

Michał Modzelewski

Jolanta Pisarek

Instytut Badań Edukacyjnych

Zastosowanie teorii detekcji sygnałów do analizy rzetelności systemu obserwacyjnego Argos

Celem niniejszego artykułu jest opis próby zastosowania teorii detekcji sygnałów oraz metody ROC do analizy rzetelności systemu obserwacyjnego Argos, czyli aplikacji komputerowej wspomagającej obserwację przebiegu lekcji w szkołach podstawowych. Badania obserwacyjne stanowiły jeden z etapów badania podłużnego szkolnych uwarunkowań efektywności kształcenia, który został zrealizowany w II semestrze roku szkolnego 2012/2013. Badanie SUEK realizowane przez Instytut Badań Edukacyjnych stanowi część projektu systemowego pt. „Badanie jakości i efektywności edukacji oraz instytucjonalizacja zaplecza badawczego”, współfinansowanego ze środków Europejskiego Funduszu Społecznego w ramach Programu Operacyjnego Kapitał Ludzki. Dane uzyskane podczas spotkań fokusowych z obserwatorami, w trakcie których 34-42 obserwatorów kodowało równoległe lekcje, bazując na tym samym materiale filmowym, pozwoliły na umieszczenie obserwatorów na płaszczyźnie ROC, czyli na wizualizację decyzji podejmowanych przez obserwatorów w trakcie klasyfikacji zachowań nauczyciela. W przedstawionej w artykule analizie każdy z obserwatorów został opisany pod względem czułości oraz specyficzności decyzji podejmowanych w trakcie równoległego kodowania lekcji na spotkaniu fokusowym. Wyniki analizy pozwoliły opisać obserwatorów pod względem zdolności identyfikacji i rejestracji zachowań nauczycieli oraz umiejętności unikania błędnych decyzji.

Potrzeba rzetelnego opisu procesu dydaktycznego

Osiągnięcia szkolne uczniów determinowane są przez różnorodne czynniki, wśród których kluczowe miejsce zajmuje środowisko szkolne. Na szczególną uwagę zasługuje klimat klasy, warunkowany przez relacje pomiędzy uczniami a nauczycielem, komunikację w grupie uczniów, a także metody wychowawcze i dydaktyczne wykorzystywane przez nauczyciela.

Interakcje uczniów z nauczycielami stanowią istotny mechanizm rozwojowy umiejętności samoregulacyjnych dzieci, w znaczący sposób wpływając na ich zaangażowanie w proces uczenia (Vygotsky, 1978). Pozytywne kontakty nauczyciela z uczniami niosą ogromne korzyści, służąc m.in. dzieciom zagrożonym niepowodzeniem w szkole (Burgess i Ladd, 2001). Istnienie konfliktu bądź nieporozumienia pomiędzy prowadzącym zajęcia a jego podopiecznymi może zwiększać ryzyko porażki (Burgess i Ladd, 2001). Z kolei dzieci, które są zmotywowane do wchodzenia w kontakt z rówieśnikami i dobrze się czują w tych relacjach, szybciej rozwijają się zarówno w społecznym, jak i akademickim obszarze (Hamre i Pianta, 2001; Ladd, Birch i Buhs, 1999; Pianta, Steinberg i Rollins, 1995).

Analiza procesów zachodzących w klasie niesie ze sobą ogromną ilość informacji pomocnych w konstrukcji programów kierowanych do nauczycieli czy pedagogów, którzy poprzez tworzenie konstruktywnych relacji z uczniami mogą przyczynić się do podniesienia efektywności kształcenia uczniów. Istnieje wiele ilościowych metod badawczych, takich jak testy czy ankiety, które mimo użyteczności i łatwości zastosowania nie zawsze pozwalają na precyzyjną analizę klimatu klasy. Uzyskanie szerszego obrazu badanego obszaru staje się możliwe dzięki połączeniu ilościowego pomiaru zjawisk wraz z metodami jakościowymi, skupiającymi się na treści przekazywanych komunikatów.

Metoda obserwacyjna jest bardzo często wykorzystywana w analizach klimatu klasy. Po pierwsze dlatego, że zapewnia badaczom bezpośredni, niedeklaratywny wgląd w jakość komunikacji pomiędzy uczestnikami lekcji. Po drugie różnorodność uzyskanych danych pozwala na opis czynników warunkujących proces uczenia w odniesieniu do metod i narzędzi wykorzystywanych przez nauczyciela oraz klimatu klasy tworzonego przez samych rówieśników. Generalny zarzut względem metody obserwacyjnej jest związany z efektem obserwatora, czyli jego wpływem na zmianę zachowania uczestników obserwacji. Im mniejszy wpływ obserwatora na obserwowanych, tym bardziej rzetelne wyniki. Z kolei obserwacja jakościowa skoncentrowana na pogłębionej analizie małej próby zachowań często nie dostarcza reprezentatywnych danych.

Celem konstrukcji aplikacji komputerowej Argos wykorzystywanej w badaniach podłużnych szkolnych uwarunkowań efektywności kształcenia było przygotowanie narzędzia, które nie będzie wymagało obecności kamery na lekcjach w celu zakodowania danych i w ten sposób przyczyni się do redukcji efektu obserwatora, a także pozwoli na ilościowe zgromadzenie danych na dużej próbie (70 oddziałów klasowych) i porównanie ich z wynikami innych narzędzi, weryfikujących wpływ pozostałych czynników na efektywność kształcenia. Ponadto system obserwacyjny Argos, dzięki automatycznej rejestracji zdarzeń (zachowań nauczyciela lub ucznia), pozwala na pełne zautomatyzowanie metodologii próbek czasowych i w przeciwieństwie do tradycyjnych metod badań (typu papier-ołówek) gwarantuje bardziej rzetelny i rozbudowany opis elementów warunkujących efektywność kształcenia.

Fundamentem koncepcji merytorycznej, wykorzystywanej w niniejszym badaniu, były dwa systemy służące obserwacji zajęć szkolnych. Systemy CLASS (*The Classroom Assessment Scoring System*) i inCLASS (*Individualized Classroom Assessment Scoring System*) stworzył zespół specjalistów z Uniwersytetu Virginia, badających proces nauczania pod kierunkiem profesora Roberta C. Pianty. Pierwszy z dwóch wystandaryzowanych narzędzi obserwacyjnych – system CLASS – wspomaga ocenę efektywności pracy nauczyciela w przedszkolach i szkołach podstawowych, zaś nauczycielom ułatwia poznanie użyteczności stosowanych metod nauczania (R.C. Pianta, K. La Paro, i B.K. Hamre, 2008). Narzędzie inCLASS umożliwia dokonanie oceny funkcjonowania dzieci podczas wchodzenia w interakcje z rówieśnikami oraz z otoczeniem edukacyjnym. W obu systemach wyszczególniono odmienne wymiary oraz odpowiadające im obserwowalne wskaźniki. W oryginalnej koncepcji, badacze, bazując na 15-minutowych cyklach badawczych (10 minut na

obserwację, 5 minut na dokonanie oceny), oceniali poszczególne wymiary (np. pozytywny klimat klasy) na 7-stopniowej skali.

System obserwacyjny Argos

Przygotowana w oparciu o systemy CLASS oraz inCLASS aplikacja komputerowa Argos umożliwia przeprowadzenie trzech rodzajów obserwacji:

- obserwacji skoncentrowanej na nauczycielu (przygotowanej w oparciu o metodę CLASS);
- obserwacji skoncentrowanej na uczniu (opierającej się na metodzie inCLASS);
- obserwacji dydaktycznej (pozwalającej na analizę metod i form pracy stosowanych przez nauczyciela)

W dwóch pierwszych trybach obserwacji została zastosowana metodologia próbek czasowych. W trakcie 25 (obserwacja nauczyciela) oraz 24 (obserwacja ucznia) próbek czasowych obserwator rejestruje konkretne zachowania nauczyciela lub uczniów i zaznacza je na osi czasu w jednym z okien aplikacji. Pojedyncza próbka czasowa trwa 1,5 minuty, z czego w trakcie 30 sekund obserwator obserwuje nauczyciela lub ucznia, a w kolejnych 60 sekundach zaznacza zaobserwowane zachowania na osi czasu. Dzięki kluczowej funkcjonalności aplikacji, jaką jest Licznik, proces próbkowania jest w pełni zautomatyzowany, co pozwala obserwatorom skoncentrować się wyłącznie na zdarzeniach, które występują na lekcji. Po zakończonych obserwacjach dane w postaci pliku xml są przesyłane za pośrednictwem panelu webowego, stworzonego specjalnie na potrzeby projektu.

Obserwacje w ramach badania szkolnych uwarunkowań efektywności kształcenia zostały zrealizowane w okresie od marca do czerwca 2013 roku. W pierwszej kolejności przeprowadzono badanie pilotażowe złożone z dwóch etapów, w trakcie których w dwudziestu klasach trzecich 37 obserwatorów zrealizowało odpowiednio 8 godzin obserwacji, następnie w dwudziestu klasach piątych w tych samych szkołach odpowiednio 10 godzin obserwacji (5 na języku polskim i 5 na matematyce).

W kwietniu w 69 oddziałach kohorty badania podłużnego szkolnych uwarunkowań efektywności kształcenia rozpoczęło się główne badanie obserwacyjne, w trakcie którego 43 obserwatorów przeprowadziło 15 godzin obserwacji w każdym z oddziałów (8 godzin języka polskiego oraz 7 godzin matematyki). Jednym z celów powyższego badania było sprawdzenie właściwości psychometrycznych systemu obserwacyjnego Argos, a przede wszystkim oszacowanie rzetelności skonstruowanego narzędzia. Trudność w określeniu rzetelności wykorzystanej metody obserwacyjnej wynika z różnorodności źródeł błędów, które należy uwzględnić w analizach, a które mają swoje źródło zarówno po stronie aplikacji (różne tryby obserwacyjne i zachowania w obrębie poszczególnych trybów, a także w stabilności działania samej aplikacji), obserwatorów (różnice w kompetencjach) czy samej lekcji (różne przedmioty czy pory lekcji, na których odbywały się obserwacje). Metodologia zastosowana w narzędziach CLASS i inCLASS, na bazie których powstał system obserwacyjny Argos, pozwala na wykorzystanie teorii uniwersalizacji do analizy rzetelności metody.

Uwzględnia ona kontekst sytuacji obserwacyjnej i pozwala określić tyle współczynników rzetelności, ile pytań odnośnie źródeł błędu potrafimy sformułować.

W celu określenia rzetelności systemu obserwacyjnego Argos przeprowadzono obserwację z wykorzystaniem procedury równoległego kodowania tej samej lekcji, zarówno podczas badań pilotażowych, jak i po każdym z etapów badania obserwacyjnego w trakcie specjalnych spotkań z obserwatorami.

Pierwsza procedura zakładała równoległe kodowanie trzech lekcji przez pary obserwatorów (obserwacja nauczyciela, obserwacja uczniów, obserwacja dydaktyczna) w trakcie badania pilotażowego w klasach trzecich oraz sześć godzin równoległego kodowania podczas badania obserwacyjnego w klasach piątych (3 godziny języka polskiego i 3 godziny matematyki). Otrzymaliśmy w ten sposób 60 podwójnie zakodowanych lekcji w klasach trzecich oraz 120 podwójnie zakodowanych lekcji w klasach piątych.

Dodatkowo, po każdym z trzech etapów badania obserwatorzy w trakcie spotkań fokusowych kodowali równoległe 3 lekcje (obserwacja nauczyciela, obserwacja ucznia, obserwacja dydaktyczna) wcześniej zakodowane przez ekspertów prowadzących szkolenie z obserwatorami, z których dane, ze względu na obecność wzorca, zostały wykorzystane w analizie rzetelności systemu Argos przedstawionej w niniejszym artykule.

Teoria detekcji sygnałów

Teoria detekcji sygnałów (TDS), która stanowiła inspirację dla opracowania koncepcji analizy rzetelności systemu obserwacyjnego Argos, wprowadziła w latach 60. rozróżnienie pomiędzy obserwatorem jako odbiorcą wrażeń zmysłowych a obserwatorem podejmującym decyzje. TDS koncentruje się na wykrywaniu słabych sygnałów na tle szumu i ma zastosowanie w każdej sytuacji, w której dane wejściowe są niejednoznaczne. Mimo iż teoria detekcji sygnałów wywodzi się z elektroniki i statystycznej teorii decyzji, Tanner i Swets (1954) przyczynili się do jej rozwoju, wykorzystując ją do analizy procesów percepcyjnych zachodzących w trakcie reagowania odbiorcy na różnego rodzaju bodźce. Dzięki stworzeniu wspólnych ram dla procesów percepcyjnych i decyzyjnych odbiorcy teoria detekcji sygnałów pozwala odpowiedzieć na pytanie, czy dana osoba jest czulszym detektorem, czy też z różnych względów preferuje konkretne odpowiedzi.

Głównym zastosowaniem teorii detekcji sygnałów jest określenie najlepszej reguły decyzyjnej, którą można przyjąć w konkretnym przypadku, jakie czynniki należy wziąć pod uwagę przy jej podejmowaniu. Zadaniem odbiorcy bodźców jest dokonanie wyboru pomiędzy opcjami: „tak, jest sygnał” i „nie, to tylko szum”. To, co jest najbardziej istotne, to uwzględnianie przez teorię detekcji sygnałów błędów, które mogą zostać popełnione w trakcie podejmowania decyzji. Mianowicie odbiorca może zaprzeczyć pojawieniu się konkretnego sygnału w sytuacji, w której realnie on wystąpił. Taki błąd jest określany mianem chybienia (*false negative*). Ponadto odbiorca może błędnie wskazać sygnał w sytuacji, w której się on nie pojawił. Taki błąd określany jest fałszywym alarmem (*false positive*).

Płaszczyzna *Receiver Operating Characteristic* (ROC) jest narzędziem, które wywodzi się z teorii detekcji sygnałów i służy do oceny poprawności klasyfikatora, na podstawie którego podejmuje się decyzje. Ilustrując związek pomiędzy czułością, czyli zdolnością odbiorcy do wykrycia sygnałów, a specyficznością klasyfikatora, czyli zdolnością odbiorcy do unikania błędnych decyzji, wspomaga proces decyzyjny ukierunkowany na wykrywanie konkretnych sygnałów. W sytuacji, gdy obiekt, który podlega analizie przez odbiorcę, opisywany jest przy pomocy zmiennej ciągłej, klasyfikacja obiektu do jednej z dwóch grup (odpowiadającym wystąpieniu sygnału lub jego braku) może przebiegać przy różnych wartościach progu, który oddziela dwie grupy od siebie. W takim przypadku możliwe jest wyrysowanie na płaszczyźnie ROC krzywej, która dostarcza informacji o efektywności klasyfikatora przy różnych wartościach progu.

Analiza rzetelności systemu obserwacyjnego Argos w ramach TDS

Pierwszym krokiem pozwalającym na przyłożenie teorii detekcji sygnałów do systemu obserwacyjnego Argos jest zdefiniowanie podstawowych pojęć obecnych w tej teorii w kontekście procesu obserwacyjnego – tzn. sygnału i odbiorcy. Przez sygnał możemy rozumieć pojedyncze zachowanie dowolnej kategorii w danej próbie czasu. Odbiorcą zaś byłby obserwator, który za pomocą aplikacji Argos odnotowuje w danej próbie czasu wszystkie zaobserwowane przez siebie zachowania (wchodzące w skład listy obserwowanych zachowań). Przy tak zdefiniowanych pojęciach, widzimy wyraźnie, że obserwator dostarcza nam tylko informacji o tym, czy w danej próbie czasu wystąpiło konkretne zachowanie. Mamy więc do dyspozycji tylko pozytywne zaklasyfikowanie sygnału. Obserwator nie określa bowiem wprost, że sygnał w danej próbie czasu się *nie* pojawił.

Do określenia, czy obserwator poprawnie zaklasyfikował sygnał, potrzebne jest odniesienie do wzorca, który dawałby nam informacje o faktycznych zdarzeniach, które w danej próbie czasu zaszły. Podczas spotkań fokusowych po pilotażu opisanym wyżej wszyscy obserwatorzy mieli za zadanie zakodować w aplikacji obserwację tej samej lekcji. Materiał filmowy z lekcji został również zakodowany przez ekspertów IBE. Zakodowanie ekspertów jest traktowane na potrzeby niniejszego artykułu jako wzorcowa obserwacja – stanowi ono najlepszy zapis tego, co wydarzyło się na lekcji w kontekście obserwowanych za pomocą systemu Argos kategorii zdarzeń.

Skonfrontowanie obserwacji dokonanych przez konkretnego obserwatora ze wzorcem pozwala nam na skonstruowanie macierzy klasyfikacji 2x2, znanej z teorii detekcji sygnałów:

Prawdziwie pozytywne wskazanie. Występuje wtedy, gdy we wzorcu mamy informację, że dane zachowanie wystąpiło, a obserwator je w programie zaznaczył.

Fałszywie negatywne wskazanie. Występuje wtedy, gdy we wzorcu mamy informację, że dane zachowanie wystąpiło, jednak obserwator go nie zaznaczył.

Fałszywie pozytywne wskazanie. Występuje wtedy, gdy według wzorca dane zachowanie nie miało miejsca, a obserwator zaznaczył, że ono wystąpiło.

Prawdziwie negatywne wskazanie. Występuje wtedy, gdy według wzorca dane zachowanie nie miało miejsca, a obserwator nie zaznaczył w aplikacji, że ono wystąpiło.

Tabela 1. Tabela krzyżowa z zestawieniem kategorii możliwych klasyfikacji sygnału w próbce

		Wzorec	
		Zachowanie wystąpiło	Zachowanie nie wystąpiło
Obserwator	Zachowanie wystąpiło	Prawdziwie pozytywne wskazanie	Fałszywie pozytywne wskazanie
	Zachowanie nie wystąpiło	Fałszywie negatywne wskazanie	Prawdziwie negatywne wskazanie

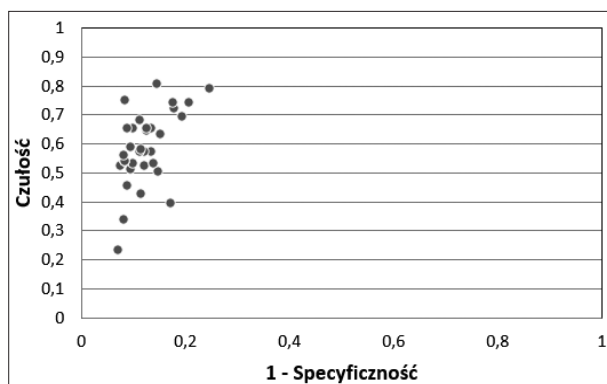
Pierwsze trzy z wymienionych kategorii nie budzą wątpliwości interpretacyjnych. Jednakże ostatnia grupa klasyfikacji sygnałów jest problematyczna. Jak powiedzieliśmy wcześniej, obserwator nie zaznacza w programie informacji o tym, że dane zachowanie nie wystąpiło. To samo zastrzeżenie dotyczy przyjętego przez nas wzorca. W związku z tym sytuacja, w której zarówno według wzorca, jak i obserwatora dane zachowanie nie miało miejsca, stanowi swoiste wyzwanie teoretyczne. Problem ten sprowadza się do tego, że obserwator ma za zadanie stwierdzić dla każdej próbki czasu nie tylko to, czy dane zachowanie wystąpiło, ale także zliczyć jego wystąpienia. Stanem przeciwnym dla wystąpienia sygnału jest brak jego wystąpienia. Co jednak możemy określić jako stan przeciwny dla pewnej liczby wystąpień konkretnego zachowania w próbce?

Dla powodzenia próby zastosowania teorii detekcji sygnałów do analizy danych obserwacyjnych systemu Argos, potrzebne było określenie maksymalnej liczby zdarzeń każdego typu, które mogły wystąpić w danej próbce, i dołączenie ich jako negatywnych wskazań do wzorca. Należy przy tym zauważyć, że im większą liczbę zachowań dołączymy do kategorii „zachowania, które według wzorca nie wystąpiły”, tym większy otrzymamy odsetek dla kategorii prawdziwie negatywnych wskazań obserwatora. Rozwiązaniem, które wydaje się najmniej arbitralne, jest określenie tego maksimum na podstawie fałszywie pozytywnych wskazań obserwatorów, tzn. tych zachowań, które według obserwatorów wystąpiły w danej próbce, ale nie wystąpiły we wzorcu. Uznajemy tym samym, że o maksymalnej liczbie wystąpień danego zachowania w danej próbce, najlepiej wnioskować wyłącznie z zaobserwowanych w tej próbce przez obserwatorów zachowań. Na takie rozwiązane zdecydowaliśmy się na potrzeby niniejszego artykułu.

Przy wyżej opisanym ujęciu, dla każdej próbki tworzymy zestawienie poszczególnych zliczeń zachowań zaobserwowanych w danej próbce przez wszystkich obserwatorów lub obecnych dla danej próbki we wzorcu, a następnie dla każdego obserwatora odnosimy wskazania dla tak określonego zbioru zachowań w danej próbce do wskazań obecnych we wzorcu. W ten sposób, dla każdego obserwatora w ramach danej próbki wypełniamy macierz zaklasyfikowań i sumujemy je po próbkach. W wyniku tej operacji dysponujemy dla każdego

obserwatora informacjami umożliwiającymi wyliczenie, opisanych wcześniej, dwóch kluczowych w teorii detekcji sygnałów współczynników: czułości i specyficzności, a także możemy każdego obserwatora umieścić na płaszczyźnie ROC. Analiza zależności pomiędzy czułością i specyficznością może dostarczyć nam bardzo dużo informacji nie tylko o obserwatorach, ale także o samej aplikacji Argos i obserwowanych kategoriach zachowań.

Poniżej (rys. 1) zaprezentowano wykres płaszczyzny ROC z naniesionymi wynikami zestawienia czułości i specyficzności dla wszystkich obserwatorów, którzy brali udział w fokusie po II etapie pilotażu narzędzia. Wykres przedstawia dane zebrane podczas trybu obserwacji nauczyciela, na jednej lekcji języka polskiego w klasie V SP.



Rysunek 1. Wykres z umiejscowieniem na płaszczyźnie ROC obserwatorów dla jednej lekcji z języka polskiego w klasie V. Tryb obserwacji nauczyciela

Patrząc na wykres, można zauważyć, że obserwatorzy są znacznie bardziej zróżnicowani pod względem czułości niż specyficzności. Czułość obserwatorów waha się bowiem od lekko ponad 0,2 do ponad 0,8, specyficzność z kolei skupia się w okolicach wartości 0,8-0,9. Niskie zróżnicowanie specyficzności nie powinno dziwić – wynika ono bezpośrednio z przyjętego sposobu wyliczenia prawdziwie negatywnych wskazań obserwatorów opisanego powyżej. Zastanówmy się, jak możemy interpretować te wskaźniki w kontekście systemu obserwacyjnego Argos.

Czułość mówi nam o odsetku wskazań zgodnych ze wzorcem spośród wszystkich zachowań obecnych we wzorcowej obserwacji. Gdy czułość obserwatora zbliży się do jedności, możemy powiedzieć, że obserwator zaznaczył co najmniej podobną liczbę wystąpień poszczególnych kategorii zachowań do eksperci we wzorcu. Jak widzimy na wykresie, większość obserwatorów zaznaczyła więcej niż połowę zachowań znajdujących się we wzorcowej obserwacji. Niska czułość wybranych obserwatorów niepokoi – może bowiem świadczyć o niskim stopniu przyswojenia założeń systemu obserwacyjnego. Informacja o wysokiej czułości nie jest jednak wystarczająca do pozytywnej oceny jakości pracy obserwatora. Czułość bowiem bardzo łatwo zmaksymalizować, zaznaczając liczne wystąpienia dużej liczby kategorii w próbce. W efekcie może się zdarzyć, że obserwator z dużą czułością to obserwator, który zbyt łatwo ulega wrażeniu, że jakieś zachowanie wystąpiło w próbce.

Specyficzność pozwala sprawdzić tę hipotezę. Miara ta mówi nam o tym, na ile obserwator powstrzymuje się od raportowania „fałszywych alarmów”, tj. zachowań, które nie są obecne we wzorcu. Przekształcenie tego wskaźnika, poprzez odjęcie go od jedności ($1 - \text{specyficzność}$), które widzimy na wykresie, umożliwia odpowiedź na pytanie o to, za jaki odsetek błędnych wskazań zachowań (tj. zachowań niewystępujących we wzorcu) dla wszystkich próbek odpowiada dany obserwator. Należy przy tym pamiętać, że błędne wskazanie wystąpienia danego zachowania w próbce mogło być udziałem kilku obserwatorów. Miara ta nie mówi nam więc o unikalnych dla danego obserwatora błędach.

Z analizy wykresu wynika, że im niższa czułość obserwatora, tym większa specyficzność. Prowadzi to do wniosku, że część obserwatorów była bardzo „wstrzemięźliwa” we wskazywaniu, że dane zachowanie miało miejsce w próbce. Można postawić hipotezę, że część z nich kierowała się restrykcyjnymi zasadami decyzyjnymi przy określaniu, czy to, co zadziało się w próbce, wypełnia znamiona poszczególnych kategorii zachowań.

Oczywiście im bliżej maksymalnych wartości tych dwóch wskaźników lokują się obserwatorzy, tym lepiej z punktu widzenia rzetelności systemu obserwacyjnego. Na wykresie zaprezentowano zestawienie dla wszystkich zaobserwowanych na lekcji kategorii zachowań. Ułożenie obserwatorów może się jednak zmienić w zależności od tego, które spośród tych kategorii weźmiemy pod uwagę. Może być bowiem tak, że niektóre kategorie sprawiają obserwatorom problem w związku z tym, że nie są odpowiednio precyzyjnie opisane. Przeprowadzenie analiz na podzbiorze kategorii może nam dużo powiedzieć o ich wpływie na jakość obserwacji prowadzonych przez obserwatorów w ramach danego trybu obserwacji. Hipotezy, które wyłonią się w ramach takich analiz, należy następnie zweryfikować, przyglądając się zależności między czułością i specyficznością dla obserwatorów w ramach pozostałych obserwacji ze spotkań fokusowych.

Podsumowanie

Przedstawiona w niniejszym artykule analiza stanowi pierwszą próbę przyjrzenia się problemowi rzetelności systemu obserwacyjnego Argos. Do jej przeprowadzenia wykorzystano pojęcia zaczerpnięte z teorii detekcji sygnałów, a w szczególności zestawienie czułości i specyficzności znane z analizy ROC. Analiza obserwacji z równoległego kodowania jednej lekcji języka polskiego przez 34 obserwatorów w trybie obserwacji nauczyciela umożliwiła zobrazowanie zróżnicowania zgodności wskazań obserwatorów z obserwacją wzorcową przeprowadzoną przez ekspertów IBE. Z pewnością analiza danych z jednej lekcji nie pozwala na orzekanie o rzetelności systemu obserwacyjnego Argos. Umożliwia jednak postawienie bardziej szczegółowych pytań związanych z jego własnościami.

Na wielkość czułości i specyficzności obserwatorów w ramach konkretnej obserwacji ma bowiem wpływ wiele czynników, które można przypisać trzem obszarom: samemu systemowi obserwacyjnemu Argos, obserwatorom oraz obserwowanej lekcji. Do kwestii związanych z systemem obserwacyjnym możemy zaliczyć adekwatność i precyzyjność obserwowanych kategorii zachowań, a także stabilność działania aplikacji komputerowej. Z kolei obserwatorzy mogą różnić się doświadczeniem i treningiem w obsłudze programu oraz stopniem

opanowania obserwowanych kategorii. Wśród kwestii związanych z obserwowaną lekcją możemy umieścić wszelkie przygodne cechy sytuacji, jaką jest obserwacja zdarzeń w konkretnej klasie, np. reakcję nauczyciela i uczniów na fakt bycia obserwowanymi, a także zdarzenia, które mają miejsce na lekcji. Sprawdzenie wpływu tych obszarów na rzetelność Argosa (a dokładniej danego trybu obserwacji) jest możliwe poprzez analizę stabilności oszacowań czułości i specyficzności obserwatorów w ramach zróżnicowanych podzbiorów obserwowanych kategorii zachowań. Drugim krokiem jest przeprowadzeniem analogicznych analiz dla wszystkich lekcji obserwowanych w ramach spotkań fokusowych.

Uzupełnieniem analiz zaprezentowanych w tym artykule może być wspomniana we wstępie analiza rzetelności danych obserwacyjnych bazująca na teorii uniwersalizacji (*generalizability theory*), która została wykorzystana przy oszacowaniu rzetelności systemów CLASS i inCLASS. Opierając się na danych obserwacyjnych z równoległego kodowania, jesteśmy w stanie uzupełnić analizy bazujące na płaszczyźnie ROC o dodatkowe wskaźniki wyliczone w ramach tej teorii.

Bibliografia

1. Birch, S.H., Buhs, E.S. Ladd. (1999). Children's social and scholastic lives in kindergarten: related spheres of influence? *Child Development*, 1, 373-400.
2. Coombs, C.H. Coombs, Dawes R.M., Tversky A. (1977) Wprowadzenie do psychologii matematycznej. Państwowe Wydawnictwo Naukowe, Warszawa.
3. Hamre, B., Pianta, R. (2001). Early teacher-child relationships and the trajectory of children's school outcomes through eighth grade. *Child Development*, 72, 625-638.
4. Ladd, G. W., Burgess, K. B. (2001). Do relational risks and protective factors moderate the linkages between childhood aggression and early psychological and school adjustment? *Child Development*, 72, 1579-1601.
5. Pianta, R. C., La Paro, K. M., Hamre, B. K. (2008). *Classroom Assessment Scoring System (CLASS): Manual K-3*. Baltimore, MD: Paul H. Brookes.
6. Pianta, R.C., Steinberg, M.S., Rollins, K.B. (1995). The first two years of school: Teacher-child relationships and deflections in children's classroom adjustment. *Development and Psychopathology*, 7, 297-312.
7. Tanner, W.P., Swets, J.A. (1954). A decision-making theory of visual detection. *Psychological Review*, 61, 401-9.
8. Vygotsky, L. (1978). Interaction between learning and development. *Mind and society*. Cambridge, MA: Harvard University Press, 79-91.