

WALIDACJA NARZĘDZIA POMIAROWEGO W NAUKACH O ZARZĄDZANIU

<https://doi.org/10.33141/po.2019.04.01>

Przegląd Organizacji, Nr 4 (951), 2019, ss. 3-10

www.przegladorganizacji.pl

©Towarzystwo Naukowe Organizacji i Kierownictwa (TNOiK)

Wojciech Czakon

Wprowadzenie

Podstawowym celem nauk społecznych jest opracowanie teorii wyjaśniających zachowania ludzi (Peter, 1981, s. 133). Realizując ten cel, nauki o zarządzaniu skupiają się na zjawiskach wewnątrzorganizacyjnych, organizacyjnych i międzyorganizacyjnych (Czakon, 2017), które zachodzą za sprawą „widzialnej ręki menedżera” (Chandler, 1977). Badacze nauk o zarządzaniu spotykają się z krytyką dotyczącą rygoru metodologicznego tej dyscypliny (Czakon, 2014). Najczęściej, wręcz potocznie, pojawiają się argumenty o: rozroście teoretycznym przy niedorozwoju operacjonalizacji (Venkatraman, Grant, 1986, s. 72), nieostrości konceptualnej (Gatignon i in., 2002, s. 1103), niejasności osadzenia empirycznego (Danneels, 2016, s. 2175), braku rzetelnego narzędzia pomiarowego (Govindarajan, Kopalle, 2006, s. 190). W rezultacie występują wątpliwości dotyczące rzetelności ustaleń, szkodliwe zarówno dla wiarygodności, jak i przydatności praktycznej nauk o zarządzaniu. Jednocześnie charakterystyczną cechą nauk o zarządzaniu jest wysoki stopień samokrytycyzmu oraz reakcja na krytykę w dążeniu do podniesienia legitymizacji akademickiej (Hatchuel, 2005). Zgłaszane wątpliwości odnośnie do jakości badań działają mobilizująco, skłaniają do dbałości o rygor metodologiczny, krytycznej oceny praktyk badaczy oraz ich rozwoju. Niniejszy artykuł wpisuje się w ten wysiłek poprzez uporządkowaną prezentację walidacji narzędzi pomiarowych¹.

Teoria, odpowiadając na pytanie, co jest, jak się przejawia oraz dlaczego przyjmuje określoną postać (Czakon, 2017), posługuje się konstruktami². Konstrukty to konceptualne abstrakcje zjawisk, które nie są bezpośrednio obserwowalne (Suddaby, 2010), jak na przykład: orientacja strategiczna (Venkatraman, 1989), innowacja dysruptywna (Govindarajan, Kopalle, 2006) czy organizacyjne zachowania obywatelskie (Suddaby, 2010). Wysoki stopień korespondencji pomiędzy konstruktami a jego przejawem empirycznym (Bacharach, 1989, s. 498) jest warunkiem niezbędnym rozwoju teorii. Umożliwia bowiem jej testowanie, wyposażając badacza w zdolność predykcji kierunku oraz wielkości zmiany kluczowych

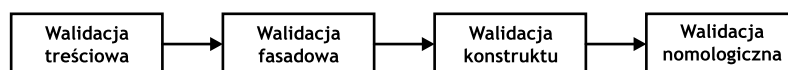
zmiennych (Peter, 1981, s. 134), a praktyków w zdolność podejmowania lepszych decyzji (Bagozzi, Edwards, 1998, s. 78). Zatem pomiar konstruktów jest niezbędny dla rozwoju nauk o zarządzaniu (Venkatraman, 1989, s. 942), natomiast brak trafnych miar stanowi główną przeszkodę (Danneels, 2016, s. 2183).

Poznanie przyczyn, przebiegu procesów oraz ich skutków w naukach o zarządzaniu wymaga udziału menedżerów, ponieważ zachowania organizacji wynikają z ich decyzji, a te z kolei ze sposobu postrzegania świata zewnętrznego i organizacji (Chen i in., 2007). Kluczowy dla zrozumienia przedmiotu badań nauk o zarządzaniu jest sam menedżer, informator posiadający odpowiednie kwalifikacje i pozycję w organizacji (Bagozzi i in., 1991, s. 423). Uchwycenie tego, co i jak postrzega menedżer, wymaga interakcji z nim właśnie, sama obserwacja nie wystarczy. Stąd znaczenie gromadzenia danych bezpośrednio od niego, które zwykle przyjmuje postać mniej lub bardziej ustrukturyzowanych wywiadów, eksperymentów czy narażeni. Niniejszy artykuł skupia się na kwestionariuszach, które stanowią narzędzie badawcze posługujące się skalami dla pomiaru konstruktów³.

Celem artykułu jest przedstawienie procedury walidacji narzędzia pomiaru w naukach o zarządzaniu. Walidacja to ocena stopnia trafności narzędzia badawczego, czyli tego czy pozwala ono zmierzyć konstrukt w sposób akceptowalny. Trafność nie jest zwykle doskonała, stąd praktyka badań oraz konsensus badaczy wyznaczają minimalne wymogi, które należy spełnić, aby narzędzie badawcze walidować.

W artykule przyjęto kolejność chronologiczną (Venkatraman, Grant, 1986), tj. najpierw przedstawiono walidację treściową, następnie fasadową, dalej konstruktów, tj. dyskryminacyjną i konwergentną (Bagozzi i in., 1991), by zakończyć walidacją nomologiczną (rys. 1).

Nauki o zarządzaniu szeroko czerpią z dorobku psychologii, zwłaszcza w zakresie prac fundamentalnych (Cronbach, Meehl, 1955), kompendiów (Nunally, 1978) oraz analiz krytycznych (Hagger i in., 2017), ale zauważyć się



Rys. 1. Sekwencja etapów procedury walidacji w naukach o zarządzaniu

Źródło: opracowanie własne na podstawie: Churchill, 1979; O'Leary-Kelly, Vokurka, 1998

daje kształtowanie własnej praktyki (O-Leary-Kelly, Vokurka, 1998; Boudreau i in., 2001; Danneels, 2016), a także żywej debaty o doskonaleniu praktyk stosowanych przez badaczy nauk o zarządzaniu (Gefen i in., 2011). Artykuł ma charakter metodologiczny, oparty na systematycznych przeglądach praktyk stosowanych w zakresie walidacji, rekomendacjach jej rozwoju oraz przykładach zastosowania procedur walidacji w naukach o zarządzaniu.

Miejsce walidacji w opracowaniu skal pomiarowych

Konstrukty to pojęcia, przedstawiające „wydestylowane” zjawiska w postaci wyróżników zrozumiałych dla wspólnoty badaczy (Suddaby, 2010, s. 346). Psychometryczne rozumienie konstruktów sprowadza się do jakiejś postulowanej cechy człowieka, którą można odzwierciedlić w testach (Cronbach, Meehl, 1955, s. 4). W naukach o zarządzaniu mogą to być także cechy interpersonalne (Glińska-Noweś, 2017), organizacji (Zakrzewska-Bielawska, 2018), diad (Mitrega i in., 2012), sieci organizacji (Kawa, 2017), ekosystemów (Klimas, 2019), a więc dotyczyć mogą wielu poziomów analizy. Występują także konstrukty wielopoziomowe (Gavetti, 2005), o różnych przejawach w zależności od poziomu analizy, co wywołuje potrzebę badania mikrofundamentów oraz mechanizmów wpływu pomiędzy poszczególnymi poziomami analizy (Czakon, 2015). W rezultacie „pożyczanie” praktyk psychometrii wiąże się nieuchronnie z potrzebą ich adaptacji do uwarunkowań empirycznych nauk o zarządzaniu.

Celem pomiaru jest dostarczenie empirycznej estymacji teoretycznego konstruktów (Gerbing, Anderson, 1988, s. 186). Konstrukty są pojęciami zaprojektowanymi po to, aby uporządkować wiedzę oraz kierunkować badania empiryczne nad wybranym aspektem rzeczywistości (Peter, 1981, s. 134). Stosowanie pojedynczej miary dla konstruktów nie powinno być praktykowane ze względu na ustalone w naukach o zarządzaniu ich poważne ograniczenia, tak istotne, iż budzą poważne wątpliwości względem rygoru (Zaichkowsky, 2015, s. 261). Obejmują one: niską korelację z mierzoną cechą przy jednoczesnej relacji wobec innych cech, tendencję do kategoryzowania ludzi w niewielką liczbę grup, odpowiadających przyjętej liczbie rang (5, 7, 9 itd.) czy obciążenie znacznym błędem pomiaru przejawiającym się niskim prawdopodobieństwem uzyskania tych samych wyników w kolejnych procedurach gromadzenia danych (Churchill, 1979). Wobec tego do pomiaru stosuje się dwie lub więcej miar nazywanych skalami, które stanowią alternatywne wskaźniki tego samego konstruktów (Gerbing, Anderson, 1988, s. 186).

Rozwój skal obejmuje procedury generowania elementów oraz opracowania ich miar, tj.: zdefiniowanie pola znaczeniowego konstruktów, procedury generowania

elementów, aprioryczną specyfikację wymiarów, ocenę wymiarów (Peter, Churchill, 1986) oraz opracowanie reguł interpretacji uzyskanych wyników (Churchill, 1979). Skale pomiarowe w zarządzaniu czerpią z dorobku psychometrii, bowiem podobnie jak w psychologii odzwierciedlają podstawowe zmienne istotne w badaniach osobniczej różnorodności (Nunnally, 1975, s. 7). Droga, którą przeszła metodologia psychometrii, jest powtarzana i adaptowana w naukach o zarządzaniu. Wręcz uderzające jest podobieństwo obserwacji stanu debaty metodologicznej w zakresie rozwoju skal pomiarowych w latach 50. ubiegłego wieku w psychologii (Nunnally, 1975) i w zarządzaniu (Venkatraman, Grant, 1986; Boudreau i in., 2001).

Gdyby na proces opracowania skal spojrzeć wynikowo, to wyróżnić można jego cztery efekty, a walidacja odbywa się pomiędzy drugim a trzecim⁴ (rys. 2).

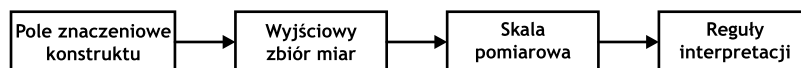
Rygor polega na ścisłym stosowaniu sekwencji etapów bez względu na stopień trudności ich realizacji (Czakon, 2014). Dlatego wzorcowe procedury⁵ opracowywania skal pomiarowych przedstawiają kolejność realizacji kluczowych czynności⁶. Można zauważyć, że walidacja jest wpleciona w procedurę opracowania skali, bowiem występuje w etapach 1, 2, 4, 6 i 7. Realizowana jest zarówno odnośnie do teoretycznego etapu opracowania skal (walidacja treściowa i fasadowa), jak i danych empirycznych (walidacja konstruktów i nomologiczna).

Rekomendacje odnośnie do reprezentowania konstruktów podkreślają uproszczenia literatury psychometrycznej, szczególnie widoczne w zakresie głębokości i szerokości konstruktów istotnych w badaniach nauk o zarządzaniu, gdzie występują (Bagozzi, Edwards, 1998): (1) cztery poziomy hierarchiczne głębokości wskaźników: pojedyncze miary, podzbiory miar, pojedyncze skale, podzbiory skal; (2) trzy poziomy hierarchiczne konstruktów: komponenty, płaszczyzny oraz globalne konstrukty.

Wobec tego badacze nauk o zarządzaniu stosują różne modele, w zależności od tego czy poszczególne miary są pojedyncze (modele całkowicie lub częściowo zdezagregowane), czy też traktowane są łącznie jako suma albo średnia (modele częściowo lub całkowicie zagregowane). Modele zdezagregowane są przydatne do szczegółowego badania składowych konstruktów oraz uzyskania drobiazgowej informacji, podczas gdy modele zagregowane do uchwycenia znaczenia całej skali (Bagozzi, Edwards, 1998).

Walidacja treściowa

Wnaukach społecznych przyjęte jest, by miary uzyskały walidację treściową, zanim przystąpi się do innych procedur walidacji (Rossiter, 2008). Dokonuje się jej, dowodząc, że elementy narzędzia pomiarowego są próbka uniwersum interesującego badacza (Cronbach, Meehl, 1955). Próbką ta powinna spełniać dwa kryteria



Rys. 2. Etapy opracowania skal pomiarowych
Źródło: opracowanie własne

(Haynes i in., 1995, s. 238): istotności, tj. odzwierciedlać wszystkie istotne wymiary konstruktów; reprezentatywności, tj. zawierać elementy wystarczające do wnioskowania o całym konstrukcie.

Walidacja treściowa odnosi się do stopnia, w jakim skala pomiarowa odzwierciedla wybrane pole znaczeniowe konstruktów, czyli jego treści (Hoskinsson i in., 1993). Przyczynia się więc do redukcji problemów semantycznych, ogranicza różnorodność terminologiczną podobnych pojęć oraz ułatwia integrację ustaleń empirycznych w nurtach badań (Venkatraman, Grant, 1986). Ustala się ją dedukcyjnie, poprzez zdefiniowanie zbioru miar opisujących uniwersum interesujące badacza oraz systematyczne próbkowanie w ramach tego uniwersum, by opracować skalę (Cronbach, Meehl, 1955). Ocena trafności treściowej jest więc procedurą osadzoną w teorii, nieopartą na badaniu empirycznym (Robins, Wiersema, 2003). Wymaga głębokiego rozpoznania teorii oraz precyzyjnej konceptualizacji w danym kontekście (Garver, Mentzer, 1999). Kryteria oceny treściowej narzędzia pomiarowego stosowane w psychologii obejmują (Haynes i in., 1995, s. 243): zbiór elementów wyjściowych skali pomiarowej, precyzję sformułowań, format odpowiedzi na elementy, sekwencję elementów, czasowe parametry odpowiedzi i in.

Walidacja treściowa zalicza się do jakościowych typów walidacji (Mitręga i in., 2012), w których rola ekspertów jest kluczowa (Hoskinsson i in., 1993), choć trwa debata czy oprócz badacza udział innego eksperta jest niezbędny, czy też nie (Rossiter, 2008). Wobec burzliwego rozwoju technik analizy statystycznej, przy jednoczesnym niedorozwoju nieilościowych technik badań, brakuje obecnie standardowej czy powszechnie akceptowanej procedury walidacji treściowej (Robins, Wiersema, 2003), a pomiędzy zwolennikami poszczególnych procedur trwa żywiołowa debata (Diamantopoulos i in., 2008).

Walidacja treściowa jest ważna dlatego, że ma istotne konsekwencje dla wnioskowania z narzędzia pomiarowego. Brak trafności treściowej przejawiać się może (Haynes i in., 1995, s. 241): (1) obniżeniem zdolności wyjaśniania narzędzia pomiarowego; (2) zniekształceniem znaczenia poszczególnych wymiarów konstruktów prowadzącym do ich przecenienia, niedocenienia lub niedostrzeżenia; (3) zniekształceniem latentnej struktury czynnikowej konstruktów ze względu na nieadekwatny dobór miar w skali; (4) zniekształceniem oceny zagregowanej wynikającej z narzędzia pomiarowego (np. zaufania, lojalności, zaangażowania); (5) zniekształceniem oceny parametrów, związków przyczynowych, predykcji i diagnozy.

Nie sposób nie zauważyć, że walidacja treściowa ma fundamentalne znaczenie dla badań⁷, a przy tym może być niezależna od walidacji konstruktów. Nawet trafne konstrukty mogą okazać się nietrafne treściowo wtedy, gdy brakuje w nich wymiarów, inne są nadreprezentowane, a jeszcze inne niedoreprezentowane (Bagozzi i in., 1991).

Podstawowym wydaje się zagadnienie określenia adekwatnej do celu badania liczby miar w narzędziu pomiarowym. Walidacja fasadowa zwykle prowadzi do redukcji liczby miar, a techniki puryfikacji statystycznej do dalszej eliminacji zbędnych miar (Rossiter, 2008). Na

etapie projektowania i walidacji narzędzia pomiarowego trudno jest jednoznacznie ustalić liczbę wymiarów oraz liczbę miar cech. Literatura nie rozwija tego wątku, debata jest szczątkowa, a zastępuje ją zasada oszczędności⁸ (*parsimony*), tj. używania tylko tylu miar, ile jest niezbędne dla uzyskania rzetelnego pomiaru (Diamantopoulos, Sigauw, 2006).

Walidacja treściowa przyciąga wiele uwagi w psychologii, tradycyjnie skupionej na psychometrii, stąd opracowane w tej dyscyplinie rekomendacje odnośnie do procedury mogą być przydatne w rozwoju praktyk badaczy nauk o zarządzaniu. Podstawowe zasady walidacji treściowej obejmują (Haynes i in., 1995):

1. Precyzyjne definiowanie pola znaczeniowego oraz wymiarów konstruktów przed przystąpieniem do innych typów walidacji.
2. Poddanie wszystkich elementów narzędzia badawczego walidacji treściowej.
3. Wykorzystanie próby populacji oraz ekspertów do wstępnego generowania elementów narzędzia pomiarowego.
4. Wykorzystanie sędziów kompetentnych, a także kwantyfikowanie ich oceny z wykorzystaniem sformalizowanych skal.
5. Ocenę proporcji reprezentatywności miar w skali.
6. Sprawozdawanie z walidacji treściowej w publikacjach przedstawiających nowe narzędzia badawcze.
7. Stosowanie analizy statystycznej dla puryfikacji narzędzia badawczego.

Zastosowanie powyższych kroków pozwala na walidację treściową skali, ale niezbędne jest przedłożenie opracowanego zestawu miar i sformułowań celowo dobranym respondentom. Pozwala to zgromadzić subiektywne opinie oraz komentarze odnośnie do sformułowań, liczby miar (zbyt dużej lub zbyt małej).

Walidacja fasadowa

Uzupełnieniem walidacji treściowej jest walidacja fasadowa, stąd obydwa te rodzaje walidacji nazywane bywają walidacją zmiennych⁹ (Bacharach, 1989). Odnosi się do fundamentalnego problemu, czy miara jest rozumiana przez respondenta zgodnie z intencjami badacza i jest niezbędna dla upewnienia się, czy zastosowana skala może być uznana za zawierającą alternatywne wskaźniki tego samego konstruktów (Gerbing, Anderson, 1988).

Walidację fasadową rozumie się jako stopień, w którym respondenci lub eksperci oceniają elementy narzędzia pomiaru jako odpowiednie do celów badania oraz stosowanych konstruktów (Hardesty, Bearden, 2004). Wbrew nieco pejoratywnej nazwie walidacja fasadowa jest niezbędnym warunkiem stosowania narzędzia, bowiem brak walidacji fasadowej uniemożliwia walidację konstruktów. Stanowi odrębny typ walidacji, który wprawdzie nie zastąpi walidacji treściowej, konstruktów czy nomologicznej (Nevo, 1985), ale ta relacja jest wzajemna, bowiem fasadowej walidacji pozostałe procedury zastąpić nie mogą.

Definiowanie walidacji fasadowej wymaga doprecyzowania czterech elementów: osoby oceniającej, przedmiotu

oceny, sposobu oceny, uwarunkowań kontekstowych (Nevo, 1985). Doprecyzowano je dla potrzeb badań nauk o zarządzaniu jako: obiekt oceny; cecha, która podlega ratingowi; respondent, który dostarcza ratingu (Rossiter, 2008). Inaczej niż w badaniach psychologicznych czy edukacji, gdzie nie rekomenduje się udziału ekspertów (Nevo, 1985), nauki o zarządzaniu standardowo się do nich odwołują (Govindarajan, Kopalle, 2002). Uznany sposobem walidacji fasadowej jest przedkładanie doświadczonym menedżerom lub ekspertom akademickim (Hardesty, Bearden, 2004): (1) projektowanego narzędzia w celu oceny każdego z elementów na trzystopniowej skali: dobrze odzwierciedla, średnio odzwierciedla, nie odzwierciedla badanej cechy; (2) definicji konstruktów, jego wymiarów oraz zestawu proponowanych elementów narzędzia badawczego, w celu uzyskania przyporządkowania, które eksperci uznają za trafne. Wystąpić może pula elementów nigdzie nieprzypisanych, stąd wprowadza się kategorię „inne”¹⁰.

Rezultatem walidacji fasadowej jest eliminacja niejasnych, zbędnych lub niepowiązanych elementów lub pozytywna opinia ekspertów o liście elementów do każdego z wymiarów konstruktów, ewentualnie przypisanie elementów do wymiarów konstruktów (Hardesty, Bearden, 2004). Decyzje o ostatecznym kształcie narzędzia podejmować można wówczas, gdy eksperci wyrażą zbieżne poglądy, jednak literatura tę zbieżność proponuje na bardzo różnych poziomach, od 60% do 80% opinii ekspertów (Hardesty, Bearden, 2004).

Mimo znaczenia walidacji fasadowej dla rygoru pomiaru nie jest ona standardowo przeprowadzana¹¹. Wobec tego formułuje się rekomendację, obserwowalną tendencję w publikacjach oraz jednoznaczne oczekiwanie recenzentów najbardziej wpływowych czasopism, aby walidację fasadową przeprowadzać, a z jej przebiegu sprawozdawać w artykułach. Szczegółowe sprawozdania z procedur walidacji fasadowej wskazują na kilkuetapowe postępowanie, np.: 5 badaczy, następnie 35 menedżerów wyższego szczebla, następnie 128 menedżerów wyższego szczebla (Govindarajan, Kopalle, 2002). Taka sekwencyjna walidacja fasadowa (Klimas, 2019) pozwala zredukować wpływ uwarunkowań kontekstowych na wyniki, a także uzyskać wiarygodne wyniki odnośnie do tego, czy narzędzie wydaje się mierzyć to, co w intencji badacza mierzyć powinno.

Walidacja konstruktów

Celem walidacji konstruktów jest ocena błędów pomiaru, losowych i systematycznych (Campbell, Friske, 1959), oraz korekta skali (Bagozzi i in., 1991). Wobec tego znaczenie walidacji konstruktów jest zupełnie podstawowe dla rozwoju każdej dyscypliny nauki, a troska o nią wyraża dążenie do podnoszenia rygoru badań (Boyd i in., 2005).

Przedmiotem walidacji konstruktów są dane i oceny oparte na danych (Haynes i in., 1995). Pomiarowi podlega zestaw miar cech, a nie sam konstrukt. W idealnym przypadku miara odzwierciedla zarówno kierunek zmienności, jak i wielkość wszystkich cech konstruktów, oraz tylko

tych cech konstruktów, które miała mierzyć (Peter, 1981, s. 143). Innymi słowy, walidacja konstruktów zależy od stopnia, w jakim konstruktowi odpowiadają miary jego cech (Churchill, 1979). Oznacza to zgodę na niedoskonałość pomiaru pod warunkiem, że skala mierzy reprezentatywną próbkę cech konstruktów, nie jest zanieczyszczona elementami pola znaczeniowego innych konstruktów ani błędami pomiaru (Peter, 1981).

Miara jest trafna wtedy, gdy różnice w uzyskanych wartościach odpowiadają różnicom mierzonych cech i żadnemu innemu czynnikowi (Churchill, 1979, s. 65). Trafna skala pomiarowa powinna wykazywać następujące własności statystyczne: jednowymiarowość, rzetelność, trafność konwergentną, trafność dywergentną oraz trafność predykcyjną (Garver, Mentzer, 1999, s. 34). Jednowymiarowość przejawia się w tym, że elementy skali mierzące ten sam jednowymiarowy konstrukt będą wykazywały kowariancję. Prosty sposób doskonalenia skali pod względem jednowymiarowości jest eliminacja pojedynczych miar albo rozbitcie konstruktów na dwa lub większą liczbę konstruktów (Hulland, 1999). Rzetelność skali pozwala sądzić, że jej przyszłe użycie da wartościowe wyniki, wolne od błędów pomiaru. Trafność konwergentna oznacza, że miary danego konstruktów są skorelowane ze zmienną latentną, którą powinny mierzyć. Trafność dywergentna uzupełnia konwergentną, oceniając, czy miary danego konstruktów różnią się od miar innych konstruktów ujętych w tym samym modelu.

Współczesne¹² instrumentarium nauk o zarządzaniu wykorzystuje bardziej zaawansowane narzędzia modelowania równań strukturalnych (Garver, Mentzer, 1999, s. 36). *Confirmatory Factor Analysis* pozwala ocenić w sposób rygorystyczny jednowymiarowość (Gerbing, Anderson, 1988), a także pozostałe cechy statystyczne skali. Model pomiarowy jest użytecznym narzędziem oceny trafności konstruktów (Jöreskog, Sörbom, 1982). Modele pomiarowe powinny spełniać kryteria dopasowania, które są przedmiotem dyskusji i ewolucji poglądów, a także wytycznych stosowania i sprawozdawania z konkretnych miar oraz ich oczekiwanych poziomów (Garver, Mentzer, 1999; Hulland, 1999; Gefen i in., 2011).

W celu obniżenia i kontroli błędów pomiaru, czyli podnoszenia poziomu trafności skal stosowanych w naukach o zarządzaniu rekomenduje się (Boyd i in., 2005):

1. Stosowanie wielu miar zamiast pojedynczych dla badanych konstruktów.
2. Stosowanie skal pomiarowych w badaniach.
3. Rozwijanie nowych skal pomiarowych, zważywszy na istotne obciążenia wcześniej opublikowanych miar, z zastosowaniem kryterium rzetelności.
4. Zwiększanie mocy statystycznej prób badawczych.
5. Przyjęcie wysokich standardów przez recenzentów, redaktorów czasopism i autorów prac.

Walidacji konstruktów należy dokonywać zawsze wtedy, gdy nie istnieją kryteria lub zbiory miar powszechnie uznane jako całkowicie adekwatne do pomiaru danej cechy (Cronbach, Meehl, 1955). Zatem każde nowe pojęcie, odpowiadający mu konstrukt i skala powinny zostać poddane walidacji¹³. Podkreślić trzeba dynamiczny¹⁴

wymiar walidacji konstruktów w naukach o zarządzaniu. Odzwierciedla ona subiektywne przekonania, stabilne w ograniczonym czasie i w niektórych kontekstach, co wymaga od badaczy walidacji konstruktów na zgromadzonych w danym badaniu danych (Peter, Churchill, 1986, s. 10). Dlatego uzasadnione jest ciągle doskonalenie skal (Zaichkowsky, 2015), a także systematyczne sprawozdawanie z oceny trafności w publikacjach.

Walidacja nomologiczna

Walidacja konstruktów przebiega w teoretycznej próżni, skupiając się na cechach miar, by ustalić ich wewnętrzną trafność. Walidacja nomologiczna odpowiada na podstawowe pytanie, czy konstrukt pozwala stawiać trafne predykcje. Konstrukty występują w sieci mniej lub bardziej wprost wyrażonych relacji względem innych konstruktów oraz zjawisk, które odzwierciedlają (Suddaby, 2010). Tworzą swoiste sieci powiązań nazywane przez psychologów sieciami nomologicznymi (Cronbach, Meehl, 1955). Dlatego niektórzy badacze włączają walidację nomologiczną w zakres walidacji konstruktów (Pennings, Smidts, 2000), choć inni uważają ją za odrębny rodzaj walidacji (Gatignon i in., 2002). Warunkiem koniecznym naukowej akceptowalności konstruktów jest usytuowanie go w sieci nomologicznej, w której co najmniej niektóre prawa dotyczą obserwowalnych zmiennych (Cronbach, Meehl, 1955, s. 22).

Walidacja nomologiczna polega na badaniu zachowania konstruktów oraz miar w postaci testowania formalnych hipotez wyprowadzonych z sieci nomologicznej (Peter, Churchill, 1986, s. 7). Opiera się na dwóch filarach: teoretycznych predykcjach oraz testach empirycznych. W ścisłym znaczeniu walidacja predykcyjna stanowi element walidacji nomologicznej, bowiem odnosi się do stopnia, w jakim pojedyncza zmienna jest związana z inną zmienną jako antecedenca lub efekt (Hagger i in., 2017). Wobec tego walidacja nomologiczna wymaga testowania wielu relacji antecedencji i konsekwencji w złożonym systemie nomologicznym (Bagozzi, 1981). Innymi słowy, predykcje to pojedyncze relacje w modelu, a testowanie predykcji całego modelu prowadzi do walidacji nomologicznej.

Współczesna krytyka praktyk walidacji¹⁵ w psychologii wskazuje na nadmierną koncentrację na testowaniu pojedynczych relacji wewnątrz teorii, a nie formalnych testów teorii jako całości (Hagger i in., 2017). Sytuacja w badaniach nauk o zarządzaniu jest podobna. Cztery dekady temu stwierdzono faktyczny brak testów nomologicznych w badaniach marketingowych (Peter, 1981), a stan ten nadal trwa w badaniach systemów informacyjnych zarządzania (Boudreau i in., 2001) czy zarządzania strategicznego (Boyd i in., 2005).

Dlatego też ważne jest zarówno wyjaśnienie oczekiwanych związków (Danneels, 2016), jak i wskazanie szerszego zakresu sieci nomologicznej (Pennings, Smidts, 2000). Testy empiryczne opierają się na mniej, np. analiza korelacji, analiza regresji (Gatignon i in., 2002; Mitrenga i in., 2012), lub bardziej wyrafinowanych, np. modelowanie równań strukturalnych (Pennings, Smidts, 2000),

narzędziach statystycznych. Zaletą modelowania równań strukturalnych jest testowanie zależności przyczynowych (Jöreskog, Sörbom, 1982), nieosiągalne mniej zaawansowanymi technikami.

Testy nomologiczne wymagają ostrożnej interpretacji, bowiem związek miar z postawami może dawać inne wyniki niż miar z faktycznymi zachowaniami (Pennings, Smidts, 2000), co tłumaczyć można celami oraz polem recepcji skal psychometrycznych. Skupiają się one na percepcji respondenta, a nie na faktycznych zachowaniach czy bezpośrednio mierzalnych zjawiskach¹⁶. Badanie faktycznych wyników wymaga zgromadzenia miar w dwóch różnych okresach (Danneels, 2016), a także zastosowania miar obiektywnych (Govindarajan, Kopalle, 2006). W znaczącym stopniu podnosi to koszty, czasochłonność oraz stopień złożoności badań, a badacze nie zawsze mają taki luksus (Venkatraman, Grant, 1986, s. 84). Praktyka badań w naukach o zarządzaniu wykazuje trwałą tendencję wzrostu rygoru metodologicznego w poszczególnych subdyscyplinach (Czakon, 2017). O ile trzy dekady temu wymóg walidacji nomologicznej nie był kluczowy w zarządzaniu strategicznym, bowiem walidacja konstruktów stanowiła pilniejszą potrzebę, o tyle w ostatnich latach liczba prac przedstawiających testy nomologiczne rośnie (Danneels, 2016).

Podsumowanie

Walidacja jest niezbędną procedurą w badaniach nauk społecznych, ponieważ prowadzi do trafnego i rzetelnego pomiaru abstrakcyjnych pojęć. Warunkuje narastanie rzetelnej wiedzy, pomiar zjawisk, ich szczegółową analizę w różnych kontekstach i na różnych poziomach analizy. Zachodzi też relacja odwrotna, bowiem brak walidacji uniemożliwia rozwój dyscyplin w dziedzinie nauk społecznych. Znaczenie walidacji podkreśla ustalone kilka dekad temu stanowisko, iż pomiar konstruktów powinien być centralną częścią projektów badawczych (Venkatraman, Grant, 1986).

Uwarunkowania rozwoju subdyscypliny zarządzania strategicznego wydają się nadal aktualne. Są nimi: trend operacjonalizacji i pomiaru konstruktów, znaczenie interpretacji miar dla interpretacji substancjalnych zależności pojęć, potrzeba doskonalenia walorów opisowych badań, pomiar strategii oraz implikacje praktyczne (Venkatraman, Grant, 1986). Wynikają stąd upowszechniające się pośród wielu subdyscyplin nauk o zarządzaniu rekomendacje (Boudreau i in., 2001):

- 1) stosowania pilotaży i testów wstępnych w celu oceny jak największego zakresu walidacji skal,
- 2) umieszczania w publikacjach sekcji poświęconej walidacji,
- 3) stosowania wcześniej walidowanych narzędzi, gdy tylko jest to możliwe,
- 4) walidacji przy wprowadzaniu nowych konstruktów.

Niniejszy artykuł przyczynia się do rozwoju rygoru metodologicznego badań w naukach o zarządzaniu na kilka sposobów. Po pierwsze, przedstawia cały zakres procedury walidacji zaczerpnięty z chronologicznie pierwszych rozważań psychometrycznych, rozwinięty o dojrzałą praktykę marketingu, a także odzwierciedlający stan wymagań w naj-

lepszycy czasopismach nauk o zarządzaniu. Po drugie, omówieniu każdego z rodzajów oceny trafności: treściowej, fasadowej konstruktu i nomologicznej, towarzyszy osadzenie w psychometrycznych korzeniach, obraz dotychczasowej praktyki oraz rekomendowane kierunki jej doskonalenia. Po trzecie, literatura referencyjna może się okazać pożyteczna do pogłębienia i rozwinięcia kompetencji badaczy nauk o zarządzaniu.

prof. dr hab. Wojciech Czakon
Uniwersytet Jagielloński w Krakowie
Wydział Zarządzania i Komunikacji Społecznej
ORCID: 0000-0002-8329-0829
e-mail: wojciech.czakon@uj.edu.pl

Przypisy

- 1) Artykuł przygotowano w ramach realizacji projektu badawczego pt. „Strategiczna krótkowzroczność a wyniki firmy”, finansowanego przez Narodowe Centrum Nauki nr UMO-2017/27/B/HS4/01150.
- 2) Stosowane w dalszej części pojęcie konstrukt odwołuje się do używanej w literaturze anglojęzycznej konwencji terminologicznej (ang. *construct*). Konstrukty to sokratejskie „pojęcia”, platońskie „idee”, nawiązujące do korespondencyjnej koncepcji prawdy, tj. zgodności pojęć z rzeczywistością.
- 3) Narzędzie badawcze obejmuje: instrukcje, pojedyncze miary, formaty odpowiedzi oraz skale odpowiedzi (Rossiter, 2008).
- 4) O potrzebie rozwoju nauk o zarządzaniu świadczy bardzo mały zbiór skal, które zostały powszechnie uznane i opracowano do nich reguły interpretacji uzyskanych wartości. Zamiast tego obserwuje się raczej powtarzane, z różnym skutkiem, walidacje skal w kolejnych badaniach.
- 5) Wyróżnia się dwa podejścia do budowania skali pomiarowej: formatywne i refleksywne, w zależności od kierunku zależności przyczynowej pomiędzy miarą a zmienną latentną. Jeśli postrzegać miary jako funkcję zmiennej latentnej, to odzwierciedlają (ang. *reflect*) one jej zmiany. I odwrotnie, jeśli zmiany wskaźników powodują zmiany zmiennej latentnej (ang. *form*), to podejście uważa się za formatywne. Dyskusja dotycząca procedury i skutków postępowania w zależności od przyjętego podejścia wykracza poza ramy niniejszego opracowania. Wskazuję jednak, że prowadzą one do różnych rezultatów zarówno w zakresie walidacji treściowej, oszczędności, jak i predykcji (Diamantopoulos, Siguaw, 2006).
- 6) Najczęściej akceptowana jest klasyczna procedura G.A. Churchilla (1979), która obejmuje osiem etapów.
- 7) Walidacja treściowa jest remedium na nieskrępowane mnożenie pojęć i zmienianie ich zakresu znaczeniowego, prowadzące do krytycznej obserwacji „dżungli teorii zarządzania”.
- 8) Rozwinięcie argumentu oszczędności prowadzi do rozwoju pierwotnie opracowanych i walidowanych skal w kierunku redukcji liczby miar, nawet o 50% (Zaichkowsky, 2015). Liczba miar uwarunkowana jest także przewidywanymi technikami analizy danych, które wymagają określonego stosunku liczby wskaźników dla konstruktów.
- 9) Stosowane tu pojęcie zmiennych odnosi się do obserwowalnej jednostki, która może przyjmować dwie lub więcej wartości.

- 10) Uwarunkowania kontekstowe badania, takie jak czas, sposób przedłożenie narzędzia, stosunek emocjonalny ekspertów względem narzędzia mogą mieć wpływ na wynik walidacji fasadowej.
- 11) W badaniach konsumenta sprawozdano z niej tylko w co piątym badaniu (Hardesty, Bearden, 2004); w mniej niż połowie badań empirycznych systemów informacyjnych zarządzania jako badanie pilotażowe (Boudreau i in., 2001). Od wyników tych odbiega optymistyczna ocena badań przedsiębiorczości, w których dostrzeżono ponad 90% informacji o walidacji fasadowej (Chandler, Lyon, 2001).
- 12) Istnieje wiele procedur oceny trafności konstruktu opracowanych na potrzeby psychometrii (Cronbach, Meehl, 1955): badanie różnic grup; macierze korelacji i analiza czynnikowa; ocena struktury wewnętrznej skali; ocena stabilności; badanie procesu i in. Klasyczne dwa kryteria: walidacji konwergentnej i dyskryminacyjnej (Campbell, Fiske, 1959) dały początek powszechnie uznawanej w psychologii, a później naukach społecznych metodzie MTMM, tj. *multi-trait multi-method*.
- 13) Psychometria przyjmuje, iż wielokrotna walidacja konstruktu przez różnych badaczy w różnych kontekstach prowadzi do uznania skali i zwalnia kolejnych badaczy z ponownych testów. Ułatwia to badania replikacyjne i stosowanie skali dla zrozumienia badanych zjawisk.
- 14) Na przykład skale innowacyjności organizacyjnej okazują się mniej lub bardziej trafne w zależności od kontekstu branżowego (Klimas, Czakon, 2018), a także ze względu na rozwój praktyki zarządzania (Hakala, 2011).
- 15) Badanie praktyk psychologów w tym zakresie wskazuje, że walidacja nomologiczna nie jest rutynową praktyką, badacze rzadko sprawozdają z testów czy formułują wnioski na podstawie testów. Rekomenduje się wobec tego każdorazowe identyfikowanie „rdzenia teoretycznego”, kluczowych relacji zmiennych, który odzwierciedla minimalny niezbędny test teorii, oraz sprawozdanie z przeprowadzonych testów nomologicznych.
- 16) Ujawniają zatem mikrofundamenty, poznawcze uwarunkowania decyzji i działań menedżerów, ale związek mikrofundamentów z rezultatem niekoniecznie jest bezpośredni.

Bibliografia

- [1] Bacharach S.B. (1989), *Organizational Theories: Some Criteria for Evaluation*, „Academy of Management Review”, Vol. 14, No. 4, pp. 496–515.
- [2] Bagozzi R.P., Yi Y., Phillips L.W. (1991), *Assessing Construct Validity in Organizational Research*, „Administrative Science Quarterly”, Vol. 39, pp. 421–458.
- [3] Bagozzi R.P., Edwards J.R. (1998), *A General Approach for Representing Constructs in Organizational Research*, „Organizational Research Methods”, Vol. 1, No. 1, pp. 45–87.
- [4] Boudreau M.C., Gefen D., Straub D.W. (2001), *Validation in Information Systems Research: A State-of-the-art Assessment*, „MIS Quarterly”, Vol. 25, No. 1, pp. 1–16.
- [5] Boyd B.K., Gove S., Hitt M.A. (2005), *Construct Measurement in Strategic Management Research: Illusion or Reality?*, „Strategic Management Journal”, Vol. 26, No. 3, pp. 239–257.
- [6] Campbell D.T., Fiske D.W. (1959), *Convergent and Discriminant Validation by the Multitrait-multimethod Matrix*, „Psychological Bulletin”, Vol. 56, No. 2, pp. 81–105.

- [7] Chandler A.D. (1977), *The visible hand: The Managerial Revolution in American Business*, Cambridge MA, Harvard Belknap.
- [8] Chandler G.N., Lyon D.W. (2001), *Issues of Research Design and Construct Measurement in Entrepreneurship Research: The Past Decade*, „Entrepreneurship Theory and Practice”, Vol. 25, No. 4, pp. 101–113.
- [9] Chen M.J., Su K.H., Tsai W. (2007), *Competitive Tension: The Awareness-motivation-capability Perspective*, „Academy of Management Journal”, Vol. 50, No. 1, pp. 101–118.
- [10] Churchill G.A. Jr (1979), *A Paradigm for Developing Better Measures of Marketing Constructs*, „Journal of Marketing Research”, Vol. 16, pp. 64–73.
- [11] Cronbach L.J., Meehl P.E. (1955), *Construct Validity in Psychological Tests*, „Psychological Bulletin”, Vol. 52, No. 4, pp. 281–302.
- [12] Czakon W. (2014), *Kryteria oceny rygoru metodologicznego badań w naukach o zarządzaniu*, „Organizacja i Kierowanie”, Vol. 1, Nr 161, s. 51–62.
- [13] Czakon W. (2015), *Mikrofundamenty rozwoju sieci międzyorganizacyjnych*, „Marketing i Rynek”, (5 (CD)), s. 798–807.
- [14] Czakon W. (2017), *Tworzenie teorii w naukach o zarządzaniu*, [w:] A. Sopińska, P. Wachowiak, *Wyzwania współczesnego zarządzania strategicznego*, Oficyna Wydawnicza SGH w Warszawie, Warszawa, s. 143–160.
- [15] Danneels E. (2016), *Survey Measures of First-and Second-order Competences*, „Strategic Management Journal”, Vol. 37, No. 10, pp. 2174–2188.
- [16] Diamantopoulos A., Siguaw J.A. (2006), *Formative Versus Reflective Indicators in Organizational Measure Development: A Comparison and Empirical Illustration*, „British Journal of Management”, Vol. 17, No. 4, pp. 263–282.
- [17] Diamantopoulos A., Siguaw J.A., Cadogan J.W. (2008), *Measuring Abstract Constructs in Management and Organizational Research: The Case of Export Coordination*, „British Journal of Management”, Vol. 19, No. 4, pp. 389–395.
- [18] Garver M.S., Mentzer J.T. (1999), *Logistics Research Methods: Employing Structural Equation Modeling to Test for Construct Validity*, „Journal of Business Logistics”, Vol. 20, No. 1, pp. 33–57.
- [19] Gatignon H., Tushman M.L., Smith W., Anderson, P. (2002), *A Structural Approach to Assessing Innovation: Construct Development of Innovation Locus, Type, and Characteristics*, „Management Science”, Vol. 48, No. 9, pp. 1103–1122.
- [20] Gavetti G. (2005), *Cognition and Hierarchy: Rethinking the Microfoundations of Capabilities’ Development*, „Organization Science”, Vol. 16, No. 6, pp. 599–617.
- [21] Gefen D., Rigdon E.E., Straub D. (2011), *Editor’s Comments: An Update and Extension to SEM Guidelines for Administrative and Social Science Research*, „MIS Quarterly”, Vol. 35, No. 2, pp. iii–xiv.
- [22] Gerbing D.W., Anderson J.C. (1988), *An Updated Paradigm for Scale Development Incorporating Unidimensionality and its Assessment*, „Journal of Marketing Research”, Vol. 25, No. 2, pp. 186–192.
- [23] Glińska-Neweś A. (2017), *Pozytywne relacje interpersonalne w zarządzaniu*, Wydawnictwo Naukowe Uniwersytetu Mikołaja Kopernika, Toruń.
- [24] Govindarajan V., Kopalle P.K. (2006), *Disruptiveness of Innovations: Measurement and an Assessment of Reliability and Validity*, „Strategic Management Journal”, Vol. 27, No. 2, pp. 189–199.
- [25] Hagger M.S., Gucciardi D.F., Chatzisarantis N.L. (2017), *On Nomological Validity and Auxiliary Assumptions: The Importance of Simultaneously Testing Effects in Social Cognitive Theories Applied to Health Behavior and Some Guidelines*, „Frontiers in Psychology”, Vol. 8, No. 1933, pp. 1–12.
- [26] Hakala H. (2011), *Strategic Orientations in Management Literature: Three Approaches to Understanding the Interaction between Market, Technology, Entrepreneurial and Learning Orientations*, „International Journal of Management Reviews”, Vol. 13, No. 2, pp. 199–217.
- [27] Hardesty D.M., Bearden, W.O. (2004), *The Use of Expert Judges in Scale Development: Implications for Improving Face Validity of Measures of Unobservable Constructs*, „Journal of Business Research”, Vol. 57, No. 2, pp. 98–107.
- [28] Hatchuel A. (2005), *Towards an Epistemology of Collective Action: Management Research as a Responsive and Actionable Discipline*, „European Management Review”, Vol. 2, No. 1, pp. 36–47.
- [29] Haynes S.N., Richard D., Kubany E.S. (1995), *Content Validity in Psychological Assessment: A Functional Approach to Concepts and Methods*, „Psychological Assessment”, Vol. 7, No. 3, p. 238.
- [30] Hoskisson R.E., Hitt M.A., Johnson R.A., Moesel D.D. (1993), *Construct Validity of an Objective (Entropy) Categorical Measure of Diversification Strategy*, „Strategic Management Journal”, Vol. 14, No. 3, pp. 215–235.
- [31] Hulland J. (1999), *Use of Partial Least Squares (PLS) in Strategic Management Research: A Review of Four Recent Studies*, „Strategic Management Journal”, Vol. 20, No. 2, pp. 195–204.
- [32] Jöreskog K.G., Sörbom D. (1982), *Recent Developments in Structural Equation Modeling*, „Journal of Marketing Research”, Vol. 19, pp. 404–416.
- [33] Kawa A. (2017), *Orientacja sieciowa przedsiębiorstw branży usług logistycznych*, Wydawnictwo Uniwersytetu Ekonomicznego w Poznaniu, Poznań.
- [34] Klimas P., Czakon W. (2018), *Organizational Innovativeness and Coopetition: A Study of Video Game Developers*, „Review of Managerial Science”, Vol. 12, No. 2, pp. 469–497.
- [35] Klimas P. (2019), *Relacje współtworzenia innowacji w ekosystemach. Kontekst ekosystemu gamingowego*, Wyd. C.H. Beck, Warszawa
- [36] Mitreğa M., Forkmann S., Ramos C., Henneberg S.C. (2012), *Networking Capability in Business Relationships – Concept and Scale Development*, „Industrial Marketing Management”, Vol. 41, No. 5, pp. 739–751.
- [37] Nevo B. (1985), *Face Validity Revisited*, „Journal of Educational Measurement”, Vol. 22, No. 4, pp. 287–293.
- [38] Nunnally J.C. (1975), *Psychometric Theory – 25 Years Ago and Now*, „Educational Researcher”, Vol. 4, No. 10, pp. 7–21.
- [39] Nunnally J.C. (1978), *Psychometric Theory*, 2nd Ed., McGraw-Hill Book Company, New York, NY.
- [40] O’Leary-Kelly S.W., Vokurka R.J. (1998), *The Empirical Assessment of Construct Validity*, „Journal of Operations Management”, Vol. 16, No. 4, pp. 387–405.



- [41] Pennings J.M., Smidts A. (2000), *Assessing the Construct Validity of Risk Attitude*, „Management Science”, Vol. 46, No. 10, pp. 1337–1348.
- [42] Peter J.P. (1981), *Construct validity: A Review of Basic Issues and Marketing Practices*, „Journal of Marketing Research”, Vol. 18, No. 2, pp. 133–145.
- [43] Peter J.P., Churchill G.A. Jr (1986), *Relationships Among Research Design Choices and Psychometric Properties of Rating Scales: A Meta-analysis*, „Journal of Marketing Research”, Vol. 23, No. 1, pp. 1–10.
- [44] Robins J.A., Wiersema M.F. (2003), *The Measurement of Corporate Portfolio Strategy: Analysis of the Content Validity of Related Diversification Indexes*, „Strategic Management Journal”, Vol. 24, No. 1, pp. 39–59.
- [45] Rossiter J.R. (2008), *Content Validity of Measures of Abstract Constructs in Management and Organizational Research*, „British Journal of Management”, Vol. 19, No. 4, pp. 380–388.
- [46] Suddaby R. (2010), *Editor’s Comments: Construct Clarity in Theories of Management and Organization*, „Academy of Management Review”, Vol. 35, No. 3, pp. 346–357.
- [47] Venkatraman N. (1989), *Strategic Orientation of Business Enterprises: The Construct, Dimensionality, and Measurement*, „Management Science”, Vol. 35, No. 8, pp. 942–962.
- [48] Venkatraman N., Grant J.H. (1986), *Construct Measurement in Organizational Strategy Research: A Critique and Proposal*, „Academy of Management Review”, Vol. 11, No. 1, pp. 71–87.
- [49] Zaichkowsky J.L. (2015), *Review and Reply to: „Why You Must Use my C-OAR-SE Method”* „Australasian Marketing Journal”, Vol. 3, No. 23, pp. 261–262.
- [50] Zakrzewska-Bielawska A. (2018), *Strategie rozwoju przedsiębiorstw. Nowe spojrzenie*, Polskie Wydawnictwo Ekonomiczne, Warszawa.

Measurement Scale Validation in Management Research

Summary

The development of the management theory requires the formulation of construct interrelations, which allows for predicting key variables behavior and helps taking better decisions. It is dependent upon construct measurement. This paper locates validity assessment in the process of scale development. Four validity procedures have been outlined: content, face, construct and nomological ones. Their respective contents, psychometric grounds, current management literature practice and recommendations have been formulated.

Keywords

validity, construct, methodology
