

Badania metaanalityczne

1. Wprowadzenie

Badania empiryczne stanowią nierozzerwalną część działań naukowych w obszarze kształcenia i wychowania. Jedną z charakterystycznych cech rozwoju badań ostatniego stulecia w omawianym obszarze jest ogromny zasięg badań empirycznych. Światowy rozwój metodologii badań pedagogicznych w ostatnich dziesięcioleciach podążał w kierunku rozwoju świadomości, dostrzegającej znaczenie i potrzebę empirycznej podbudowy badań. W czasie, gdy rozwijało się wiele nowych metod (np. badania w działaniu, nowe modele eksperymentów pedagogicznych, różnorodne złożone analizy statystyczne itp.), powstało wiele możliwości ich masowego wykorzystania. Dziś trudno wyobrazić sobie np. przedsięwzięcia w systemie szkolnym bez opierania ich na wynikach analiz rzeczywistości oświatowej, czyli badań empirycznych. Jako przykład należy wymienić dobrze rozwiniętą teorię eksperymentu szkolnego (Schulversuch) i jednocześnie wyjątkowo dobrze rozwiniętą praktykę badawczą wymienionych doświadczeń w szkolnictwie Europy Zachodniej. Podobny przykład stanowić może wprowadzanie innowacji dydaktycznych i metodycznych, zmiany programów nauczania itd. Omówione wyżej tendencje odnoszą się nie tylko do badań empirycznych, ale dotyczą wszystkich prowadzonych w pedagogice dociekań.

2. Integracja badań empirycznych

W ostatnich dziesięcioleciach szczególnie wyraźnie wystąpił problem masowości publikowania rezultatów badawczych. Uwzględniając liczby doktoratów, np. w Europie, oczywistym staje się zasięg tego obszaru badawczego. Ilość wyników badań ulega ciąglemu zwiększeniu, co w konsekwencji prowadzi do ograniczenia możliwości dokonywania ich przeglądów. Zjawisko to można dokumentować światowymi danymi z publikacji amerykańskich. W roku 1988 informacje ujawnione przez Current Index to Journals in Education – to 17000 artykułów, publikowanych w 700 czasopismach. Research in Education posiada spis 9000 tytułów. Jeszcze większe wrażenie robi obraz z obszaru ważniejszych prac naukowych. Comprehensive Dissertation Abstracts – w 1988 roku wymienia 6000 dysertacji doktorskich

z obszaru wychowania i kształcenia. Powyższe przykłady dotyczą jedynie jednego wybranego roku. Można więc otrzymać przybliżone dane ostatniego dziesięciolecia, mnożąc przedstawione wyżej wyniki przez liczbę dziesięć.

Oczywistym jest fakt, że dane te nie prezentują całości problemu. Nie zostały tu uwzględnione książki i samodzielne publikacje z obszaru kształcenia i wychowania. Nadmienić należy również, że omówione dane dotyczą jedynie publikacji w języku angielskim. Liczba publikowanych prac w świecie, które dostarczają określonych rezultatów badawczych z obszaru pedagogiki, jest nieporównywalnie większa. Francuski Centre de Documentation Sciences Humaines w Paryżu w roku 1988 przyjął 1240 naukowych czasopism z całego świata, które publikowały rezultaty badań prowadzonych w pedagogice, psychologii i pokrewnych dyscyplinach (oczywiście w różnych językach).

Badacze i użytkownicy prac naukowo-wychowawczych z obszaru wychowania i kształcenia już od dawna są świadomi problemów istniejącej sytuacji. Naukowcy podejmujący te problemy w obszarze swoich specjalności starają się dorównać kroku pojawiającym się wyzwaniom rzeczywistości. Prace naukowe – badawcze przeznaczone są przecież także dla nauczycieli, wychowawców, dyrektorów, pracowników oświaty itp. Konieczna i pilna jest więc lepsza systematyzacja tej mnogości literatury.

Od dawna już w światowej naukowej i specjalistycznej literaturze istnieją przeglądy (angielskie: review), które próbują streszczać najważniejsze naukowe osiągnięcia i wyniki o określonej problematyce. Liczne z nich stanowią integrację rezultatów szeregu badań. Ponad dwadzieścia lat temu Glass odkrył niedociągnięcia ówczesnych naukowych przeglądów (Glass 1976). Stwierdził on, że autorzy raportów i przeglądów, prace naukowe do swoich integracji wybierają w większości zupełnie przypadkowo. Rezultaty i wyniki ich opracowań są nieprecyzyjne i zbyt ogólne. Metodologiczny model takich przeglądów jest za każdym razem inny. W wyniku wymienionych błędów stwierdzenia tych przeglądów są mało rzetelne, niesystematyczne i często nieporównywalne. Ponadto autorzy przeglądów zbyt mało czytelnie opisują swoją metodologię (zwłaszcza opracowania materiału empirycznego).

Słaba przejrzystość przeglądów uniemożliwia użytkownikom dokonanie oceny adekwatności prezentowanych stwierdzeń. Większość przeglądów ma zbyt prosty model metodologiczny o małej nośności.

Główną przyczynę tego stanu Glass upatrywał w słabo rozwiniętej metodologii integracji wyników publikowanych badań. Dowodził on, że „potrzebujemy metod, które będą umożliwiały systematyczne wartościowanie wyników badań, aby można było wyciągnąć istotę poznania z wielkiego zbioru poszczególnych badań” (Glass 1976, s. 4). Według Glassa przeglądy dokonywane ówczesnie nie mogą tego zadania zrealizować wystarczająco dobrze. Dlatego też wyodrębnił on trzy typy analiz niezbędnych do progresu prac badawczych w obszarze kształcenia i wychowania:

1. pierwotna analiza,
2. wtórna analiza,
3. metaanaliza.

Pierwotna analiza stanowi podstawową czynność opracowania danych, którą zwykle wykonuje badacz, planujący badanie i zbieranie danych. Analiza wtórna jest reanalizą danych, ukierunkowaną na poszukiwanie odpowiedzi na pierwotne pytania badawcze przy pomocy doskonalszych i lepszych metod statystycznych. Wykonują ją na ogół badacze, nie będący członkami grupy badań pierwotnych. Posiadają oni jednak dostęp do materiału empirycznego. Metaanaliza to z kolei ilościowe opracowanie wyników badań, a nie danych pochodzących z pierwotnych i wtórnych analiz. Metaanalityk zatem dokonuje analiz statystycznych ilościowych wyników poszczególnych badań.

Analizie pierwotnej badacze poświęcają zwykle najwięcej uwagi i energii. Sponsorzy badań finansują ją najbardziej szczerze i najczęściej, natomiast czasopisma pedagogiczne dokładają najwięcej starań o pozyskanie artykułów, dokumentujących wyniki opisanych wyżej analiz.

Tylko niektórym analizom wtórnym poświęca się tak wiele uwagi. Jeśli nawet zdarzają się takie przypadki, wymagają one wiele trudu i wysokich kryteriów jakościowych, aby mogły rywalizować z analizami pierwotnymi. W czasie, gdy Glass jako pierwszy wyodrębnił metaanalizę (wystąpienie z okazji otrzymania funkcji prezydenta stowarzyszenia AERA), analizy wtórne i przeglądy ilościowe nie otrzymywały wysokich ocen, a tym samym oczekiwanego prestiżu. W ówczesnej literaturze metodologicznej problematyka ta nie została całkowicie i systematycznie opracowana.

Rozważania klasyfikacji wyodrębnionej przez Glassa, dotyczącej trzech analiz, wyraźnie wskazują na brak istotnej i precyzyjnej różnicy pomiędzy analizą pierwotną i wtórną. Obydwie one zajmują się początkowymi danymi empirycznymi. Pytania badawcze są takie same, różnica polega jedynie na użyciu odrębnych metod statystycznych. Nieistotne i niezasadnicze jest to, że wtórna analiza przebiega później – gdy pierwotna jest już zakończona ostatecznie, czyli zinterpretowana i posiadająca sformułowane teoretycznie wnioski. Podobnie nie zasadnicze jest również to, że pierwotnej i wtórnej analizy nie dokonują najczęściej ci sami badacze. Wymienione różnice są zbyt małe, aby można było używać określeń „pierwotna” i „wtórna”. W świetle tych rozważań granica pomiędzy dwiema pierwszymi analizami a metaanalizą jest zupełnie oczywista. Metaanaliza nie zajmuje się początkowymi danymi empirycznymi, lecz wynikami uzyskanymi z tych danych. Lepiej byłoby sporządzić klasyfikację tak, aby wyodrębniała ona:

1. pierwotną empiryczną analizę, czyli wykorzystywanie znanych metod i technik do przeprowadzenia badań zjawisk z obszaru wychowania (np. eksperyment, techniki socjometryczne, skale postaw, inwentarze osobowościowe, testy zachowań, metody projekcyjne itd.),
2. wtórną empiryczną analizę, czyli procedury, którymi nie bada się bezpośrednio zjawisk, lecz opracowuje wyniki uzyskane w analizach pierwotnych.

W tym ujęciu metaanaliza jest jedynie jedną z form analizy wtórnej, która zajmuje się integracją wyników zbioru analiz pierwotnych (dotyczy to procesu opracowywania wyników publikowanych analiz pierwotnych).

W każdym przypadku, bez względu na miejsce metaanalizy w obydwu klasyfikacjach, definicja metaanalizy Glassa pozostaje aktualna: „metaanaliza jest statystyczną analizą zbioru analitycznych wyników poszczególnych badań w celu integracji wniosków naukowych” (Glass 1976, s. 3). Dla metaanalizy charakterystyczne jest (Glass 1976, s. 3–4):

1. wykorzystywanie najbardziej obiektywnych metod selekcji badań wybranych do syntezy,
2. opisywanie cech wybranych badań z ilościowymi kategoriami (tzn. duża próba, mała próba; publikowanie w periodykach, monografie itp.),
3. wyniki wszystkich badań przekształcane są na jednakową skalę, zwaną „moc efektu” (ang.: effect size, ES),
4. używanie ilościowych metod statystycznych do pomiaru stosunku między cechami i wynikami badań.

Zaprezentowana charakterystyka dowodzi, że zdaniem Glassa nie każdy przegląd empiryczny jest metaanalizą. Dotyczy to głównie wcześniejszych przeglądów. Obok przedmiotu badań istotna jest również metodologia, czyli droga od pytań do konstatacji. Przedmiot badań zarówno w metaanalizie, jak i w przeglądzie może być taki sam, w przeciwieństwie do metodologii, która w obydwu przypadkach jest bardzo zróżnicowana.

3. Główne cechy metaanalizy

Wystąpienie Glassa spowodowało dwa zasadnicze skutki. Pierwszym z nich było ogromne zainteresowanie praktyką badań metaanalizy. Drugą korzyścią okazało się upowszechnienie potrzeby teoretycznego rozwoju i uzasadnienie metaanalizy jako metody badawczej.

Zamierzeniem autora jest dokonanie streszczenia głównych cech metaanalizy w oparciu o dotychczasowe źródła metodologiczne. Uwzględnione zostaną zasadnicze punkty wyjścia metaanalizy Glassa i prace powstałe w późniejszych latach.

Wśród nowszych opracowań w prezentacji problemu istotne okażą się prace Slavina, Hedgesa, Rosenthala, Kulika i Kulik, McGawa, Huntera i Schmidta.

Główne cechy metaanalizy można sformułować następująco:

1. metaanaliza obejmuje wyniki badań (rezultaty z raportów badań). Oznacza to, że w badaniu metaanalizy nie zbiera się początkowego materiału empirycznego. Do metaanalizy potrzebne są badania pierwotne i niektóre ich rezultaty,
2. metaanaliza jest aplikacją metod statystycznych na ilościowe rezultaty badań. Empirycznym materiałem metaanalizy są np. średnie (arytmetyczna, mediana, dominanta), miary rozrzutu (wariancja i odchylenie standardowe), wskaźniki korelacji (Pearsona, Spearmana, itd.), rezultaty weryfikacji hipotez statystycznych (test Kołmogorowa i Smirnowa, test $-\chi^2$ itd.), prze-

działy ufności i inne. Rezultaty początkowe są przedmiotem jedynie pierwotnej analizy,

3. metaanaliza obejmuje zbiór poszczególnych badań. Niektóre z metaanaliz integrują kilkaset, a nawet i kilka tysięcy badań,
4. w metaanalizie wylicza się moc efektu, a nie tylko jego kierunek lub ustalony poziom istotności. Procedura wyliczenia mocy efektu powinna umożliwić porównywanie efektów z poszczególnych badań,
5. przedmiotem metaanalizy jest również stosunek wyników badań do zasadniczych cech badań. Uwzględnia ona również wpływ cech badań na ich wyniki.

Należy zaznaczyć, że taka forma metaanalizy (odpowiadająca wymienionym wymogom) nie jest ogólnie akceptowana. Niektórzy autorzy używają terminu „metaanaliza” w ujęciu szerszym niż Glass.

Pierwszym dylematem zakresu metaanalizy jest ilość badań, które powinna ona objąć swym zasięgiem. Dowodzenie Glassa, wskazujące na konieczność występowania zbioru, nie jest wystarczająco precyzyjne i jednoznaczne.

Rodzi się proste pytanie: czy dla metaanalizy wystarczające jest użycie tylko dwóch badań? Rosenthal (1990) używa określenia „metaanaliza” również w odniesieniu do badań, w których badacz dokonuje porównania i zestawiania poziomu istotności z dwóch własnych eksperymentów. Jednakże przyjmuje się w większości, że warunkiem właściwej metaanalizy jest konieczność użycia zbioru badań, a nie tylko dwóch, trzech lub kilku. Dlatego też analiza, która integruje wyniki dwóch lub kilku badań, nie może pretendować do określenia „metaanaliza”. Oczywistym jest, że pytanie postawione wcześniej pozostaje teoretycznie otwarte, natomiast w praktyce przy obecnym stanie badań kwestia ta posiada już swoje ustalenia.

Drugi problem dotyczy przebiegu granicy między przeglądem i metaanalizą. Odnosi się to głównie do dwóch ostatnich cech metaanalizy. Dopuszcza się możliwość różnorodnych interpretacji. Pierwsza z nich zakłada, że wyliczanie mocy efektu i analiza statystyczna wpływu cech badań na ich wyniki są konieczne i bezwarunkowe. Inną interpretacją jest przypuszczenie, że omawiane czynności są jedynie kwestią metodologicznego poziomu metaanalizy. Wcześniejsze metaanalizy nie obejmowały wyliczania mocy efektu i analiz statystycznych wpływu cech badań na ich wyniki. Dopiero w ostatnim czasie staje się to na ogół warunkiem koniecznym. Wyliczanie mocy efektu wpływa na trafność i rzetelność wniosków metaanalitycznego badania – potęgując je. Podobne znaczenie ma również uwzględnianie wpływu cech poszczególnych badań (reprezentatywność prób, wielkość prób, model metodologiczny, forma publikowania itd.) na ich wyniki. Poziom istotności drugiej cechy metaanalizy nie jest jednak tak wielki, jak pierwszej. Tezy tej dowodził Glass, który empirycznie ustalił, że wpływy cech w większości są niewielkie (Glass, McGaw i Smith 1981).

Określenie „metaanaliza” było i jest przedmiotem dyskusji oraz krytyk. Niektórzy autorzy zarzucają temu sformułowaniu przywołanie skojarzenia z wyższym gatunkiem analizy w porównaniu z pierwotną i wtórną analizą, czyli z nadanalizą w sensie poziomu wartości naukowej wyników. Drugi problem tego określenia do-

tyczy jego brzmienia. Wskazuje ono na rozproszenie i rozdzielenie, choć w swojej istocie metaanaliza jest procedurą odwrotną – integrującą. Dlatego też niektórzy autorzy dowodzą, że do tej procedury bardziej odpowiednia jest nazwa synteza niż analiza. Żadne jednak z proponowanych określić, np. badawcza integracja, syntetyczne badanie, metasynthese, nie przyjęło się w praktyce.

4. Kontrowersje wokół metaanalizy

Proponowana metodologia wywołała falę krytyki. W krótkim czasie po pierwszej metaanalizie dotyczącej skuteczności psychoterapii pojawiły się zarzuty podważające założenia metodologiczne (Eysenck 1978, Mansfield i Busse 1977, Presby 1978). Nowe zarzuty pojawiły się również, gdy Glass i Smith opublikowali rezultaty metaanalizy, dotyczące wielkości klas szkolnych (Slavin 1984, Educational Research Service 1980).

Omawiana krytyka zawarta została w trzech zasadniczych punktach:

1. dokonane metaanalizy wiążą z sobą sprzeczne zjawiska (w konkretnych przypadkach sprzeczne badania); nie istnieje możliwość otrzymania logicznych i naukowo uzasadnionych wniosków przy pomocy integracji badań, które zawierają rozmaite czynności pomiaru, różnorodne zmienne i zupełnie różne badane populacje,
2. przeprowadzone metaanalizy zbyt wiele uwagi poświęcają badaniom o niskim poziomie jakości,
3. metaanalizy zbyt mocno opierają się na publikowanych rezultatach (rezultaty publikowane w zasadniczy sposób różnią się od niepublikowanych).

Wymieniona krytyka wymaga głębszej analizy. Pierwszy z zarzutów jest uzasadniony z punktu widzenia trudności, które pojawiają się w integracji bardzo odmiennych rezultatów. Istnieje wiele możliwości wyboru dróg rozwiązań metodologicznych, dotyczących badania określonej tematyki. Stąd też pojawia się wiele różnorodnych podejść w poszukiwaniu rozwiązań na następujące pytania: jak badać? jak przebyć drogę od postawienia hipotez do ich weryfikacji? jak otrzymać empiryczne dane o zjawiskach wychowawczych? itd. W pierwszej kolejności należy podkreślić, że we wcześniejszych przeglądach kwantytatywnych badacze również korzystali z bardzo różnorodnego materiału empirycznego. Nie używali oni jednak skomplikowanych metod statystycznych. Dlatego też problem sprzeczności rezultatów nie był tak bardzo widoczny. W przeglądach i w metaanalizie nie można jednak ograniczyć się tylko do dociekań, które stanowią wielokrotną replikę określonego badania. Drugim argumentem przeciw omawianej krytyce jest rozwój metaanalizy w czasie pojawienia formułowanych zarzutów (Mansfield i Busse 1977, Presby 1978). Okazało się, że istnieją drogi rozwiązań problemów i dylematów integracji bardzo różnorodnych rezultatów. Przejrzystym dowodem takich możliwości jest procedura obliczania miary mocy efektu. Jednak pomimo ogromnego postępu w rozwoju metaanalizy ciągle otwarte pozostają pytania i wątpliwości, powiązane z precyzyjnymi metodami integracji rozmaitych rezultatów.

Druga krytyka, która dotyczy różnorodnego poziomu badań wybranych do metaanalizy, jest raczej krytyką niektórych metaanaliz Glassa niż ogólną krytyką nowej metodologii. Należy stwierdzić, że istnieją możliwości uwzględniania poziomu metodologicznego („jakości”) poszczególnych badań. Jedną z możliwych dróg rozwiązania tego problemu jest metoda Slavina. Jednak i ona okazuje się wadliwa, gdy staje się dogmatem.

Krytyka właściwości metaanalizy polegającej na zbyt mocnym opieraniu się na publikowanych rezultatach jest uzasadniona. Nie istnieje możliwość otrzymania danych o wszystkich niepublikowanych badaniach. Nigdy też nie staną się one tak dostępne, jak badania publikowane. Dowiedziono empirycznie, że sam fakt publikowania ma wpływ na rezultaty. Stwierdzenia krytyków wskazujące na fakt, że istnieją różnice między rezultatami niepublikowanych i publikowanych badań są więc prawdziwe. Rodzi się zatem pytanie: czy istnieje konieczność rezygnacji z integracji rezultatów dostępnych z powodu braku uwzględnienia rezultatów niedostępnych, nawet w przypadku, gdy rezultaty te są niezwykle istotne? Krytykę tę należy rozumieć dwutorowo. W pierwszym sensie, że metaanalityk powinien dołożyć wszelkich starań, aby otrzymać także badania niepublikowane i ich rezultaty. Drugi sens wskazuje na potrzebę odnalezienia możliwości uwzględniania wpływu publikowania na rezultaty już w trakcie dokonywanych czynności statystycznych (a nie tylko w ostatecznych interpretacjach). Metaanaliza powinna rozwinąć procedurę precyzyjnego uwzględniania omawianego wpływu. Nie wystarczającym jest interpretowanie ostatecznych wyników jedynie w świetle faktu, iż wszystkie rezultaty nie odzwierciedlają rzeczywistości wychowawczej jednakowo dobrze i rzetelnie.

5. Podsumowanie

Rozwój badań empirycznych jest jedną z najbardziej charakterystycznych cech rozwoju pedagogiki ostatniego stulecia. Szczególnie mocno zaznaczył się on w minionych dziesięcioleciach wraz z wystąpieniem problemu masowości rezultatów badawczych. Konsekwencją tego kierunku rozwoju stała się potrzeba integracji rezultatów badań empirycznych. Przeglądy kwantytatywne, istniejąc w literaturze od dawna, próbują streszczać najważniejsze osiągnięcia naukowe i wyniki badań o określonej problematyce. Posiadają one jednak pewne słabości. Do najistotniejszych z nich należy przypadkowy wybór prac do integracji, ich niska rzetelność i mała systematyczność oraz brak istnienia możliwości komparacji z powodu różnorodności modelu metodologicznego.

Istotę integracji badań empirycznych w zasadniczy sposób zmieniła metaanaliza. Możliwość kwantytatywnej integracji umożliwiła rozszerzenie granic badań empirycznych rozmaitych zjawisk wychowawczych, na skalę nie spotykaną wcześniej.

Nowa metodologia stała się jednak także obiektem krytyki. W krótkim czasie po opublikowaniu pierwszych metaanaliz, pojawiły się zarzuty podważające założe-

nia metodologiczne. Zaprezentowana w omówionym rozdziale krytyka i towarzyszące jej wątpliwości wskazują nie tylko na możliwość, ale wręcz na konieczność dalszego rozwoju metodologii integracji kwantytatywnej.

Bibliografia

- Cooper H.M., *Integrating research: A guide to literature reviews*, Newbury Park 1989.
- Cooper H.M. i Rosenthal R., *Statistical versus traditional procedures for summarizing research findings*, *Psychological Bulletin*, 1980, s. 442–449.
- Glass G. V., *Primary, secondary, and meta-analysis of research*, *Educational Researcher*, 1976, s. 3–8.
- Glass G.V., *Integrating findings: The meta-analysis of research*, *Review of Research in Education*, 1977, s. 351–379.
- Glass G.V. i Smith M.L., *Meta-analysis of research on class size and achievement*, *Educational Evaluation and Policy Analysis*, 1979, s. 2–16.
- Glass G.V., McGaw B. i Smith M.L., *Meta-analysis in social research*, Beverly Hills 1981.
- Hedges L.V. i Olkin I., *Analyses, reanalyses, and meta-analysis*, *Contemporary Education Review*, 1982, s. 157–165.
- Kulik J.A. i Kulik C.-L.C., *Meta-Analysis in Education*, *International Review of Educational Research*, 1989, s. 223–340.
- Rosenthal R., *Meta-analytic procedures for social research*, Beverly Hills 1991.
- Slavin R.E., *Best-evidence synthesis: An alternative to meta-analysis and traditional reviews*, *Educational Researcher*, 1986, s. 5–11.
- Smith M.L. i Glass G.V., *Meta-analysis of research on class size and its relationship to attitudes and instruction*, *American Educational Research Journal*, 1980, s. 419–433.
- Wolf F.M., *Meta-analysis: Quantitative methods for research synthesis*, Beverly Hills 1986.