nie umniejszają wysokiej pozytywnej ogólnej oceny badania i publikacji 3.

Rycina 1 prezentuje model teoretyczny, który jest też modelem hipotetycznym weryfikowanym w badaniu. Niestety, ani z tego modelu (ani z opisu analizy statystycznej) nie jest jasne, czy (jakie) założono kowariancje w modelu. Kowariancje będą miały znaczący wpływ na dopasowanie modelu. Dodatkowo, ich usunięcie wpływa na możliwość osiągnięcia efektów głównych i pośrednich.

Nie jest jasne, czy zaprezentowane w wynikach wskaźniki dopasowania modelu dotyczą modelu już zmodyfikowanego (z usuniętymi ścieżkami) czy też bez modyfikacji (podejrzewam, że bez, warto zawsze przedstawić pełne dopasowanie modelu hipotetycznego). Na oko, mógł też wystąpić błąd w podawanych współczynnikach: ogólny model ma wartość  $\chi^2/kwadrat/df = 54.103$ , a delta  $\chi^2/kwadrat/DF$  po usunięciu ścieżek wynosi  $= 51.83$  i wskazane, jest że taka zmiana  $\chi^2/kwadrat/df$  byłaby nieistotna? Wydaje mi się to mało prawdopodobne?

Podobnie jak w badaniu 2, można mieć obawy, iż usunięcie części ścieżek z modelu hipotetycznego (z powodu braku ich istotności) nie jest optymalnym rozwiązaniem analitycznym. Na marginesie, warto wspomnieć, iż nie jest jasne jak prowadzono porównania między modelem hipotetycznym a modelem z usuniętymi ścieżkami? Standardowym rozwiązaniem jest ustalenie iż dane ścieżki w modelu konkurencyjnym są równe zero i porównanie tego modelu z modelem hipotetycznym. Nie jest jasne, czy taką właśnie procedurę zastosowano czy też 'fizycznie' usunięto ścieżki z modelu (zamiast wprowadzić dodatkowe założenie, że ich efekty każdej z nich są równe zero).

Podobnie jak w badaniu 2, brakuje pewnych informacji, na przykład o sposobie randomizacji uczestników do warunków badania, oceny czy randomizacja była efektywna.

Brakuje informacji o skali missingów i o tym czy pojawiały się one losowo (MCAR) oraz argumentu dla usunięcia osób z brakami danych nie są przedstawione (standardem jest raczej uzupełnianie braków np. za pomocą takich procedur jak multiple imputation czy full information maximum likelihood).

Ponownie, nie jest dla mnie jasne jak oszacowano wielkość próby; czy brano pod uwagę schemat analityczny z modelami zagnieżdżonymi i liczbą predyktorów w równaniach analizy ścieżkowej? Jak rozumiem z metaanalizy Behnke i Kaczmarek (2018), spodziewane efekty zależności między reakcją kardiowaskularną a wyzwaniem/zagrożeniem są małe, więc być może lepiej byłoby założyć, iż efekt jest mały, czyli sigma-kwadrat = około 0.02. Sigma-kwadrat = r-kwadrat / (1-r-kwadrat), czyli dla  $r = .10$  (Behnke i Kaczmarek, 2018) sigma-kwadrat wynosi  $< 0.02$ .

Brak wykrycia relacji między zmiennymi dotyczącymi odpowiedzi kardiowaskularnej a innymi (mierzonymi samoopisowo) zmiennymi w badaniu wynika najprawdopodobniej ze stosunkowo niskiej mocy do wykrycia małych efektów, które prawdopodobnie pojawiają się w tej relacji. Stąd, być może sensownym uzupełnieniem (jako dodatkowe analizy) byłoby przetestowanie prostych efektów manipulacji na mediatory i zmienną wynikową (np. za pomocą ANOVA). Oczywiście, zwiększa to liczbę wykonywanych analiz, ale za to pozwala na przetestowanie potencjalnego efektu na najbardziej interesujący mediator, odpowiedź kardiowaskularną, w warunkach adekwatnej mocy analiz.

Szkoda, iż badanie nie było prerejestrowane (lub post-rejestrowane).

## Uwagi końcowe

Choć recenzja zawiera dużą liczbę uwag, są to jednak uwagi o nieco mniejszej wadze i znaczeniu, niż ogólne uwagi pozytywne, dotyczące nowatorstwa i ogólnego znaczenia przeprowadzonych badań, jakości stosowanych procedur badawczych (np. manipulacji), staranności opisu procedur pozwalających na replikację.

Przytoczone uwagi dotyczące braków (czy wskazane niejasności) można uznać za dyskusyjne i część z nich wynika ze stale ewoluujących standardów badań. Część zaprezentowanych badań była zapewne projektowana i prowadzona w czasie, gdy te standardy nie były jeszcze tak silnie podkreślane w literaturze. W ocenie zbioru cyklu publikacji, należy wziąć pod uwagę, iż w wielu czasopismach, w tym i tych z 1 kwartyła w psychologii, wymagania w procesie publikacyjnym odbiegają od złotych standardów, np. stosowanych np. w światowej medycynie klinicznej (np. standardy CONSORT dla badań eksperymentalnych), jak i również presja środowiska na doskonałość metodologiczną (np. w zakresie tzw. sensitivity analysis, missing data analysis and treatment, etc) jest niewielka, co w konsekwencji może

powodować publikowanie manuskryptów które nie zawierają wszystkich informacji, np. sugerowanych w złotych standardach dotyczących badań ilościowych w psychologii (Consort, Applebaum i in., 2018, American Psychologist). Jest to szczególnie ważne w procesie kształcenia młodych badaczy. Nie wątpię, iż gdyby mgr Behnke napotkał bardziej wnikliwych (ale przyjaznych) recenzentów i edytorów, jego publikacje zawierałyby nieco więcej metodologicznych szczegółów, które przedstawiono powyżej.

### **Konkluzja**

Z pełnym przekonaniem wnioskuję do Rady Dyscypliny Psychologia Wydziału Psychologii i Kognitywistyki UAM w Poznaniu o dopuszczenia mgr Macieja Behnke do dalszych etapów postępowania. **Rozprawa odpowiada warunkom określonym w odpowiedniej ustawie o tytule i stopniach naukowych** (Dz.U nr 65 poz. 595 z późniejszymi zmianami) i w pełni kwalifikuje mgra Behnke do nadania stopnia doktora w zakresie psychologii. Jednocześnie, ze względu na nowatorstwo badań i wkład wniesiony w naukę uzyskany za pomocą badań, przedstawionych w ocenianym zbiorze publikacji, **wnoszę o wyróżnienie mgra Macieja Behnke za przedstawioną rozprawę doktorską.**

Wrocław, 30 grudnia 2020

*Aleksandra Tuszczyńska*