

Część III

EKSPERYMENT W BADANIACH GLOTTODYDAKTYCZNYCH

Katarzyna Rokoszevska
Akademia im. Jana Długosza w Częstochowie

PLANOWANIE I PRZE- PROWADZANIE EKSPER- RYMENTU NAUKOWEGO W BADANIACH GŁOTTO- DYDAKTYCZNYCH

Planning and conducting experiments in glottodidactic research

The classic experiment is one of the most important research methods in glottodidactics. The classic experiment must fulfill two conditions, namely control and experimental groups must be formed and subjects must be assigned to these groups at random. If one of these conditions, namely randomization, is not met, the research method is called quasi-experimental. If none of the conditions is met the method is called pre-experimental. In fact, the pre-experiment is probably most often used in research on teaching and learning foreign languages because, just like the true experiment, it allows the researcher to draw conclusions about the cause and effect relationship between particular variables and, at the same time, it is more practical than the true experiment. The aim of this paper is to present the experimental procedure consisting of ten steps proposed by Clegg (1994).

1. Badania ilościowe, a badania jakościowe

Eksperyment naukowy należy do badań ilościowych, które różnią się od badań jakościowych pod wieloma względami. Jak wskazują Larsen-Freeman i Long (1991), celem badań ilościowych jest zrozumienie procesu przyswajania drugiego języka i zachowań jego uczestników, podczas gdy celem badań ilościowych jest przewidywanie i wyjaśnianie ich zachowań oraz ustalanie związków

przyczynowo-skutkowych pomiędzy licznymi czynnikami wpływającymi na ten proces. W badaniach jakościowych, które cechuje bliskość danych, badacz stosuje naturalistyczną, niekontrolowaną obserwację z perspektywy wewnętrznej, co może prowadzić do subiektywizmu. W badaniach ilościowych, badacz, oddalony od danych, przeprowadza kontrolowane pomiary z perspektywy zewnętrznej, co zwiększa szansę na obiektywizm. Badania jakościowe można określić jako induktywne, opisowe, odkrywczе, holistyczne i ukierunkowane na proces, podczas gdy badania ilościowe jako deduktywne, redukcjonistyczne, ukierunkowane na weryfikację hipotez i produkt. O ile cechą badań jakościowych jest trafność, to badań ilościowych rzetelność. Ważną różnicą jest także to, że badania jakościowe nie dają podstaw do uogólnienia wniosków otrzymanych w badaniu z próbki dla populacji macierzystej, z której ta próbka została pobrana. Warto też dodać, że badania jakościowe zakładają dynamiczny, a ilościowe stały charakter rzeczywistości. Należy także zaznaczyć, że granica pomiędzy tymi dwoma typami badań jest płynna. Innymi słowy, badania jakościowe i ilościowe tworzą kontinuum rozciągające się od introspekcji, obserwacji uczestniczącej i nieuczestniczącej poprzez badania korelacyjne do schematu przed-eksperymentalnego, pseudo-eksperymentu i eksperymentu klasycznego.

W klasycznym eksperymencie należy spełnić dwa warunki. Należy utworzyć grupę kontrolną i eksperymentalną, a obiekty należy przypisać do tych grup w sposób przypadkowy, co nazywamy randomizacją. Według Hatch i Farhady (1982), zaletą eksperymentu klasycznego jest możliwość ustalenia związku przyczynowo-skutkowego oraz uogólnienia wniosków dla populacji macierzystej. Jednakże, eksperyment klasyczny jest czasem nieodpowiednią metodą do badania zachowań ludzkich, ponieważ badacz często w nie-naturalny sposób manipuluje zmiennymi w warunkach laboratoryjnych, co powoduje, że wnioski z takiego badania trudno odnieść do prawdziwej klasy uczniów. Ponadto, drugi warunek klasycznego eksperymentu jest trudny do spełnienia w badaniach glottodydaktycznych, które są często prowadzone we wcześniej utworzonych klasach uczniów. Dlatego też w tej dziedzinie często stosuje się pseudo-eksperyment, który różni się od eksperymentu klasycznego tym, że obiekty nie są losowo przypisane do grupy kontrolnej i eksperymentalnej. Pseudo-eksperyment umożliwia badanie zachowań uczniów w naturalnym kontekście. Pod względem rzetelności, pseudo-eksperyment nie jest tak adekwatny jak prawdziwy eksperyment, ale jest lepszy od schematu przed-eksperymentalnego, w którym żaden z warunków eksperymentu klasycznego nie jest spełniony. Schemat ten ukazuje pewien obraz procesu przyswajania drugiego języka, pozwala na generowanie hipotez, ale nie daje podstaw do wyciągania związków przyczynowo-skutkowych. Przykładem takiego schematu jest badanie „przed i po”, w którym uczniowie są badani pod względem danej zmiennej, np. znajomości słownictwa, przed i po wprowadzeniu działania eksperymentalnego, np. treningu ze strategii uczenia się słownictwa.

2. Plan postępowania eksperymentalnego

2.1. Określenie zmiennych i postawienie hipotezy

Plan postępowania eksperymentalnego składa się według Clegg'a (1994) z dziesięciu kroków, z których pierwszy to pomysł o wpływie jednej zmiennej na drugą. Pomysł ten może powstać na podstawie wiedzy, doświadczenia i intuicji badacza, albo też na podstawie informacji, jakiej dostarczają testy ustalające korelację pomiędzy danymi zmiennymi. Można powiedzieć, że to ten pierwszy oczywisty krok decyduje o odkrywczości naszego badania.

Drugi krok to zdefiniowanie zmiennych i zdecydowanie, w jakich jednostkach będą one mierzone. Zmienna to ilościowa reprezentacja konstruktów, czyli rzeczywistej cechy lub umiejętności człowieka (Brown 1991). Zmienne należy zoperacjonalizować, czyli zdefiniować w stosunku do danego konstruktów tak, aby był on możliwy do zaobserwowania, zmierzenia i przetestowania w obiektywny sposób. Przykładem konstruktów może być biegłość ucznia w języku obcym, a zmienną wyniki z odpowiedniego testu językowego. Wyróżnia się różne typy zmiennych (Brown 1991). Zmienna zależna (ZZ) to zmienna odnośnie, której stawiana jest hipoteza, że będzie się ona zmieniać wskutek działania eksperymentalnego. Zmienna niezależna (ZN) to wybrana zmienna systematycznie manipulowana w celu zbadania jej wpływu na zmienną zależną, np. inteligencja, motywacja, osobowość. Zmienna interwencyjna to abstrakcyjna, teoretyczna nazwa, która określa związek pomiędzy zmienną zależną i zmienną niezależną, np. rola różnic indywidualnych w procesie przyswajania drugiego języka. Zmienna zakłócająca to zmienna, która wpływa na związek między zmienną zależną i zmienną niezależną, np. wiek uczniów, osobowość eksperymentatora. Istnieją także tzw. zmienne „nieprzyjemne”, które są nieszkodliwe, ponieważ wpływają jednakowo na obie grupy, np. niepokój, posiłki, zmęczenie (Clegg 1994). W końcu, zmienne kontrolne to zmienne, które są zneutralizowane po to, aby nie wpływały na wynik, np. dłuższy pobyt w kraju języka docelowego (Brown 1991). W praktyce oznacza to, że uczniowie, którzy nie spełniają warunków wyrażonych w zmiennych kontrolnych są wykluczani z badania bądź też biorą w nim udział wraz z resztą klasy, ale ich wyniki nie są brane pod uwagę.

Aby zmienne mogły być określone w sposób ilościowy, muszą być wyrażone w odpowiednich jednostkach. Rodzaje jednostek, stosowanych do oceniania wielkości zmiennych, mają istotne znaczenie w teście statystycznym wybranym do analizy danych. Istnieją cztery skale pomiarowe w rosnącym porządku dokładności (Clegg 1994). W skali nominalnej lub symbolicznej, liczby nominalne oznaczają nazwy kategorii. Są one używane do klasyfikowania rzeczy i mogą być zastąpione przez inne symbole, np. litery, kształty, kolory czy opis słowny. Własnością skali nominalnej jest równoważność, co oznacza, że liczba przyporządkowana danym obiektom oznacza, że są one do siebie podobne czy

też inaczej mówiąc, że można pomiędzy nimi postawić znak równości. W skali tej największa liczba pokazuje liczbę kategorii w danym układzie, a działania arytmetyczne nie są dozwolone. Skala porządkowa, oparta na liczbach porządkowych, pokazuje proste uporządkowanie danej zmiennej i związek pomiędzy poszczególnymi pozycjami. Własnością tej skali jest równoważność i względny rozmiar wskazywany przez symbole „większy niż” i „mniejszy niż”. Różnica pomiędzy dwoma stopniami w różnych miejscach skali nie jest taka sama, co oznacza, że działania arytmetyczne są niedopuszczalne. W skali przedziałowej liczby są uporządkowane w taki sposób, że odpowiednie przedziały we wszystkich miejscach skali mają jednakowe rozmiary. Operacje arytmetyczne są tu dozwolone. Jednakże, porównania między skalami są trudne, ponieważ różne skale mają różne przedziały i punkty zerowe, które są tu dowolnie ustalone. Skala ta może zawierać liczby ujemne i dodatnie. Skala przedziałowa jest stosowana w większości testów językowych. Skala stosunkowa lub ilorazowa zawiera uszeregowanie liczb z równymi przedziałami, ale posiada zero absolutne, dzięki czemu działania arytmetyczne są dozwolone, a porównania między skalami proste. Skala ta nie może zawierać liczb ujemnych. Ze względu na zero absolutne, skala ta nie jest stosowana w badaniach glottodydaktycznych, ponieważ nawet uczeń, który nie zna języka obcego wnosi do jego nauki istotne informacje z zakresu języka pierwszego (Brown 1991).

Trzeci krok w postępowaniu eksperymentalnym to przedstawienie pomysłu w sposób formalny, za pomocą jednostronnej lub dwustronnej hipotezy eksperymentalnej (Clegg 1994). Hipoteza to każda idea lub teoria, która podaje pewne prowizoryczne przewidywania. Planowanie eksperymentu należy zacząć od postawienia hipotezy zerowej, która jest zdefiniowana jako „*hipoteza statystyczna o braku różnicy*” (Clegg 1994:72). Zakłada ona, że zmienna niezależna nie wpływa na zmienną zależną w oczekiwany sposób. Inaczej mówiąc, eksperyment należy zacząć od założenia, że się on nie uda, czyli że dwa zbiory wyników nie będą się różnić. Następnie należy sformułować hipotezę badawczą lub eksperymentalną (H_1), która jest hipotezą alternatywną (H_A) dla hipotezy zerowej. Hipoteza badawcza to wstępna idea badacza w pracy eksperymentalnej. Może ona być jednostronna lub dwustronna. Hipoteza jednostronna lub kierunkowa wskazuje przewidywany kierunek wyniku, co jest widoczne w użyciu takich słów jak „zmniejszyć się”, „zwiększyć się”, „obniżony”, „wzrosnąć” itp. Hipoteza dwustronna lub niekierunkowa jest bardziej niesprecyzowana, a słowa charakterystyczne w jej sformułowaniu to „wpływ”, „stosunek”, „zmiana” itp.

Typ hipotezy wpływa na sposób wnioskowania oraz na interpretację wyników (Clegg 1994). Jeśli badania pokazały istotną różnicę między dwoma zbiorami wyników, a przewidywana zmiana zaszła w oczekiwany sposób, to należy przyjąć hipotezę eksperymentalną, a jeśli badania pokazały brak różnicy to należy ją odrzucić. Jednakże, w przypadku, gdy badania wykazały różnicę pomiędzy zbiorami liczb, ale zmiana zaszła w odwrotnym kierunku, hipotezę

kierunkową należy odrzucić i wyciągnąć wniosek, że eksperyment się nie udał. Można go przeprowadzić ponownie (replikacja) przewidując nowy kierunek lub nie. W tej samej sytuacji, w przypadku hipotezy niekierunkowej, można odrzucić hipotezę zerową i przyjąć hipotezę eksperymentalną mówiącą o tym, że zmienna niezależna mogła zmienić zmienną zależną. Ponadto, typ hipotezy wpływa na interpretację wskaźnika prawdopodobieństwa w tabelach statystycznych dla danego testu. Wartość poziomu istotności w przypadku hipotezy dwustronnej jest dwa razy wyższa niż dla hipotezy jednostronnej. Oznacza to, że ze statystycznego punktu widzenia, łatwiej jest wykazać istotną różnicę pomiędzy zbiorami wyników w przypadku hipotezy jednostronnej. Może się też zdarzyć tak, że pomiędzy dwoma zbiorami wyników nie będzie istotnej różnicy, jeśli zostaną one zinterpretowane jako rezultaty dwustronne, ale będzie istotna różnica, jeśli zostaną one zinterpretowane jako rezultaty jednostronne. Dlatego też, każdy badacz powinien być świadomy tych różnic i wziąć je pod uwagę przy formułowaniu hipotezy, ponieważ jej zmiana w trakcie analizy statystycznej po przeprowadzonym badaniu jest niedopuszczalna.

2.2. Wybór analizy statystycznej

Krok czwarty w postępowaniu eksperymentalnym to zdecydowanie, jaki rodzaj analizy statystycznej będzie odpowiedni. Zasadniczo wyróżniane są statystyki opisowe i inferencyjne (Hatch i Lazarton 1991). Statystyki opisowe zawierają miary tendencji centralnej, rozkłady częstości oraz miary rozrzutu. Miary tendencji centralnej to średnia arytmetyczna, mediana i moda. Średnia arytmetyczna, czyli suma wyników podzielona przez ich ilość, jest reprezentatywna wtedy, gdy wyniki znajdują się w pobliżu wartości centralnej. Jeśli wyniki są rozrzucone bardzo szeroko, to może ona wprowadzać w błąd. Jej wadą jest też to, że może nie być równa żadnej z liczb w danym zbiorze. Ze średnią arytmetyczną wiąże się zjawisko zbędnej dokładności, które polega na użyciu zbyt wielu liczb po przecinku, podczas gdy liczby te powinny reprezentować rzeczywistą dokładność (Clegg 1994). Mediana, czyli środkowa liczba w szeregu liczb ustawionych według wielkości, jest łatwa do odnalezienia jedynie w małym zbiorze liczb. Jest stosowana, jeśli pojawiają się tzw. „samotniki”, czyli pojedyncze wyniki odbiegające od reszty (Clegg 1994). Należy pamiętać o tym, że zmiana jednej liczby bliżej środka powoduje znaczną zmianę mediany, a nie średniej arytmetycznej. Jednocześnie, zmiana liczby krańcowej nie powoduje zmiany mediany, tylko średniej arytmetycznej. Moda, czyli najczęściej pojawiająca się wartość w danym zbiorze, jest stosowana, jeśli liczby w rozkładzie są równomiernie rozmieszczone. Jednakże, jest ona rzadko stosowana, ponieważ jest wielkością niestabilną w takim sensie, że zmiana jednej liczby może ją bardzo zmienić. Ponadto, duże zbiory liczb mają wiele modalnych.

Diagramy czy też rozkłady częstości pokazują, jak często dana wartość występuje (Brown 1991). Zasadniczo wyróżnia się trzy rodzaje rozkładów.

Rozkład normalny, zwany krzywą Gaussa, jest symetryczny i ma kształt dzwonu. W rozkładzie tym większość wyników znajduje się w środkowej części, a wyników ekstremalnych jest niewiele. Średnia arytmetyczna, mediana i moda mają tu taką samą lub prawie taką samą wartość, a końce krzywej nie dotykają osi poziomej, tylko zbliżają się do niej w nieskończoności. W rozkładzie dodatnio skośnym, liczby są rozłożone dość równomiernie poniżej średniej arytmetycznej, natomiast w rozkładzie ujemnie skośnym powyżej średniej arytmetycznej. Można tu posługiwać się wynikami surowymi, procentowymi, ważonymi, przy których różne wagi są przypisane do różnych części testu, oraz standardowymi. Wyniki standardowe umożliwiają porównanie wyników z różnych testów. Są one stosowane tylko do liczb obliczonych na podstawie rozkładów normalnych. Wyniki $-z$ pokazują odchylenie wyniku od średniej arytmetycznej wyrażone w jednostkach odchylenia standardowego. Ponieważ mogą one być liczbami ujemnymi lub ułamkami dziesiętnymi, można je przeliczyć na wyniki $-T$ za pomocą prostego wzoru ($T = 10z + 50$) po to, aby stały się bardziej czytelne dla ucznia.

Miary rozrzutu lub dyspersji pokazują jak bardzo rozsypane są liczby (Brown 1991). Rozstęp pokazuje przez ile liczb rozciągnięty jest rozkład. Jest on obliczany przez odjęcie najmniejszego wyniku od największego. Jest on stosowny wtedy, gdy wszystkie wartości są podobne. Wadą rozstępu jest to, że wartości ekstremalne mają na niego bardzo duży wpływ, a obecność samotników może powodować to, że podobne rozkłady mają różne rozstępy. Odchylenie średnie wskazuje to, jak wyniki w rozkładzie różnią się od średniej arytmetycznej. Można je zdefiniować jako „średnią arytmetyczną odchyień od średniej arytmetycznej” (Clegg 1994:34). Im większe odchylenie średnie, tym bardziej rozproszony rozkład wyników. Odchylenie średnie, które bazuje na wszystkich liczbach w rozkładzie, jest bardziej stabilną statystyką niż rozstęp, który bazuje tylko na dwóch liczbach. Odchylenie standardowe (OS) podsumowuje średnie odległości wszystkich wyników od średniej arytmetycznej danego zbioru. Z kolei wariancja to kwadrat odchylenia standardowego.

Najczęściej stosowane statystyki opisowe to średnia arytmetyczna i odchylenie standardowe. Dane te krótko podsumowują wyniki otrzymane w badaniu, ale nie odpowiadają na pytanie czy hipoteza eksperymentalna jest prawdziwa. Aby to ustalić należy zastosować statystyki inferencyjne, czyli testy wnioskowania statystycznego, które podadzą prawdopodobieństwo tego, czy różnica pomiędzy zbiorami wyników jest rzeczywista czy przypadkowa, informując tym samym o marginesie błędu.

Testy statystyczne to ciągi prostych obliczeń arytmetycznych, które należy wykonać w odpowiedniej kolejności. Testy statystyczne dzielą się na parametryczne i nieparametryczne. Testy różnią się od siebie tzw. mocą testu, czyli zdolnością do wykrywania istotnych różnic w zbiorach wyników (Clegg 1994). Moc testu zależy od jego precyzji. Im bardziej wyszukany test, im więcej informacji wykorzystuje, tym jest on mocniejszy. Testy parametryczne są testami

mocniejszymi od testów nieparametrycznych, ale dla ich zastosowania należy spełnić trzy warunki (Hatch i Lazarton 1991). Pierwszy warunek jest taki, że obie próbki muszą mieć rozkład normalny. Aby ustalić czy stopień odchylenia od normalnego kształtu jest do zaakceptowania, można zastosować tzw. test „rzutu oka” lub testy chi-kwadrat. Zastosowanie testu „rzutu oka” polega na zaobserwowaniu czy te same względne proporcje wyników znajdują się pomiędzy konkretnymi wartościami liczb. Testy chi-kwadrat to testy dobrego dopasowania lub zgodności. Porównują one kształt rozkładu nie na podstawie rzeczywistych wyników, tylko ich częstości. Działają one na obszarze, na którym kształty rozkładów nakładają się. Skala pomiarowa nie ma tu znaczenia, a rozmiar próbek może być różny. Należy pamiętać o tym, że testy te nie wykazują związku przyczynowo-skutkowego. Drugi warunek dla zastosowania testów parametrycznych jest taki, że wariancje próbek powinny być podobne. Jednorodność wariancji oznacza, że dwa zbiory wyników są rozproszone mniej więcej w równej ilości, co można sprawdzić za pomocą testu ilorazowego wariancji, czyli testu *F*. Trzeci warunek dotyczy skali. Mówi on o tym, że próbki powinny obejmować wyniki o skali co najmniej przedziałowej.

Piąty krok postępowania eksperymentalnego to określenie poziomu istotności czy też prawdopodobieństwa próbki (Clegg 1994). Istotność jest wyrażona w formie dokładnej liczbowej wartości prawdopodobieństwa. Wyróżniane są trzy opisy prawdopodobieństwa: ułamkowy lub stosunkowy (1 na 20), procentowy (50% na 50%) lub w postaci ułamka dziesiętnego (5% – 0,05). Jednocześnie wyróżniane są trzy poziomy istotności. Najniższy akceptowany poziom ufności to $p = 0,05$, $p < 0,05$ (5%). Oznacza on to, że jeżeli działanie na dane następuje z pewnością 0,05 (5%), to możliwość szczęśliwego trafu wynosi 1 na 20 lub 5 na 100 przypadków. Inaczej mówiąc, w 20% przypadków wyniki są przypadkowe, a w 80% przypadków wyniki odzwierciedlają rzeczywistą różnicę spowodowaną działaniem zmiennej niezależnej na zmienną zależną. Pozostałe dwa poziomy istotności to $p < 0,01$ (1%; 1 na 100 przypadków) i $p < 0,001$ (0,1%, 1 na 1000 przypadków). Należy zauważyć, że im mniejsza liczba wyraża prawdopodobieństwo, tym wyższy jest poziom istotności. Ogólnie mówiąc, poziom istotności powinien być ustalony przed przeprowadzeniem eksperymentu, a jego wysokość zależy od tego, na ile bezbłędny powinien być wniosek badacza. W badaniach glottodydaktycznych najczęściej stosuje się pierwszy z wymienionych tu poziomów.

2.3. Wybór próbki badawczej

Szósty krok postępowania eksperymentalnego to wybranie próbki z populacji macierzystej będącej przedmiotem badań (Clegg 1994). Populacja to zbiór ludzi, zdarzeń lub wyników mających pewną wspólną cechę, z którego pobierana jest próbka do badań. Populacja macierzysta to populacja, z której została pobrana próbka. Populacja generalna to typowi członkowie społeczeństwa lub co

najmniej 80% badanej populacji. Wyniki otrzymane od osób biorących udział w eksperymencie uważane są za próbki z populacji o praktycznie nieskończonym rozmiarze. Jednakże, prowadzenie eksperymentów na całej populacji jest bardzo rzadkie. Celem pracy eksperymentalnej jest całościowy opis populacji i nieograniczenie wniosków do próbki, która została przebadana.

Pobieranie próbek oznacza otrzymywanie wyników lub obserwacji od ludzi (próbki zdarzeń) w pewnych warunkach (próbki zmiennych środowiskowych) w konkretnym czasie (próbki czasowe) (Clegg 1994). Wszystkie te elementy zmieniają się z dużymi konsekwencjami, co oznacza, że eksperyment w naukach społecznych jest trudniejszy niż w naukach ścisłych. Na pobieranie próbek mają wpływ takie czynniki jak rozmiar, procedura i trafność doboru. Im większa jest próbka tym lepsza, ponieważ lepiej odzwierciedla ona charakterystyczne cechy populacji macierzystej. Procedura pobierania próbek powinna być obiektywna, czyli publicznie sprawdzalna, a dobór elementów tworzących próbkę powinien być trafny. Próbki powinny być pobierane w sposób losowy, tzn. na podstawie starannego planu, który gwarantuje to, że wszystkie elementy populacji macierzystej mają taką samą szansę na pojawienie się w próbce. W przeciwnym razie, próbka jest uważana za obciążoną.

Istnieją trzy sposoby pobierania próbek (Hatch i Lazartou 1991). Systematyczne pobieranie próbek polega na tym, że każdemu obiektowi przydziela się numer, np. według alfabetycznej listy nazwisk, a następnie dobiera się podgrupę do badania za pomocą tablic liczb losowych, które zawierają ciągi cyfr od 1 do 9 wymieszanych w taki sposób, że każda z tych cyfr pojawia się tak samo często. W przypadku większej grupy, cyfry te można odczytywać parami. Warstwowe pobieranie próbek, gdzie warstwy to różne wartości zmiennych, takich jak wiek, płeć, klasa społeczna, dochody czy rasa, polega na tym, że próbka składa się z podgrup, które zawierają istotne zmienne w takich proporcjach jak w populacji macierzystej. Tworzenie próbek w taki sposób wymaga dobrej znajomości danej populacji. Jednocześnie, szanse odchylenia próbki są tu mniejsze. Pobieranie próbek wiązkami opiera się na istnieniu naturalnych grup, np. dzieci w klasie, które to grupy są numerowane i wybiera się z nich losową próbkę. Następnie w każdej wiązce wyróżnia się podgrupy i z jednej lub więcej z nich pobiera się elementy do próbki. Należy zdawać sobie sprawę z tego, że badacz jest częścią próbki i źródłem jej obciążenia. Obecnie uważa się, że czynniki takie jak płeć, rasa, cechy fizyczne czy osobowość badacza mogą wpływać na rodzaj wyników otrzymanych od obiektów. Dlatego też zaleca się przeprowadzanie eksperymentu przez jednego lub wielu przeszkolonych badaczy.

2.4. Zgromadzenie danych

Siódmy krok omawianej tu procedury to zastosowanie działania eksperymentalnego do jednej części próbki i potraktowanie drugiej jako grupy kontrolnej.

Działanie eksperymentalne jest stosowane w grupie eksperymentalnej. Grupa kontrolna jest traktowana w podobny sposób, ale przy użyciu innych wartości zmiennej niezależnej. W badaniu „przed i po” występuje jedna grupa. Jest to schemat przed-eksperymentalny, który, jak już wcześniej wspomniano, zobowiązuje badacza do innego wnioskowania niż w eksperymencie.

Krok ósmy to zgromadzenie danych w wyniku, którego badacz otrzymuje dwa zbiory wyników odzwierciedlające różne wartości zmiennej niezależnej. Wyniki te są wartościami zmiennej zależnej, np. wyniki uczniów w nauce słownictwa. Gromadzenie danych wiąże się z projektowaniem eksperymentów, które dotyczy metod dopasowania grup lub pojedynczych wyników.

Wyróżniane są trzy rodzaje projektów (Clegg 1994). Projekt powtarzanych pomiarów, zwany także projektem wewnątrz obiektów, jest oparty na obiektach samokontrolujących się, z których każdy dostarcza dwóch wyników. Obiekty powtarzają swoje działanie w nieznacznie zmienionych warunkach, a wyniki są porównywane wewnątrz każdego obiektu. W projekcie tym występuje skutek praktyki i efekt znużenia, które polegają na tym, że obiekty biorące udział w eksperymentach przez dłuższy czas, lepiej wykonują polecenia, ale mogą też być znużone. Skutek praktyki i efekt znużenia znoszą się wzajemnie i nie powodują błędu zakłócającego, który mógłby zaciemnić konsekwencje zmiennej niezależnej. Jest to jeden z najtrafniejszych prostych projektów, ale nie może być często używany. Nie jest on najlepszy dla eksperymentów z uczeniem się, ponieważ materiału testowego można się uczyć tylko raz nawet, jeśli przerwa pomiędzy sesjami jest długa. Pewnym rozwiązaniem jest przygotowanie dwóch bardzo podobnych zbiorów do nauki. Przykładem projektu powtarzanych pomiarów jest dosyć często stosowane w glottodydaktyce badanie „przed i po”.

W projekcie obiektów dopasowanych, obiekty są połączone w pary ze względu na istotne zmienne w taki sposób, że wynik w jednej grupie jest porównywany z konkretnym wynikiem w drugiej grupie. W celu osiągnięcia dokładnego łączenia w pary, najpierw tworzy się jedną grupę, a potem poszukuje się partnerów do każdego jej elementu. Najlepszym przykładem dopasowanych partnerów są bliźnięta jednojajowe, których identyczny kod genetyczny redukuje liczbę zmiennych. Obiekty takie posiadają prawie takie samo doświadczenie społeczne i podobną inteligencję, osobowość czy inne cechy psychiczne. Dopasowywanie niespokrewnionych osobników odbywa się na podstawie istotnych zmiennych.

W projekcie obiektów niezależnych, zwanym także projektem pomiędzy obiektami, obiekty są dzielone na dwie grupy, które poddaje się różnym działaniom. Podstawowe założenie tego projektu jest takie, że grupy powinny być porównywalne od początku pod względem wszystkich istotnych zmiennych. W projekcie tym to nie poszczególne obiekty, ale całe grupy są dopasowane całościowo pod względem istotnych zmiennych. Projekt ten z mniejszym prawdopodobieństwem wykazuje istotną różnicę w wynikach. Jest on mniej czuły na drobne zmiany pod wpływem działania eksperymentalnego, ponieważ

mogą wystąpić duże zmienności pomiędzy wszystkimi obiektami w dwóch grupach. Z punktu widzenia statystyki zaleca się dopasowywanie wyników. Może się zdarzyć tak, że różnica pomiędzy zbiorami liczb będzie istotna, jeśli będą one dopasowane, a nieistotna, jeśli będą potraktowane jako niedopasowane. Należy też zauważyć, że w badaniach glottodydaktycznych najczęściej stosuje się jednak projekty obiektów niezależnych.

2.5. Analiza danych i wyciągnięcie wniosków

Dziewiąty krok postępowania eksperymentalnego to przeanalizowanie danych, które składa się z trzech etapów, tj. postawienia hipotezy zerowej, zastosowania odpowiedniego testu statystycznego i przyjęcia lub odrzucenia na jego podstawie hipotezy zerowej (Clegg 1994). Wybór testu zależy od rodzaju próbki. W przypadku próbek związanych, można użyć dwóch testów nieparametrycznych, tj. testu znaków i testu Wilcozona, oraz jednego testu parametrycznego, tj. testu t Studenta dla próbek związanych, jeśli zostaną spełnione trzy, omówione wcześniej, warunki dla zastosowania testów parametrycznych (Hatch i Lazarton 1991, Clegg 1994). Test znaków, który dostarcza statystyki S można zastosować, gdy wyniki wyrażone są co najmniej w skali nominalnej. Test ten pokazuje ile jest wyników w jednym zbiorze, które są mniejsze od swoich odpowiedników w drugim zbiorze i ile jest tych, które są większe. Może być użyty do danych nieliczbowych, wyrażonych na przykład za pomocą takich słów jak „lepszy”, „taki sam”, „gorszy”, „zgadzam się”, „nie wiem”, „nie zgadzam się”. W teście tym liczy się różnice pomiędzy wynikami połączonymi w pary, przypisując każdej znak plus lub minus, a następnie liczy się ile jest różnic ze znakiem plus, a ile ze znakiem minus, przy czym mniejsza z tych liczb to statystyka S pokazująca ile razy występuje znak rzadziej pojawiający się. Dodatkowo, liczy się sumę par wyników, pomiędzy którymi jest różnica (N). Za pomocą tych dwóch danych odczytywany jest poziom istotności dla testu jedno- lub dwustronnego. W przypadku danych nieliczbowych, elementy należące do jednej kategorii mają plusy, do drugiej minusy, a te, które się nie zmieniają są wyłączone z analizy.

Test Wilcozona, który dostarcza statystyki T , może być użyty dla danych wyrażonych, co najmniej w skali porządkowej. W teście tym obliczamy różnice pomiędzy parami liczb pamiętając o znaku plus lub minus, rangujemy te różnice pod względem wielkości, przy czym najmniejsza wartość ma rangę 1. Następnie dodajemy osobno rangi różnic dodatnich i ujemnych. Wartość statystyki T jest mniejszą z tych sum. Rozmiar próbki (N) liczymy tak, jak w teście znaków. Za pomocą tych dwóch danych odczytujemy poziom istotności dla testu jedno lub dwustronnego. Test Wilcozona jest testem mocniejszym, ponieważ przy próbie o tym samym rozstępie prawidłowo odrzuca hipotezę zerową, a test znaków tego nie czyni. Dzieje się tak, dlatego że test Wilcozona wykorzystuje więcej informacji niż test znaków.

Parametryczny Test t związany lub Test t Studenta, dostarczający wartości t , ułożony przez Gosseta, zastępuje nieparametryczny test Wilcozona i jest od niego mocniejszy. Testy t są oparte na liczeniu średnich i rozkładu wyników wokół średniej. Rozmiar próbki (N) oblicza się za pomocą stopni swobody (SS). Stopnie swobody to sumaryczna liczba elementów próbki, która musi być znana, gdy znana jest ogólna suma, aby można było uzupełnić pozostałe elementy brakujące. Na przykład, jeśli mamy podane cyfry 1, 5, 4, 3 i wiemy, że suma wynosi 16, to brakującą cyfrą jest 3. Stopnie swobody w statystyce t to suma par wyników minus 1.

W przypadku próbek niezwiązanych, można użyć testu nieparametrycznego, jakim jest test U Manna-Whitneya lub testu parametrycznego, jakim jest Test t niezwiązany (Hatch i Lazarton 1991, Clegg 1994). Test U Manna-Whitneya, dający statystykę U , może być zastosowany przy skali co najmniej przedziałowej. Test ten jest równy testowi Wilcozona, ale nie ma tu potrzeby łączenia wyników w pary, a grupy wyników mogą być nierównych rozmiarów. Analizę rozpoczyna się od krótszej listy wyników, a następnie dokonuje się rangowania obu list razem. W wyniku tego testu powstają dwie statystyki: U (mniejszy wynik) i U' (większy wynik). Aby odczytać poziom istotności w tablicy statystycznej, bierzemy pod uwagę rozmiar próbek (N_A i N_B). W miejscu ich skrzyżowania, tablica pokazuje dwie wartości, które nie mogą być przekroczone przez wartość U . Na górze podana jest wartość dla testu dwustronnego, a na dole dla jednostronnego. Jeśli zostały spełnione trzy warunki dla zastosowania testów parametrycznych, a wyniki są przedstawione w skali co najmniej przedziałowej, to można zastosować Test t niezwiązany, który zastępuje test U Manna-Whitneya i jest od niego mocniejszy. Należy pamiętać o tym, że w teście tym każda lista ma swoje stopnie swobody, które oblicza się sumując elementy z obu list i odejmując 2. Szczegółowe plany obliczeń dotyczące wszystkich wymienionych tu testów statystycznych są w przejrzysty sposób przedstawione przez Clegga (1994), a obliczenia wykonywane przy pomocy programu Statistica przez Augustyńską (2003).

Na podstawie wskaźnika prawdopodobieństwa odczytanego w tabeli statystycznej dla hipotezy jednostronnej lub dwustronnej oraz na podstawie zakładanego poziomu prawdopodobieństwa należy podjąć decyzję o przyjęciu lub odrzuceniu hipotezy zerowej. Przy czym prawidłowe wnioskowanie jest następujące: „Hipoteza zerowa nie może być odrzucona” (por. „Hipoteza zerowa została przyjęta”) (Clegg 1994:75).

Dziesiąty, a zarazem ostatni krok postępowania eksperymentalnego, to podjęcie decyzji czy hipoteza eksperymentalna może być przyjęta czy odrzucona. Hipoteza eksperymentalna może być przyjęta, jeśli hipoteza zerowa jest odrzucona. Jeśli hipoteza zerowa nie może być odrzucona, należy odrzucić hipotezę eksperymentalną. Przykład prawidłowego sposobu zapisania wniosku jest następujący: „Rezultaty analizy statystycznej były istotne na poziomie $p < 0,05$ (jednostronny test Wilcozona, $T = 6$, $N = 9$), a zatem hipoteza eksperymentalna została przyjęta. Wyciągnięto wniosek, że...” (Clegg 1994:82).

Należy jednakże pamiętać o tym, że w przeprowadzaniu eksperymentu badacz jest narażony na dwa rodzaje błędu (Clegg 1994). Błąd I rodzaju to zdecydowanie, że zbiory wyników są różne i przyjęcie hipotezy eksperymentalnej, podczas gdy w rzeczywistości ZN nie wpłynęła na ZZ w przewidywany sposób, lecz wystąpił 1 przypadek losowy na 20, czy 1 na 100 przypadków losowych. Błąd II rodzaju to zdecydowanie, że zbiory wyników w rzeczywistości się nie różnią, zaakceptowanie hipotezy zerowej, wyciągnięcie wniosku, że eksperyment się nie udał, podczas gdy w rzeczywistości się udał tylko, że metoda wartościowania ZZ nie była wystarczająco precyzyjna, aby wykazać subtelne zmiany.

BIBLIOGRAFIA

- Augustyńska, U. 2003. *Statystyczna analiza danych w badaniach pedagogicznych z wykorzystaniem programu Statistica*. Częstochowa: Wydawnictwo Wyższej Szkoły Pedagogicznej w Częstochowie.
- Brown, H. D. 1991. *Understanding Research in Second Language Learning. A Teacher's Guide to Statistics and Research Design*. Cambridge: Cambridge University Press.
- Clegg, F. 1994. *Po prostu statystyka. Kurs dla studentów nauk społecznych*. Warszawa: WSiP.
- Hatch, E. i H. Farhady. 1982. *Research Design and Statistics for Applied Linguistics*. Rowley, Mass.: Newbury House.
- Hatch, E. i A. Lazarson. 1991. *The Research Manual. Design and Statistics for Applied Linguistics*. Rowley, Mass.: Newbury House.
- Larsen-Freeman, D. and M. Long. 1991. *An Introduction to Second Language Acquisition Research*. New York: Longman.