

Anna Michońska-Stadnik

Trafność i rzetelność w badaniach glottodydaktycznych

Lingwistyka Stosowana / Applied Linguistics / Angewandte Linguistik nr 4,
31-40

2011

Artykuł został opracowany do udostępnienia w internecie przez Muzeum Historii Polski w ramach prac podejmowanych na rzecz zapewnienia otwartego, powszechnego i trwałego dostępu do polskiego dorobku naukowego i kulturalnego. Artykuł jest umieszczony w kolekcji cyfrowej bazhum.muzhp.pl, gromadzącej zawartość polskich czasopism humanistycznych i społecznych.

Tekst jest udostępniony do wykorzystania w ramach
dozwolonego użytku.

Anna MICHONSKA-STADNIK

Uniwersytet Wrocławski

Trafność i rzetelność w badaniach glottodydaktycznych

1. Wstęp

Trafność i rzetelność są głównymi cechami wiarygodności każdego badania, a więc także badania glottodydaktycznego (GD). Nie możemy mówić o poprawności i wiarygodności badania, jeśli marna będzie jakość samych danych, a także ich analiza. Sposób traktowania tych danych powinien być pełny i ścisły, tzn. wystarczająco pełny i wystarczająco ścisły w stosunku do potrzeb celu badania i związanych z nim pytań badawczych. Ograniczenie takie bierze się z niemożności zapewnienia, szczególnie w glottodydaktyce, takiej idealnej sytuacji, w której grupa badana i narzędzia badawcze odzwierciedlałyby *wszystkie* aspekty procesu nauczania i uczenia się języka drugiego (N/U J2). Stąd też pytanie badawcze: *Jak wiek ucznia wpływa na proces N/U J2?* musi pozostać bez jednoznacznej odpowiedzi. Aby spróbować na nie w pełni odpowiedzieć, potrzeba kilkudziesięciu narzędzi badawczych, albowiem powinniśmy najpierw zdefiniować, jak rozumiemy proces N/U J2, czyli określić tenże konstrukt w stosunku do rzeczywistości dydaktycznej. Co więcej, sama zmienna wieku ucznia też wymagałaby analizy bardziej szczegółowej, co najmniej w trzech odrębnych grupach: dzieci, młodzieży i dorosłych. Z kolei sformułowanie *Nauczanie sprawności czytania ze zrozumieniem* nie może być uznane za problem badawczy, ponieważ jego specyfika wymaga opisowego potraktowania zagadnienia, a nie rozwiązania problemu. Trudno jest więc również w takim przypadku rozmawiać o trafności i rzetelności w stosunku do pytań badawczych, które będzie niezwykle trudno do takiego ogólnego zagadnienia sformułować. Dlatego też wskazane jest, aby badacz wybrał pewien wycinek rzeczywistości dydaktycznej, zgodnie ze swoimi realnymi warunkami i możliwościami. Dopiero wtedy możemy określić, czy badanie takie może być wiarygodne i poprawne w stosunku do tego właśnie wybranego wycinka rzeczywistości. Miarą poprawności i wiarygodności badania jest stopień jego trafności i rzetelności.

2. Pojęcie trafności

U podstaw trafności badania leży skojarzenie abstrakcyjnego pojęcia (konstruktu) z jego empirycznym przejawem. *Trafność teoretyczna* (ang. *construct validity*) określa, na ile wskaźniki (ang. *exponents*), odpowiadające im narzędzia gromadzenia danych oraz interpretacja wyników badania odzwierciedlają pojęcie teoretyczne, będące w polu zainteresowania badacza. Dla przykładu: test teoretycznie dotyczący historii Europy, a w rzeczywistości uwzględniający przede wszystkim historię Wielkiej Brytanii, nie jest testem trafnym; z kolei test teoretycznie badający rozumienie ze słuchu tekstów obcojęzycznych, a w rzeczywistości sprawdzający rozumienie słownictwa, także nie jest testem trafnym; kwestionariusz, którego celem jest zbadanie poziomu lęku językowego w klasie, a w rzeczywistości pytający uczniów wyłącznie o atmosferę w klasie, także nie będzie trafnym narzędziem badawczym. Zadanie testowe na ustnym egzaminie maturalnym, które wymaga od ucznia reakcji na określoną przewidzianą sytuację językową, także nie jest testem trafnym, ponieważ nie sprawdza spontanicznej umiejętności użycia języka w sytuacjach dnia codziennego; nie jest więc to test faktycznie oceniający sprawność mówienia w języku obcym. Trafność samego instrumentu badawczego jest zależna od tego, czy i w jakim stopniu narzędzie mierzy to, co powinno mierzyć.

Trafność empiryczna z kolei jest przedstawiana jako cecha właściwa wskaźnikom i określa, na ile wyczerpują one dane zjawisko. Na przykład, wskaźnikami *płynności wypowiedzi* ucznia w języku obcym będą: długość frazy, długość pauz, liczba powtórzeń, liczba zawahań, liczba wypowiedzianych sylab na sekundę. Nie będą nimi na pewno liczba błędów gramatycznych i błędów w wymowie, ponieważ te ostatnie są już wskaźnikami innego konstruktów, a mianowicie poprawności językowej. Badacz powinien umieć określić, na ile ta pierwsza grupa wskaźników wyczerpuje ocenę wypowiedzi i być może dodać jeszcze inne. Z kolei poziom *przyswojenia* słownictwa zostanie wyczerpany, gdy badacz zastosuje dwa wskaźniki: stopień zapamiętania, czyli rozpoznania słów i ich znaczenia oraz umiejętność ich poprawnego użycia w komunikacji ustnej i pisemnej. Zastosowanie tylko jednego z tych wskaźników nie wyczerpuje konstruktów przyswojenia, a tylko zapamiętanie lub użycie słownictwa. Ten temat rozwija bardzo dogłębnie Michel Paradis w obu swoich książkach (2004, 2009). Na marginesie warto wspomnieć, dla uwagi początkujących badaczy, że słownictwo ma dwojaki charakter i wymaga zaangażowania dwóch różnych rodzajów wiedzy. Jeśli odnosimy się tylko do odtworzenia z pamięci żądanych słów lub ewentualnego podania ich znaczenia, korzystamy wtedy z wiedzy deklaratywnej. Natomiast jeśli spodziewamy się użycia danej grupy słów w mowie lub w piśmie, mamy wtedy do czynienia z zupełnie innym procesem, który jest aktywizowany w pamięci proceduralnej. O tych ważnych aspektach istoty przyswojenia słownictwa należy zawsze pamiętać podczas układania wszelkich testów sprawdzających znajomość słów, czyli stopień ich przyswojenia. Tak więc test typu „połącz słowa z ich definicjami” będzie sprawdzał konstrukt zapamiętania, a test wymagający napisania krótkiego tekstu z użyciem wymaganych słów, będzie

sprawdzał umiejętność ich użycia. Najczęściej, aby rzeczywiście sprawdzić stopień przyswojenia słownictwa stosuje się oba rodzaje testów, o czym wspomniałam powyżej, i wtedy dopiero takie narzędzie będzie rzeczywiście trafne.

Zoltan Dörnyei (2007) pisze też o trafności *wewnętrznej* i *zewnętrznej*. Ta pierwsza oznacza, że wynik badania został ustalony na podstawie właściwych danych i nie jest efektem zakłóceń wynikających z innych czynników niż zmienna niezależna. O tych czynnikach piszemy bardziej szczegółowo gdzie indziej (W. Wilczyńska, A. Michońska-Stadnik 2010), tu natomiast podam tylko kilka przykładów:

- Zmiana liczby uczestników grupy badanej; może wynikać po prostu z nieobecności uczniów na lekcji, ale także z niechęci do brania udziału w samym badaniu. Sytuacja taka nie musi być groźna dla wiarygodności badania, chyba że grupa jest niewielka i „wypadnięcie” nawet jednego uczestnika może już powodować zakłócenia.
- Efekt nabierania doświadczenia (praktyki); uzyskiwanie coraz lepszych wyników, zwłaszcza w badaniu podłużnym, nie musi być koniecznym rezultatem naszego postępowania, ale po prostu uczniowie mogą nabierać doświadczenia w rozwiązywaniu samych testów. Bardzo często dostrzegamy to zjawisko podczas wszelkiego rodzaju egzaminów państwowych, np. gimnazjalnych lub maturalnych. Uczniowie, którzy zdali egzamin maturalny na wysoką ocenę, bardzo często nie są w stanie zrozumieć wykładu z gramatyki opisowej czy literatury angielskiej na I roku studiów anglistycznych, ponieważ ich dobra ocena była rezultatem nabytej umiejętności rozwiązywania testów, a nie rzeczywistej znajomości języka angielskiego.
- Efekt dojrzewania; rozwój psychiczny i społeczny uczniów, zwłaszcza młodych, daje w rezultacie coraz lepsze wyniki testów językowych albo coraz pełniejsze odpowiedzi na pytania w ankietach otwartych.
- Efekt obserwatora; w obecności obserwatora osoby badane mogą zachowywać się inaczej, co może zakłócać wyniki badania.
- Efekt Hawthorne’a: osoba badana w obecności obserwatora stara się zachowywać lepiej, bardziej standardowo.
- Efekt aureoli: osoba badana podświadomie spełnia oczekiwania badacza, domyśla się, co należy zrobić lub napisać, aby zadowolić osobę przeprowadzającą badanie, zwłaszcza jeśli jest to sam nauczyciel prowadzący klasę.

Badacze często nie są w ogóle świadomi istnienia takich czynników zakłócających i dobre wyniki badania przypisują wyłącznie działaniu zmiennej niezależnej, co jest oczywistym niedopatrzaniem.

Trafność zewnętrzna dotyczy możliwości uogólniania wyników poza daną zbiorowość lub moment badania. W badaniach GD jest to również utrudnione, gdyż warunki potrzebne do uogólnienia wyników rzadko kiedy pozostają do końca spełnione. Na przykład tzw. losowy dobór próby nie będzie łatwy w przypadku badania wpływu różnic indywidualnych na proces N/U J2. Już samo istnienie różnic indywidualnych, na przykład takich, jak tolerancja/nietolerancja wieloznaczności, ogranicza możliwości losowego doboru próby. Jesteśmy zainteresowani uczniami charakteryzującymi się obecnością lub nieobecnością pewnej cechy indywidualnej,

a więc nie będzie to już próba czysto losowa, ale celowa. W przypadku chęci użycia typowej statystyki indukcyjnej, trudno będzie, rozpatrując wyniki takiego badania, uzyskać ich tzw. rozkład normalny, o czym nieco więcej powiemy w dalszej części tego artykułu.

Podsumowując, trafność jest zawsze rozumiana jako adekwatność czegoś do czegoś:

- konstrukt do rzeczywistości badanej,
- wskaźników (eksponentów) do konstruktów,
- narzędzi badawczych do wskaźników.

3. Pojęcie rzetelności

Rzetelność dotyczy spójności, czyli stopnia zgodności wyników, jakie można uzyskać za pomocą narzędzi i procedur, badając daną grupę czy populację w różnych okolicznościach.

Teorie rzetelności rozwinęły się przede wszystkim w dyscyplinach związanych z psychologią, a podwaliny pod nie położył Charles Spearman na początku XX wieku. Rzetelność daje badaczom informację na temat stopnia przypadkowej niespójności oraz niesystematycznej fluktuacji indywidualnych wyników w pomiarze. Trzeba pamiętać, że teoretycznie trafny pomiar jest zawsze rzetelny, natomiast rzetelny pomiar nie zawsze musi być trafny. Inaczej mówiąc, narzędzie pomiaru może być rzetelne, ale nie musi w sposób trafny odzwierciedlać danego konstruktów. Szczególnie widać to w badaniach GD, gdy np. rzekomy test na rozumienie tekstu ze słuchu, zawierający elementy poprawności gramatycznej i leksykalnej jest rzetelny, ale nie odzwierciedla trafnie wybranego konstruktów. Więcej na temat rzetelności powiemy w konkretnych przykładach jej interpretacji w poszczególnych typach badań.

4. Trafność i rzetelność w badaniach ilościowych

Obserwowana często fascynacja badaniami ilościowymi w psychologii, pedagogice oraz innych naukach społecznych przyczyniła się także do ich popularności w GD. W glottodydaktyce pojawiły się jednak konkretne problemy związane z aplikacją badań statystycznych, tak często stosowanych w opracowaniu wyników badań uważanych za eksperymentalne. Nieobecność tzw. zera bezwzględnego nie pozwala, na przykład, na zastosowanie jednej ze skal porządkowania danych, tzw. skali rangowej. Trudno jest przecież mówić o zerowej kompetencji komunikacyjnej w języku obcym, zerowej znajomości słownictwa lub zerowej wiedzy metakognitywnej. Oprócz braku zera bezwzględnego pojawiają się trudności z uzyska-

niem rozkładu normalnego wyników (krzywa Gaussa) w grupie badanej. Badania korelacyjne, w takich warunkach polegają głównie na określeniu tendencji w występowaniu badanej zależności pomiędzy zmiennymi, ale nie są w stanie określić związku przyczynowo-skutkowego pomiędzy nimi. Podobnie porównanie średnich statystycznych metodą *testu t* powinno być właściwie wykonywane w warunkach istnienia rozkładu normalnego wyników. Pamiętać należy, że im większa grupa badana, tym większa szansa na uzyskanie prawidłowej krzywej Gaussa. Tymczasem w badaniach glottodydaktycznych najczęściej mamy do czynienia z grupą wielkości klasy szkolnej, którą określa się jako małą lub wręcz niewystarczającą. Co więcej, rozkład normalny wyników w przypadku typowych testów językowych z punktu widzenia nauczyciela jest wręcz niepożądany. Dla nauczyciela i klasy korzystniejszy będzie rozkład skośny negatywnie (lewostronny), czyli ten z przewagą wysokich wyników grupy. Podobnie, losowy dobór próby nie zawsze jest możliwy, czy pożądanym, o czym pisałam powyżej.

W badaniach ilościowych przyjmuje się zazwyczaj następujące kryteria poprawności:

1. Rzetelność wyników (określana za pomocą różnych technik, o których poniżej).
2. Trafność konstruktywna, czyli zaufanie do interpretacji wyników, iż faktycznie dotyczy ona badanego pojęcia.
3. Trafność badania, czyli trafność wewnętrzną i zewnętrzną, które opisane zostały powyżej.

Istnieją trzy główne techniki oceny rzetelności w badaniu ilościowym:

- a. *Technika powtarzania pomiaru.* U jej podstaw leży przekonanie, że jeśli powtórzymy w tej samej grupie pomiar tym samym narzędziem po jakimś okresie czasu, powinniśmy otrzymać podobne albo wręcz takie same wyniki. Odchylenie wzajemne tych wyników mierzy się za pomocą korelacji. Współczynnik korelacji powinien wtedy osiągnąć wartość co najmniej 0,75, mierzony z użyciem wzoru α Cronbacha albo z użyciem wzoru r Pearsona. Wadą tej techniki jest to, że w tej samej grupie nigdy nie uda się stworzyć idealnie takich samych warunków w pierwszym oraz w powtórzonym pomiarze. Każdy badacz powinien być świadomy tego, że uzyskany wynik tego samego testu może być lepszy albo gorszy od poprzedniego z powodów czysto pozajęzykowych, np. pora dnia, pogoda, test z innego przedmiotu tego samego dnia, wreszcie to, że badani mogli po prostu zapamiętać wyniki poprzedniego testu. Natomiast rzetelność kwestionariuszy można z powodzeniem badać tą techniką.
- b. *Technika połówkowa.* Jest to skuteczny sposób sprawdzania rzetelności, zwłaszcza testów, ponieważ nie muszą one być powtarzane. Wystarczy badaną grupę podzielić na dwie części, np. według numerów parzystych i nieparzystych w dzienniku, a następnie obliczyć współczynnik korelacji testu w tych dwóch częściach. Istotna jest w tym przypadku ilość zadań testu. Im będzie ich więcej, tym współczynnik korelacji może być wyższy.
- c. *Technika zgodności wewnętrznej.* Do obliczenia rzetelności służy w tym przypadku wzór Kudera-Richardsona. Określa on, w jakim stopniu odpowiedzi poszczególnych pozycji testu są zgodne z wynikiem całego testu. Poziom takiej

rzetelności określa się w skali od 0 do 1; oczywiście, im bliżej 1, tym wyższa jest rzetelność narzędzia badawczego. Przyjmuje się, że w taki sposób mierzona rzetelność nie powinna być niższa od 0,60 (por. J. Rysiewicz 2010). Co ciekawe, zanim przystąpi się do obliczania współczynnika zgodnie z tym wzorem, należy odrzucić wszystkie zadania testu, które były albo za łatwe albo za trudne, bo na przykład wykonali je poprawnie wszyscy testowani uczniowie albo z kolei nie wykonał ich nikt. Procedury tzw. analizy jednostkowej (ang. *item analysis*) można znaleźć w instrukcjach towarzyszących opisowi stosowania wzoru Kudera-Richardsona. Już sama taka wstępna analiza jest pożyteczną wskazówką dla badacza. Jeśli bowiem okaże się, że 50% zadań testowych zostało odrzuconych już w analizie jednostkowej, nie ma sensu wykonywanie dalszych obliczeń na przykładzie tego narzędzia, ponieważ nie jest ono rzetelne.

Osobnym problemem w badaniach wykorzystujących kwestionariusze, a więc przede wszystkim w badaniach ilościowych, jest używanie narzędzi skonstruowanych w innych krajach, a więc dostosowanych do innych warunków społecznych i kulturowych. Częste zafascynowanie rezultatami badań przeprowadzonych w Stanach Zjednoczonych czy Wielkiej Brytanii powoduje to, że polski badacz nie zwraca uwagi na zupełnie inny kontekst tych procedur. Przede wszystkim, są to badania przeprowadzane na osobach uczących się angielskiego jako języka drugiego, gdzie nauczanie sformalizowane w klasie wspomaganie jest znacznie poprzez obecność języka angielskiego poza szkołą. Takich warunków nie ma osoba ucząca się tego języka w Polsce, gdzie np. język angielski jest nauczany jako obcy, a nie jako drugi. Ankiety i testy przeprowadzane wśród polskich uczniów, nawet jeśli będą powielały wyniki tych samych narzędzi stosowanych w krajach anglosaskich, powinny być inaczej interpretowane, ponieważ wykonywane są w zupełnie innych warunkach, w zupełnie innej populacji. Co więcej, nie będą odzwierciedlały tych samych tendencji ponieważ zostały skonstruowane dla innych celów. Szczególnie niebezpieczne jest bezkrytyczne stosowanie kwestionariuszy dotyczących tzw. różnic indywidualnych, inaczej zmiennych ucznia, czyli różnic poznawczych, emocjonalnych, opiniotwórczych. Takie ankiety czy testy powinny być najpierw starannie przetłumaczone na język polski, potem każdy punkt powinien być zweryfikowany pod względem jakościowym, tzn. sprawdzony czy wyrażony za jego pomocą aspekt danego konstruktów jest zrozumiały dla polskiego ucznia, czy faktycznie da się umiejscowić w polskiej rzeczywistości badawczej, czy nie jest nam obcy kulturowo. Dopiero potem taką ankietę należałoby sprawdzić poprzez badanie pilotażowe i wtedy zdecydować o jej przydatności dla badania właściwego. Gdy takie warunki nie zostaną spełnione, można poddać w wątpliwość poprawność narzędzia badawczego, a tym samym poprawność całego badania. Dalsze wskazówki w tej kwestii zawiera książka J. Brzezińskiego (2002).

Na podstawie własnego doświadczenia z badaniami na temat stosowania strategii uczenia się, mogę stwierdzić, że bardzo popularna i często stosowana ankieta R. Oxford (1990) SILL (ang. *Strategy Inventory for Language Learning*), w swojej części dotyczącej używania strategii emocjonalnych jest dla naszych uczniów obca

kulturowo. Nie jest bowiem popularne w naszym kraju stosowanie technik relaksacyjnych przed zajęciami językowymi, przed testami czy sprawdzianami. Nie jesteśmy też przyzwyczajeni formułować dla zachęty pozytywne stwierdzenia na temat samych siebie i swoich postępów w nauce języka obcego. Uczniowie poproszeni o napisanie czegoś pozytywnego o sobie w takim kontekście, przeważnie poprzestają na jednym lub dwóch zdaniach i zakłopotani przyznają się, że nic więcej nie potrafią napisać.

5. Trafność i rzetelność w badaniach jakościowych

Według SAGE Research Methods Online (SRMO), trafność badań jakościowych stanowi pewne wyzwanie. W badaniach ilościowych trafność określa, na ile wyniki adekwatnie odzwierciedlają rzeczywistość. Natomiast w badaniu jakościowym trafność oznacza stopień adekwatności wyników do opisywanego kontekstu badawczego. To oznacza, że nawet jeśli nie mamy bezpośredniego dostępu do całości badanej rzeczywistości i nasze sądy mogą być jednostronne, wciąż możemy twierdzić, że nasze badanie jest trafne i wiarygodne, bo odzwierciedla prawdziwość tego danego i dostępnego nam wycinka rzeczywistości. Badanie musi być wykonane zgodnie z naszą najlepszą wiedzą, z uwzględnieniem wszelkich możliwych okoliczności. Niektórzy twierdzą, że trafność takich badań należy bardziej rozpatrywać w kategoriach moralnych niż epistemologicznych.

Z kolei rzetelność w badaniach jakościowych może mieć także różne interpretacje. Badanie jakościowe jest rzetelne, jeśli:

1. Różne techniki gromadzenia danych, np. wywiady lub obserwacje, wykazują istnienie podobnych lub wręcz tych samych zjawisk.
2. Dokładna replikacja badania jest niemożliwa, ponieważ warunki zawsze będą nieco inne. W związku z tym potrzebny jest bardzo dokładny opis sposobów gromadzenia danych i ich szczegółowa analiza.
3. Według niektórych badaczy, rzetelność nie musi mieć określonego statusu w badaniach jakościowych, ponieważ i tak inny badacz zgromadzi w podobnym kontekście inne dane i inaczej je zinterpretuje. Tego poglądu nie podzielam, ponieważ moim zdaniem nadmierny relatywizm nie prowadzi do niczego i neguje sensowność jakichkolwiek prac badawczych.

Jak można, w takim razie, osiągnąć wiarygodność badań jakościowych i zapewnić ich odpowiednią trafność i rzetelność? Podam tutaj kilka możliwości:

1. Rzetelność narzędzia badawczego określa się poprzez zastosowanie techniki kilku oceniających. Jeśli oceny ich są zbieżne, narzędzie jest rzetelnie zaplanowane. Ta uwaga dotyczy przede wszystkim takich narzędzi, jak: ankiety otwarte, dłuższe wypowiedzi ustne oraz pisemne.
2. Aby zapewnić rzetelność, badanie musi być prowadzone w sposób systematyczny, zwykle w określonych odstępach czasowych. Odstępy te nie muszą być regularne, lepiej jeśli wyznacza je rytm, np. roku szkolnego.

3. Trzeba zapewnić ciągłość badań z tymi samymi osobami. W przypadku niewielkich grup inny skład grupy badanej może całkowicie zmienić wyniki badania.
4. Trzeba zapewnić staranny dobór osób badanych pod względem celu badania. Należy uwzględnić czynniki mogące zakłócić w taki czy inny sposób zbieranie danych. Na przykład, osoby ze skłonnościami do ciągłych żartów, do fantazjowania, do nieszczerości, mogą brać udział w badaniu, ale interpretacja ich odpowiedzi powinna uwzględniać ich cechy osobowościowe.
5. Należy angażować osoby badane w możliwie jak najwięcej sytuacji badawczych z użyciem różnych technik. Jest to tak zwana *triangulacja*, bardzo pożyteczna technika, gdzie badaną grupę obserwuje się trzykrotnie, na różne sposoby (np. wywiady, obserwacja, testy).
6. W sprawozdaniu z badania autor powinien w sposób szczególny uwypatnić te aspekty, które świadczą o wiarygodności jego wyników, o bezstronności badacza, o dojrzałości sądów. Należy też podać wszystkie sytuacje nietypowe, trudności i problemy, co uwiarygodni dany opis (W. Wilczyńska, A. Michońska-Stadnik 2010:141-142).

Wobec szczególnego charakteru GD, której materiał badawczy nie zawsze poddaje się analizie ilościowej ze względu na wielowymiarowość problematyki (badania ludzi w sytuacji formalnego uczenia się w klasie, zmienne warunki, problemy z rozkładem normalnym wyników), widzimy potrzebę uzupełniania się wzajemnego obu głównych typów badań. Jak pisze K. Łęska (2010: 220), „Niewątpliwą zaletą badań połączonych jest fakt, że przyczyniają się one do lepszego zrozumienia zjawisk złożonych”.

6. Badania mieszane

W badaniach mieszanych dominuje zasada triangulacji, ponieważ w GD jako nauce społecznej, nie można bezpiecznie polegać na wynikach uzyskanych za pomocą jednej tylko techniki gromadzenia danych, kojarzonej zwykle z jedną określoną metodologią. Na przykład wywiad otwarty czy notatki z lekcji kojarzą się zwykle z metodą obserwacji, a więc z jakościowym typem badań. Z kolei testy czy ankiety zamknięte (np. oceniane według skali Likerta), kojarzą się z metodami eksperymentalnymi, czyli z ilościowym typem badań. Badania mieszane, czyniąc użytek z różnych sposobów gromadzenia danych i ich analizy, zwiększają wiarygodność otrzymanych wyników.

Podam teraz przykład badania mieszanego z użyciem triangulacji: Chcąc zbadać postępy uczniów w przyswajaniu słownictwa języka obcego po wprowadzeniu szkolenia w zakresie wykorzystania strategii pamięciowych, możemy wykorzystać następujące narzędzia zbierania danych:

1. Testy rozpoznawania i użycia wybranych słów (ilościowe) przed i po szkoleniu w zakresie strategii (postępy uczniów).

2. Technikę głośnego myślenia podczas czytania tekstu obcojęzycznego ze zrozumieniem. Wykaże nam użycie tych strategii (jakościowe).
3. Wywiad z uczniami (jakościowe) na temat przydatności strategii pamięciowych w rozpoznawaniu i używaniu słownictwa.

Oczywiście, każde z tych narzędzi wymaga osobnego podejścia do analizy otrzymanych danych.

Podam teraz przykłady badań połączonych, których jest 4 rodzaje (por. SAGE Research Methods Online). Kursywą zaznaczono badanie, które ma przewagę, to znaczy jest bardziej dogłębnie wykorzystane do realizacji danego celu badawczego. Badanie zapisane zwykłym drukiem jest badaniem pomocniczym. Kolejność wykonywania działań badawczych zależy od celu badania.

1. *Jakościowe/ ilościowe*. Badacz prowadzi wywiad półotwarty, którego celem jest wybranie konkretnych problemów do opracowania w ankiecie zamkniętej, która będzie potem analizowana z wykorzystaniem statystyki (np. sposoby uczenia się słownictwa/ ankietę szczegółową na temat stosowanych strategii).
2. *Ilościowe/ jakościowe*. Przeprowadzamy kwestionariusz, na podstawie którego wybieramy jednostki o szczególnych cechach charakterystycznych, które potem poddaje się intensywnej obserwacji (np. na podstawie ankiety dotyczącej stopnia tolerancji wieloznaczności wybieramy jednostki ze szczególnie niewielką tolerancją i obserwujemy ich postępy w mówieniu w języku obcym).
3. *Ilościowe/ jakościowe*. Kwestionariusze wskazują na istnienie pewnych zależności pomiędzy zmiennymi. Następnie jest prowadzona obserwacja lub wywiad otwarty celem zbadania czynników, które mogą być odpowiedzialne za tę zależność (np. kwestionariusz i test wykazały, że uczniowie z małą tolerancją wieloznaczności mają problemy z gramatyką J2. Obserwacja wskazuje na dodatkowe czynniki, np. niepewność, uzależnienie od języka ojczystego).
4. *Jakościowe/ ilościowe*. W obserwacji widać występowanie pewnych interesujących zależności. Badanie ankietowe ma na celu uchwycenie, na ile te zależności występują systematycznie (np. zaobserwowano, że osoby ciche i spokojne czytają lepiej w J2. Test na introwersję i na czytanie ze zrozumieniem, skorelowane ze sobą, być może pokażą systematyczną zależność pomiędzy tymi zmiennymi).

Korzyści wynikające z podejścia mieszanego są niewątpliwe, co wykazane zostało na przykładach. Obie tradycje badawcze reprezentują, co prawda, odrębne podejścia do rzeczywistości: (a) ilościowe: pozytywizm i obiektywizm; (b) jakościowe: interpretacyjność i konstruktywizm, ale dopiero razem dają one najbardziej adekwatny obraz rzeczywistości badanej.

Niektórzy uważają, że nie powinno się tych badań łączyć. Niemniej jednak, w GD wiele można osiągnąć w rezultacie połączenia mocnych stron obu tradycji badawczych, co wiąże się z pewnością ze zwiększonymi kosztami, większym zaangażowaniem badacza, ale wartość takich badań nie ulega wątpliwości. Z pewnością większa będzie ich wiarygodność i poprawność.

BIBLIOGRAFIA

- BRZEZIŃSKI J. (2002), *Metodologia badań psychologicznych*. Warszawa: PWN.
- DÖRNYEI Z. (2007), *Research Methods in Applied Linguistics*. Oxford: Oxford University Press.
- ŁĘSKA K. (2010), *Zalety stosowania połączonych metod badawczych w analizie interakcji na lekcjach języka obcego*, (w:) *Neofilolog*, 34 (2010), 217-228.
- OXFORD R. (1990), *Language learning strategies: What every teacher should know*. Boston, M.A.: Heinle.
- PARADIS M. (2004), *A Neurolinguistic Theory of Bilingualism*. Amsterdam/Philadelphia: John Benjamins.
- PARADIS M. (2009), *Declarative and Procedural Determinants of Second Languages*. Amsterdam/Philadelphia: John Benjamins.
- RYSIEWICZ J. (2009), *Uchybienia metodologiczne w badaniach ilościowych w językoznawstwie stosowanym*, (w:) *Neofilolog*, 32 (2009), 177-188.
- SAGE Research Methods Online (2010), *Encyclopedia of Social Science Research Methods*. DD: 20.12.2010.
- WILCZYŃSKA W., MICHONSKA-STADNIK A. (2010), *Metodologia badań w glottodydaktyce. Wprowadzenie*. Kraków: AVALON/ Flair.

Validity and reliability in second/foreign language acquisition research

For years second language acquisition (SLA) studies have been using research methods originally employed in other social sciences, e.g. in psychology and sociology. A unique character of SLA investigations, resulting from their specific conditions and actions (i.e. researching language skills versus knowledge), allowed for a number of developments in research methods, especially in the area of validity and reliability of their procedures and instruments. Statistical measures, exceptionally popular about ten years ago, appeared to be not fully adequate to give a true picture of the complex process of language acquisition and its variables. What is proposed nowadays are mixed research types, both qualitative and quantitative, successfully applying instruments appropriate for each type respectively. Consequently, the problem of research credibility should be considered with more attention. The concepts of validity and reliability of research studies and instruments should be regarded from both the qualitative and quantitative perspective. Each research type requires a different approach to these two basic constructs, and studies of mixed character will have to make an adequate use of qualitative and quantitative interpretations of validity and reliability.