

Karolina Świst
Paulina Skórska
dr Artur Pokropek
Filip Kulon

Instytut Badań Edukacyjnych w Warszawie

Pomiar umiejętności rozwiązywania problemów złożonych – znaczenie oraz możliwości i sposoby wykorzystania w badaniach edukacyjnych

Wprowadzenie

Wśród umiejętności niezbędnych dla funkcjonowania współczesnej gospodarki wymienia się rozwiązywanie złożonych problemów, myślenie krytyczne, zarządzanie swoim zachowaniem, umiejętności z zakresu technologii informacyjno-komunikacyjnych (TIK) oraz pracę w zespole i skuteczną komunikację (Autor, Levy i Murnane, 2003; Griffin, Care i Mcgaw, 2011). Szczególny nacisk kładzie się na umiejętność rozwiązywania problemów złożonych (*complex problem solving*). Problemy te cechują się wzajemnymi powiązaniem pomiędzy zmiennymi oraz odzwierciedlają dynamiczny i niejednoznaczny charakter sytuacji, w których funkcjonuje jednostka. Ich rozwiązanie wymaga zrealizowania wielu celów jednocześnie (często wzajemnie ze sobą sprzecznych), a wynik poszczególnych działań jest odroczone w czasie (Funke, 1991; 2001). Sprawdzanie umiejętności rozwiązywania problemów złożonych stało się ważną częścią międzynarodowych badań w obszarze edukacji takich jak PISA (*Programme for International Student Assessment*) czy PIAAC (*Programme for the International Assessment of Adult Competencies*). Uwzględnienie tego zagadnienia wymagało zmian w stosowanych metodach pomiaru edukacyjnego.

Do trafnego wnioskowania na temat poziomu umiejętności rozwiązywania problemów złożonych nie wystarcza wyłącznie pomiar pojedynczej cechy ukrytej (najczęściej operacjonalizowanej jako zmienna ciągła). Kluczowe staje się pomiar interakcji struktur wiedzy oraz strategii (procesów) wykorzystywanych w celu rozwiązywania problemów (Masters i Mislavy, 1993). Ten typ pomiaru wymaga zaawansowanych technologii komputerowych umożliwiających dokładną rejestrację zachowań uczniów w czasie rozwiązywania problemów w postaci np. dzienników aktywności (*activity logs*) (Hmelo-Silver, Chinn, Chan i O'Donnell, 2013). Wyprowadzenie trafnych wniosków na podstawie tak zebranych danych wymaga zastosowania skomplikowanych technik statystycznej analizy danych. Konieczna staje się integracja psychometrii i metod wykorzystywanych w innych dziedzinach nauki, np. uczenia się maszynowego (*machine learning*) czy drażenia danych (*data mining*).

Celem niniejszego opracowania jest przybliżenie Czytelnikowi zagadnienia rozwiązywania problemów złożonych w ujęciu psychologicznym, sposobów jego pomiaru (także w badaniu PISA), modelowania statystycznego tego typu danych oraz kontrowersji i obszarów dotychczas niezbadanych.

Definicja problemów złożonych

Badanie procesów zachodzących w ludzkim umyśle podczas rozwiązywania problemów złożonych jest domeną psychologii poznawczej. Jedną z najpopularniejszych teorii opisujących procesy dokonywane przez ludzki umysł podczas rozwiązywania problemów jest teoria Newella i Simona (1972). Zakłada się w niej, że problem można zdefiniować jako układ stanów wiedzy, opisany poprzez cztery charakterystyki: (1) reprezentację stanu początkowego (punkt wyjścia), (2) reprezentację celu, który chcemy osiągnąć (punkt dojścia), (3) reguły przekształcenia stanu wiedzy (możliwe do zastosowania operacje przybliżające nas do osiągnięcia celu) oraz (4) restrykcje (przekształcenia niedopuszczalne podczas rozwiązywania problemu). Newell i Simon (1972) zakładają też, że u podstaw rozwiązywania problemu leży myślenie sekwencyjne. Z tego założenia wynika, że procesy poznawcze występujące podczas rozwiązywania problemów można zidentyfikować np. poprzez protokoły głośnego myślenia.

Według Sternberga i Williamsa (2002) proces rozwiązywania problemów można podzielić na szereg etapów, wśród których wyróżniają oni: (1) identyfikację problemu, (2) stworzenie jego definicji i reprezentacji, (3) opracowanie strategii rozwiązywania problemu, (4) organizowanie dostępnych informacji, (5) alokację zasobów mentalnych niezbędnych do rozwiązania problemu, (6) monitorowanie procesu rozwiązywania problemu i (7) ocenę uzyskanych rezultatów.

Badania nad umiejętnością rozwiązywania problemów przeprowadzono zazwyczaj w laboratorium przy pomocy takich klasycznych scenariuszy, jak np. problem 9 kropek (połączenie 9 kropek z wykorzystaniem 4 lub mniej prostych linii, bez unoszenia ołówka z papieru i przeprowadzania tej samej linii więcej niż raz), problem świecy i pudełka (jak przykleić zapaloną świecę do ściany, mając do dyspozycji takie elementy jak pinezki i pudełko zapałek) czy problem rentgena (czyli jak promieniowanie może zostać wykorzystane do zniszczenia guza, bez jednoczesnego uszkodzenia zdrowych tkanek) (Duncker, 1945). Takie problemy można było rozwiązać relatywnie szybko oraz istniały dla nich optymalne rozwiązania. Zakładano, że tego typu problemy przejawiały podobne charakterystyki oraz uruchamiały zbliżone procesy poznawcze co złożone dylematy, z którymi ludzie zmagają się na co dzień. W pewnym momencie zakwestionowano jednak trafność ekologiczną tych badań, czyli możliwość bezpośredniego uogólnienia wyników laboratoryjnych na sytuację rozwiązywania złożonych problemów w prawdziwym życiu.

Badania nad bliższym rzeczywistości rozwiązywaniem problemów złożonych prowadzono na przykład w kontekście podejmowania decyzji: politycznych (Voss, Wolfe, Lawrence i Engle, 1991), biznesowych (Wagner, 1991) czy sądowych (Amsel, Langer i Loutzenhiser, 1991). Inne badania obejmowały też tematykę nabywania umiejętności eksperckich podczas rozwiązywania problemów (Anderson, Boyle i Reiser, 1985), rozróżnienia pomiędzy świadomymi

i nieświadomymi procesami poznawczymi związanymi z rozwiązywaniem złożonych problemów (Broadbent, 1977) oraz interakcji pomiędzy poznawczymi, motywacyjnymi i społecznymi czynnikami wpływającymi na umiejętność rozwiązywania problemów złożonych (Dörner, 1983).

Nowsze koncepcje zwracają uwagę zarówno na właściwości rozwiązywanego problemu, jak i na charakterystyki rozwiązującego. Frensch i Funke (1995) przedstawiają wieloaspektową teorię rozwiązywania problemów złożonych. Zgodnie z tą teorią podkreśla się interakcję pomiędzy jednostką a problemem oraz pokonywanie barier pomiędzy danym stanem wyjściowym i stanem docelowym poprzez odpowiednie zachowania i procesy poznawcze. Dany stan wyjściowy, stan docelowy oraz przeszkody utrudniające osiągnięcie stanu docelowego są złożone, nieprzejrzyste oraz ulegają dynamicznym zmianom. Dokładne dane na temat stanu obecnego, stanu docelowego oraz przeszkód są nieznanymi dla rozwiązywanego problemu. Rozwiązywanie problemów złożonych zakłada więc skuteczne i efektywne spełnienie przez rozwiązującego wymagań niezbędnych do rozwiązania problemu. Wymaga to od rozwiązującego zaangażowania wiedzy, zdolności poznawczych, emocjonalnych, osobistych i społecznych.

Dostrzeżenie znaczenia wielu czynników charakteryzujących osobę rozwiązującą problem, jak i naturę samego problemu, skutkowało próbami zidentyfikowania korelatów umiejętności skutecznego rozwiązywania problemów złożonych. Wydawało się, że główną cechą, która może mieć związek ze skutecznością rozwiązywania problemów złożonych, jest poziom inteligencji, jednak wyniki dotychczasowych badań są sprzeczne. Według niektórych badaczy (Gonzalez, Thomas i Vanyukov, 2005; Danner i in., 2011) istnieje związek pomiędzy inteligencją i umiejętnością rozwiązywania problemów złożonych, według innych (Kluwe, Misiak i Haider, 1991; Beckmann i Guthke, 1995, Joslyn i Hunt, 1998; Wenke, Frensch i Funke, 2005) takiego związku nie ma. Sprzeczne wyniki mogą być spowodowane tym, że miary inteligencji mogą być mniej trafne niż wskaźniki skuteczności rozwiązywania problemów złożonych. Przede wszystkim inteligencję definiuje się w zdecydowanie szerszy sposób, co utrudnia predykcję międzyosobniczych różnic w umiejętności rozwiązywania problemów złożonych. Co więcej, tradycyjne testy inteligencji koncentrują się zazwyczaj na szybkości i skuteczności przetwarzania symboli (czyli na umiejętności rozumowania płynnego), jak również na pojemności pamięci roboczej. Oznacza to, że tego typu testy koncentrują się na wynikach, nie zaś na procesie dochodzenia do danego rozwiązania czy zastosowanej strategii, co jest kluczowe w przypadku rozwiązywania problemów złożonych. Dodatkowo może istnieć szereg moderatorów relacji pomiędzy inteligencją a umiejętnością rozwiązywania problemów złożonych. Wyniki mogą być modyfikowane przez takie czynniki jak: przejrzystość i kontekst zadania czy zgodność warunków testowania inteligencji i umiejętności rozwiązywania problemów złożonych. Rozwiązywanie problemów złożonych zależy także od czynników zewnętrznych, czyli doświadczenia w danej dziedzinie (Krems, 1995), umiejętności poznawczych (np. wiedzy, stylu poznawczego, strategii monitorowania i ewaluacji, a także ogólnego poziomu inteligencji; Beckmann i Guthke, 1995; Berry i Broadbent, 1995) oraz pewnych umiejętności pozapoznawczych (pewności siebie, motywacji, wytrwałości, cech osobowości; Dörner i Wearing, 1995).

Na skuteczność rozwiązywania problemów złożonych wpływają też czynniki zewnętrzne: struktura problemu (np. złożoność, przejrzystość; Funke, 1995), kontekst problemu (znany bądź nieznan rozwiązującemu; Huber, 1995) oraz inne czynniki środowiskowe (informacje zwrotne, oczekiwania, współpraca; Brehmer, 1995).

Pomiar umiejętności rozwiązywania problemów złożonych przez uczniów

Podczas pomiaru umiejętności rozwiązywania problemów złożonych nie liczy się wyłącznie efekt działania, czyli wynik rozwiązania danego zadania, którego często nie można jednoznacznie sklasyfikować jako poprawną lub niepoprawną odpowiedź. Za kluczowy uznać można wybór strategii rozwiązywania problemu oraz kroki podejmowane przez ucznia w czasie jego rozwiązywania, uwzględniające charakterystykę zagadnienia i kontekst, w którym odbywa się jego rozwiązywanie. Dlatego też do modelowania umiejętności rozwiązywania problemów złożonych potrzeba znacznie więcej danych niż w przypadku modelowania umiejętności mierzonych przez klasyczną psychometrię. Do rozwoju metodologii pomiaru umiejętności rozwiązywania problemów złożonych niezbędny był więc rozwój technologiczny, umożliwiający radzenie sobie z dużymi zbiorami danych.

Dzięki komputerom możliwe jest zarejestrowanie wszystkich zachowań ucznia podczas rozwiązywania problemu w postaci dzienników aktywności (Shute, 2011; Mislevy i in., 2014). Do analizy tak złożonych danych kluczowe jest przeprowadzenie dokładnej konceptualizacji i operacjonalizacji ukierunkowujących pomiar (Williamson, Mislevy i Bejar, 2006). Jednym z najdokładniej opracowanych sposobów pomiaru umiejętności rozwiązywania problemów złożonych jest schemat oceny oparty na dowodach (*evidence-centered assessment design* – ECD; Almond, Steinberg i Mislevy, 2002; Mislevy, Steinberg i Almond, 2003; Mislevy, Almond i Lukas, 2003) opracowany przez ETS (*Educational Testing Service*). Schemat ten opiera się na koncepcji pomiaru Messicka (1990; 1994; 2000), która zakłada, że należy rozpocząć od konceptualizacji, czyli zadania pytania, jaka złożona wiedza, umiejętności czy inne atrybuty powinny zostać ocenione. Następnie powinno się określić, jakie zachowania powinny ujawniać istnienie badanej umiejętności i jakie zadania powinny wywołać następujące zachowania. Dlatego też dokładna konceptualizacja badanej umiejętności powinna poprzedzać konstrukcję i selekcję adekwatnych zadań, a także sformułowanie kryteriów ich oceny. Choć są to tradycyjne elementy procesu badawczego, są one szczególnie ważne podczas operacjonalizacji umiejętności rozwiązywania problemów złożonych ze względu na ich dynamikę i wieloaspektowy charakter. Wybór zachowań ucznia rejestrowanych podczas pomiaru musi być głęboko uzasadniony przez badacza.

W schemacie ECD mieści się zarówno model koncepcji pomiaru (*Conceptual Assessment Framework*, CAF), jak i architektura służąca dostarczaniu dowodów, składająca się z 4 procesów: (1) selekcji zadania (*activity selection process*), (2) prezentacji zadania w danej formie (*presentation process*), (3) przetwarzania odpowiedzi, której dostarcza badany (*response process*), oraz (4) podsumowania

punktowania (*summary scoring process*). W CAF mieszczą się trzy modele: ucznia/badanego (*student*), dowodów (*evidence*) i zadania (*task*). W modelu ucznia/badanego analizuje się zmienne określające poziom umiejętności jednostki; w modelu zadania – zmienne charakteryzujące właściwości zadania. Model dowodów łączy ze sobą zmienne mierzone na obydwu poziomach. Składa się on z dwóch części: (1) ewaluacyjnej, która wskazuje na to, jak identyfikować i oceniać kluczowe elementy zadania, oraz (2) pomiarowej, która wskazuje, jak modelować zebrane dane (Mislevy i in., 2014). Schemat ECD wydaje się być rozbudowany, jednak jego konstrukcja pozwala na wykreowanie szeregu realistycznych, złożonych i interaktywnych zadań z dokładnym schematem punktowania (Mislevy, Steinberg, Breyer, Almond i Johnson, 2002).

Wykorzystanie narzędzi pomiaru rozwiązywania złożonych problemów w badaniu PISA

OECD (2013) deklaruje, że w związku ze zmianami społecznymi oraz technologicznymi, coraz częściej będzie się wymagać wysokiego poziomu umiejętności adaptacyjnych, uczenia się przez całe życie, otwartości na doświadczenia i uczenia się na własnych błędach. W zawodach specjalistycznych i na stanowiskach kierowniczych wymagane jest posiadanie umiejętności rozwiązywania problemów złożonych. Niewielu takich pracowników opiera swoją pracę na powtarzalnych czynnościach. Dlatego w międzynarodowych, porównawczych badaniach edukacyjnych (np. PISA) uwzględnia się komponent pomiaru umiejętności rozwiązywania problemów złożonych. W badaniach tych nacisk kładzie się na zdolność i chęć jednostki do angażowania się w procesy poznawcze w celu zrozumienia i rozwiązania sytuacji problemowych, których metoda rozwiązania nie jest oczywista (OECD, 2013).

W badaniu PISA procesy kognitywne zaangażowane podczas rozwiązywania problemów podzielono na 4 grupy: (1) eksplorację i zrozumienie (czyli obserwację sytuacji problemowej, wyszukiwanie informacji, identyfikację przeszkód), (2) reprezentację i formułowanie (wykorzystanie tabel, wykresów i symboli w celu modelowania aspektów sytuacji problemowej, formułowania hipotez o związkach między nimi i budowanie reprezentacji problemu), (3) planowanie i wykonanie (formułowanie planu/strategii rozwiązywania problemu oraz wprowadzenie jej w życie) oraz (4) monitorowanie i refleksja (monitorowanie postępu, reakcja na informację zwrotną). Przykładowo, zadanie mierzące umiejętność rozwiązywania problemów złożonych polega na samodzielnym uruchomieniu urządzenia (odtwarzacza MP3) bez korzystania z instrukcji. Uczeń może zmieniać typ muzyki, głośność urządzenia i poziom basów. W prezentowanym zadaniu odtwarzacz MP3 posiada cztery przyciski kontrolne oraz więcej niż 20 okien odzwierciedlających wartości przybieranych przez zmienne. Funkcje przycisków nie są stałe, lecz są uwarunkowane stanem urządzenia oraz kolejnością wciskanych przycisków (Poddiakov, 2012). Używanie terminu „rozwiązywanie problemów złożonych” (*complex problem solving*) w odniesieniu do tego typu zadań może sugerować, że pomiar odwołuje się bezpośrednio do dornerowskiej definicji problemów złożonych. Jednakże część badaczy (Funke, 2001; Novick i Bassok, 2005), wskazuje, że

procesy leżące u podłoża rozwiązywania problemów złożonych nie powinny zależeć od kontekstu danego zadania – technicznego, społecznego lub osobistego (zob. Greiff, 2012). Tymczasem jeśli uczeń ma opanowaną umiejętność obsługi odtwarzacza MP3, podczas rozwiązywania zadania nie angażuje wszystkich procesów poznawczych związanych z rozwiązywaniem problemów złożonych, lecz jedynie swoje uprzednie doświadczenie.

W badaniu PISA 2012 średni wynik rozwiązywania problemów złożonych osiągnięty przez polskich uczniów wynosił 481 (przy średniej dla krajów OECD wynoszącej 500 i odchyleniu standardowym 100) (OECD, 2013). Niepokojący może się wydawać fakt, że w Polsce było więcej niż przeciętnie w krajach OECD uczniów osiągających najłabsze wyniki i mniej uczniów o najwyższym poziomie tej umiejętności. Co więcej, uczniowie radzili sobie gorzej z zadaniami wymagającymi rozwiązywania problemów złożonych niż w matematycznej części testu. Wskazuje się (OECD, 2013), że w Polsce wynik w zadaniach wymagających rozwiązywania problemów złożonych jest związany z przynależnością do konkretnej szkoły. Może to sugerować istnienie zjawiska segregacji międzyszkolnej w tym aspekcie. Wynik polskich uczniów poniżej oczekiwań próbuje się tłumaczyć komputerową formą badania, z którą polscy uczniowie nie są dobrze zaznajomieni. Sposób przeprowadzania testu wyjaśnia więcej niż 5% zróżnicowania wyników w zakresie rozwiązywania problemów. Dodatkowo, nie odnaleziono związku pomiędzy statusem społeczno-ekonomicznym uczniów a wynikami w części dotyczącej rozwiązywania problemów złożonych (OECD, 2013).

Statystyczne modele szacowania poziomu umiejętności rozwiązywania problemów złożonych

Jak wspomiano, dane z komputerowych testów mierzących umiejętność rozwiązywania problemów złożonych są rejestrowane zazwyczaj w formie dziennika aktywności. Dzięki temu zebrane zostają informacje na temat szeregu czasowego wydarzeń oraz oceny danego działania jako poprawne lub niepoprawne (von Davier i Halpin, 2013). Choć dziennik aktywności dostarcza ważnych informacji na temat ucznia, ich ilość może utrudniać modelowanie i wnioskowanie. Innymi czynnikami utrudniającymi interpretację są wielowymiarowość mierzonych umiejętności oraz lokalna zależność danych, z którymi nie radzi sobie tradycyjna psychometria (Quellmalz, Timms i Schneider, 2009).

Von Davier i Halpin (2013) wskazują, że istnieje szereg potencjalnych metod modelowania tego typu danych, jak np. dynamiczna analiza czynnikowa, modelowanie wielopoziomowe, dynamiczne modele liniowe, modele równań różniczkowych, nieparametryczne modele eksploracyjne (takie jak analiza sieci społecznych), ukryte modele Markowa, sieci bayesowskie, metody uczenia się maszynowego, analiza klas ukrytych, sieci neuronowe. Ze względu na rozmiary niniejszego opracowania, przedstawione zostaną tylko trzy techniki oparte na: sieciach neuronowych, sieciach bayesowskich oraz ukrytych modelach Markowa. Każda z nich ma swoje zalety i wady, biorąc pod uwagę możliwości obliczeniowe oraz koncepcyjne.

Sztuczne sieci neuronowe są nieliniowymi, parametrycznymi modelami używanymi w celach regresji i klasyfikacji złożonych wzorów (MacKay, 1995). W celu wykorzystania sieci konieczne jest najpierw dokonanie procesu jej uczenia poprzez użycie próbki danych, które mają zostać poddane analizie. Wydają się one być szczególnie przydatne w wielu aspektach rozpoznawania i predykcji wzorców. Sztuczne sieci neuronowe pozwalają na szybki i ciągły pomiar umiejętności w obrębie problemów, które są źle zdefiniowane lub gdy wzory wykonania (np. strategia rozwiązywania problemów zastosowana przez badanego) są ukryte w danych. Sieci neuronowe nadają się też do modelowania nabywania umiejętności eksperckich przez badanego w miarę rozwiązywania problemów oraz do adekwatnego modelowania zależności pomiędzy umiejętnościami niezbędnymi do rozwiązania problemu (Stevens, Lopo i Wang, 1996; Hurst, Casillas i Stevens, 1997). Jednakże sztuczne sieci neuronowe opierają się na klasycznym rozumieniu prawdopodobieństwa – zakładają dokładną specyfikację wszystkich założeń modelowania. Dodatkowo, są podatne na problem przeuczenia, tj. zwiększenia błędu predykcji, gdy analizowane są nowe dane (nie z próby uczącej). Problem ten pojawia się szczególnie w przypadku skomplikowanych modeli z dużą liczbą parametrów (Lawrence, Giles i Tsoi, 1997).

Problem przeuczenia można rozwiązać poprzez wykorzystanie metod bayesowskich, umożliwiających kontrolę tego zjawiska. Modele sieci bayesowskich są graficznymi modelami (skierowanymi grafami acyklicznymi; zob. Pearl, 1988) reprezentującymi warunkową niezależność pomiędzy losowym zestawem zmiennych (węzłami). Węzły mogą reprezentować dowolne zmienne: parametr modelu lub też zmienną ukrytą (Correa, Bielza i Pamies-Teixeira, 2009). Połączenia (łuki) pomiędzy węzłami reprezentują założenia o warunkowej niezależności – determinują, jaka informacja na temat prawdopodobieństwa jest potrzebna, by wyspecyfikować rozkład prawdopodobieństwa zmiennych znajdujących się w sieci. Połączenia determinują także kierunek przyczynowości (Charniak, 1991). Modele sieci bayesowskich służą do modelowania specyficznych sytuacji – zdajemy sobie sprawę, że przyczynowość jest w nich kluczowa, jednak nasze zrozumienie relacji pomiędzy zmiennymi jest niekompletne, więc musimy opisać je w sposób probabilistyczny (Charniak, 1991). Taką sytuacją jest rozwiązywanie problemów złożonych oraz rozwiązywanie problemów złożonych we współpracy z innymi ludźmi. Jeśli obserwujemy dowody w postaci uzyskania wartości danej zmiennej, możemy dokonać aktualizacji rozkładów prawdopodobieństw brzegowych zmiennych w całej sieci. Przekładając to na sytuację pomiaru umiejętności złożonych, poprzez twierdzenie Bayesa możemy aktualizować nasze przekonania na temat poziomu umiejętności badanego, biorąc pod uwagę jego obserwowalny poziom wykonania zadania w teście (Mislevy i Gitomer, 1996). Aktualizowane przekonania mogą być podstawą selekcji następnego zadania, dostarczania informacji zwrotnych czy też raportowania wyników osobom zainteresowanym wynikiem pomiaru. W związku z tym, że sieci bayesowskie zakładają warunkową niezależność zmiennych, nadają się idealnie do modelowania rozwiązywania problemów złożonych, w których do celu można dojść różnymi drogami (Mislevy i in., 2014).

Dla modelowania rozwiązywania problemów złożonych w czasie w największym stopniu nadają się ukryte modele Markowa (*hidden Markov model*, HMM). Jest to specjalny przypadek sieci bayesowskiej dostosowanej do analizy szeregów czasowych. W modelowaniu szeregów czasowych założenie, że wydarzenie może być przyczyną późniejszego zdarzenia, jednak nie na odwrót, upraszcza schemat sieci bayesowskiej – relacje przyczynowe determinuje przede wszystkim czas (Ghahramani, 2001). Ukryty model Markowa jest narzędziem służącym reprezentacji rozkładów prawdopodobieństwa względem sekwencji obserwacji. Model ten posiada dwie kluczowe charakterystyki. Pierwszą jest założenie, że obserwacja w czasie t została wygenerowana przez proces, którego dyskretny stan S_t jest ukryty dla obserwatora (Ghahramani, 2001). Drugą charakterystyką modelu jest własność Markowa – w tym przypadku oznacza to, że modelowanie przyszłej sekwencji stanów modelu zależy wyłącznie od obecnego stanu, w którym przebywa model, nie zaś przeszłej sekwencji stanów (Anderson, 2012). Ukryte modele Markowa mogą posłużyć do wykonywania trzech podstawowych typów analiz w przypadku danych pochodzących z badań dotyczących rozwiązywania złożonych problemów podczas współpracy (Soller i Stevens, 2007). Po pierwsze, jest to estymacja modelu charakteryzującego w najlepszy sposób zestaw obserwacji. Po drugie, umożliwia wyjaśnienie sekwencji obserwacji, wydarzeń lub zachowań w terminach przynależności do klas ukrytych. Po trzecie, pozwala na przewidywanie prawdopodobieństwa wystąpienia przyszłych obserwacji, wydarzeń, zachowań. W związku z tym, że HMM umożliwia przypisanie prawdopodobieństw bycia w danym stanie bądź przejścia z jednego stanu do drugiego, umożliwia on np. modelowanie zmian strategii rozwiązywania problemów.

Podsumowanie

Modelowanie umiejętności rozwiązywania problemów złożonych wymaga od badacza odpowiednio dostosowanych narzędzi pomiarowych oraz wiedzy statystycznej wychodzącej poza klasyczną teorię testu i modelowanie w ramach teorii odpowiedzi na pozycje testowe (*item response theory*, IRT). Dodatkowo, badania nad trafnością i rzetelnością opisanych metod nadal nie są wystarczająco rozwinięte.

W kontekście późniejszej oceny trafności narzędzi nie zawsze wiadomo, jaki model rozwiązywania problemów przyjąć przy projektowaniu pomiaru umiejętności rozwiązywania problemów złożonych. Wprawdzie w wytycznych PISA (OECD, 2013) zapisane jest, że przyjęto dornerowski model rozumienia problemów złożonych, jednakże wśród badaczy nie ma zgody na temat definicji i konceptualizacji modelu rozwiązywania problemów. Część badaczy (Funke, 2001; Novick i Bassok, 2005; Greiff, 2012) uważa, że umiejętność rozwiązywania problemów złożonych nie powinna zależeć od kontekstu danego zadania – technicznego, społecznego lub osobistego. Tymczasem zadanie sprawdzające umiejętność np. obsługi odtwarzacza MP3, może angażować częściowo zdolność rozwiązywania problemów, ale też uprzednie doświadczenie. Jeśli przyjmiemy definicję, w której kontekst ma znaczenie, to można zastanowić się, czy zadania PISA w odpowiednim stopniu odzwierciedlają złożoność codziennych sytuacji życiowych. Można zadać sobie pytanie, czy odpowiednie

zaprogramowanie odtwarzacza MP3 lub zakup biletu w biletomacie różni się jakościowo od laboratoryjnego problemu wież Hanoi. Z drugiej strony należy pamiętać, że stopień złożoności problemów prezentowanych w badaniu PISA musiał być ograniczony – autorzy musieli wykreować zadanie, które mogło być eksplorowane i rozumiane przez uczniów w ograniczonym czasie oraz jednocześnie spełniało wymogi „złożonego problemu”. W zadaniach spełniono przynajmniej dwa wymogi: nieprzejrzystość zadania (niekompletny zestaw wejściowych danych, które musiały zostać skompletowane w sposób niezależny, metodami prób i błędów) oraz wiele wzajemnych powiązań (Greiff, Wüstenberg, Holt, Goldhammer i Funke, 2013).

Istnieje także problem z badaniem trafności zewnętrznej, czyli korelatów umiejętności rozwiązywania problemów złożonych – np. ciągle nie ma zgodności, czy inteligencja jest w jakikolwiek sposób z nią związana. Można także zadać pytanie na temat trafności predykcyjnej. Wskazuje się, że umiejętność rozwiązywania problemów złożonych przekłada się na sukces w miejscu pracy, jednak nie wiadomo, co miałyby być takim kryterium w szkole. Coraz ważniejszym aspektem rozwiązywania problemów złożonych jest też rozwiązywanie problemów złożonych we współpracy z innymi ludźmi (jeśli założyć, że rozwiązywane zadania mają w jak największym stopniu odzwierciedlać realne, codzienne sytuacje). Poprawne i trafne modelowanie rozwiązywania problemów złożonych w grupie wymaga jednak znacznie szerszej wiedzy dotyczącej zjawisk takich jak np. wpływ społeczny, myślenie grupowe lub innych fenomenów badanych w obrębie psychologii społecznej i mikrosocjologii. Dodatkowo, współczesne podejścia zakładają, że istnieje czynnik inteligencji danej grupy (definiowany jako uogólniona zdolność grupy do rozwiązywania różnych zadań). Nie musi być on skorelowany z inteligencją poszczególnych jej członków, lecz jego cechą jest stabilność względem poszczególnych zadań rozwiązywanych przez dany zespół. Oznacza to, że zdolność grupy do rozwiązania konkretnego typu zadań jest skorelowana ze zdolnością grupy do wykonywania zupełnie innych zadań (Woolley, Chabris, Pentland, Hashmi i Malone, 2010). Wszystkie te czynniki powinny być wzięte pod uwagę podczas modelowania statystycznego, co oczywiście czyni je coraz bardziej złożonym i trudniejszym.

Ze względu na współczesne wyzwania edukacji związane z przygotowaniem uczniów do funkcjonowania w społeczeństwie i na rynku pracy w dorosłym życiu można zakładać, że nurt badania umiejętności rozwiązywania problemów złożonych stanie się koniecznym aspektem pomiaru edukacyjnego. Wyzwaniem dla badaczy jest projektowanie trafnych narzędzi pomiaru umiejętności złożonych. Rozwój technologiczny, wykorzystanie komputerów w testowaniu uczniów i rozwój statystycznych modeli analizy danych sprawił, że badanie umiejętności rozwiązywania problemów złożonych stało się możliwe. Wciąż brakuje jednak dowodów na trafność i rzetelność wniosków płynących z tych badań. Umiejętność rozwiązywania problemów złożonych jest jedną z fundamentalnych kompetencji, stanowiącą **podstawę rozwoju gospodarczego społeczeństw**. Należy zatem spodziewać się, że będzie to jeden z dominujących kierunków poszukiwań teoretycznych i praktycznych w badaniach edukacyjnych w najbliższych dekadach.

Bibliografia

1. Almond, R., Steinberg, L. i Mislevy, R. (2002). Enhancing the design and delivery of assessment systems: A four-process architecture. *The Journal of Technology, Learning and Assessment*, 1(5), 1-64.
2. Amsel, E., Langer, R. i Loutzenhiser, L. (1991). Do lawyers reason differently from psychologists? A comparative design for studying expertise. W R. J. Sternberg i P. A. Frensch (red.), *Complex problem solving: Principles and mechanisms* (s. 223-250). Hillsdale, NJ: Lawrence Erlbaum Associates.
3. Anderson, D. F. (2012). An efficient finite difference method for parameter sensitivities of continuous time Markov chains. *SIAM Journal on Numerical Analysis*, 50(5), 2237-2258.
4. Anderson, J. R., Boyle, C. F. i Reiser, B. J. (1985). Intelligent tutoring systems. *Science*, 228, 456-462.
5. Autor, D. H., Levy, F. i Murnane, R. J. (2003). The skill content of recent technological change: An empirical exploration. *Quarterly Journal of Economics*, 118(4), 1279-1333.
6. Beckmann, J. F. i Guthke, J. (1995). Complex problem solving, intelligence, and learning ability. W P. A. Frensch i J. Funke (red.), *Complex problem solving: The European perspective* (s. 177-200). Hillsdale, NJ: Lawrence Erlbaum Associates.
7. Berry, D. C. i Broadbent, D. E. (1995). Implicit learning in the control of complex systems. W P. A. Frensch i J. Funke (red.), *Complex problem solving: The European perspective* (s. 131-150). Hillsdale, NJ: Lawrence Erlbaum Associates.
8. Brehmer, B. (1995). Feedback delays in complex dynamic decision tasks. W P. A. Frensch i J. Funke (red.), *Complex problem solving: The European perspective* (s. 103-130). Hillsdale, NJ: Lawrence Erlbaum Associates.
9. Broadbent, D. E. (1977). Levels, hierarchies, and the locus of control. *Quarterly Journal of Experimental Psychology*, 29, 181-201.
10. Charniak, E. (1991). Bayesian networks without tears. *AI magazine*, 12(4), 50-63.
11. Correa, M., Bielza, C. i Pamies-Teixeira, J. (2009). Comparison of Bayesian networks and artificial neural networks for quality detection in a machining process. *Expert Systems with Applications*, 36(3), 7270-7279.
12. Danner, D., Hagemann, D., Schankin, A., Hager, M. i Funke, J. (2011). Beyond IQ. A latent state trait analysis of general intelligence, dynamic decision making, and implicit learning. *Intelligence*, 39(5), 323-334.
13. Dörner, D. (1983). Heuristics and cognition in complex systems. W R. Groner, M. Groner i W. F. Bischof (red.), *Methods of heuristics* (s. 89-108). Hillsdale, NJ: Lawrence Erlbaum Associates.
14. Dörner, D., i Wearing, A. J. (1995). Complex problem solving: Toward a (computer simulated) theory. W P. A. Frensch i J. Funke (red.), *Complex problem solving: The European perspective* (s. 65-99). Hillsdale, NJ: Lawrence Erlbaum Associates.
15. Duncker, K. (1945). On problem solving. *Psychological Monographs*, 58(270), 1-113.
16. Frensch, P. A. i Funke, J. (1995). Definitions, traditions, and a general framework for understanding complex problem solving. W P. A. Frensch i J. Funke (red.), *Complex problem solving: The European perspective* (s. 3-25). Hillsdale, NJ: Lawrence Erlbaum Associates.
17. Funke, J. (1991). Solving complex problems: Exploration and control of complex systems. W R. J. Sternberg i P. A. Frensch (red.), *Complex problem solving: Principles and mechanisms* (s. 185-222). Hillsdale, NJ: Lawrence Erlbaum Associates.
18. Funke, J. (2001). Dynamic systems as tools for analysing human judgement. *Thinking and Reasoning*, 7(1), 69-89.

19. Funke, U. (1995). Using complex problem solving tasks in personnel selection and training. W P. A. Frensch i J. Funke (red.), *Complex problem solving: The European perspective* (s. 219-240). Hillsdale, NJ: Lawrence Erlbaum Associates.
20. Ghahramani, Z. (2001). An introduction to hidden Markov models and Bayesian networks. *International Journal of Pattern Recognition and Artificial Intelligence*, 15(1), 9-42.
21. Gonzalez C., Thomas, R. i Vanyukov, P. (2005). The relationships between cognitive ability and dynamic decision making. *Intelligence*, 33(2), 169-186.
22. Greiff, S. (2012). Assessment and theory in Complex Problem Solving. A continuing contradiction? *Journal of Educational and Developmental Psychology*, 2, 49-56.
23. Greiff, S., Wüstenberg, S., Holt, D. V., Goldhammer, F. i Funke, J. (2013). Computer-based assessment of Complex Problem Solving: concept, implementation, and application. *Educational Technology Research and Development*, 61(3), 407-421.
24. Griffin, P., Care, E. i McGaw, B. (2011). The changing role of education and schools. W P. Griffin, B. McGaw i E. Care (red.), *Assessment and teaching 21st century skills*. Dordrecht: Springer.
25. Hmelo-Silver, C., Chinn, C., Chan, C. i O'Donnell, A. (2013). *International Handbook of Collaborative Learning*, Routledge.
26. Huber. O. (1995). Complex Problem Solving as Multistage Decision Making. W P. A. Frensch i J. Funke (red.), *Complex problem solving: The European perspective* (s. 151-176). Hillsdale, NJ: Lawrence Erlbaum Associates.
27. Hurst, K., Casillas, A. i Stevens, R. H. (1997). *Exploring the dynamics of complex problem-solving with artificial neural network-based assessment systems*. Los Angeles, CA: National Center for Research on Evaluation, Standards and Student Testing (CRESST).
28. Joslyn, S. L. i Hunt, E. (1998). Evaluating individual differences in response to emergency situations. *Journal of Experimental Psychology: Applied*, 4(1), 16-43.
29. Kluwe, R. H., Misiak, C. i Haider, H. (1991). The control of complex systems and performance in intelligence tests. W H. Rowe (red.), *Intelligence: Reconceptualization and measurement* (s. 227-244). Hillsdale, NJ: Lawrence Erlbaum Associates.
30. Krems, J. F. (1995): Cognitive Flexibility and Complex Problem Solving. W: P. A. Frensch i J. Funke (red.), *Complex problem solving: The European perspective* (s. 201-218). Hillsdale, NJ: Lawrence Erlbaum Associates.
31. Lawrence, S., Giles, C. L. i Tsoi, A. C. (1997). *Lessons in neural network training: Overfitting may be harder than expected*. Proceedings of the Fourteenth National Conference on Artificial Intelligence, AAAI-97, 540-545.
32. MacKay, D. J. (1995). Probable networks and plausible predictions-a review of practical Bayesian methods for supervised neural networks. *Network: Computation in Neural Systems*, 6(3), 469-505.
33. Masters, G. N. i Mislevy, R. J. (1993). New views of student learning: Implications for educational measurement. W N. Frederiksen, R. J. Mislevy i I. I. Bejar (red.), *Test theory for a new generation of tests* (s. 219-242). Hillsdale, NJ: Lawrence Erlbaum Associates.
34. Messick, S. (1990). *Validity of Test Interpretation and Use*. Princeton, NJ: Educational Testing Service.
35. Messick, S. (1994). The interplay of evidence and consequences in the validation of performance assessments. *Educational researcher*, 23(2), 13-23.
36. Messick, S. (2000). Consequences of test interpretation and use: The fusion of validity and values in psychological assessment. W: R. D. Goffin i E. Helmes (red.), *Problems and solutions in human assessment: Honoring Douglas N. Jackson at seventy* (s. 3-20). Boston: Kluwer.

37. Mislevy, R. J., Almond, R. G. i Lukas, J. F. (2003). *A brief introduction to evidence-centered design*. ETS Research Report Series, 2003(1), 1-29.
38. Mislevy, R. J. i Gitomer, D. H. (1996). The role of probability-based inference in an intelligent tutoring system. *User-Modeling and User-Adapted Interaction*, 5, 253-282.
39. Mislevy, R. J., Oranje, A., Bauer, M. I., von Davier, A., Hao, J., Corrigan, S., Hoffman, E., DiCerbo, K. i John, M. (2014). *Psychometric considerations in game-based assessment*. Redwood City, CA: GlassLab. Pobrano z http://www.instituteofplay.org/wp-content/uploads/2014/02/GlassLab_GBA1_WhitePaperFull.pdf.
40. Mislevy, R. J., Steinberg, L. S. i Almond, R. G. (2003). Focus article: On the structure of educational assessments. *Measurement: Interdisciplinary research and perspectives*, 1(1), 3-62.
41. Mislevy, R. J., Steinberg, L. S., Breyer, F. J., Almond, R. G. i Johnson, L. (2002). Making sense of data from complex assessments. *Applied Measurement in Education*, 15(4), 363-389.
42. Newell, A. i Simon, H. A. (1972). *Human problem solving*. Englewood Cliffs, NJ: Prentice-Hall.
43. Novick, L. R. i Bassok, M. (2005). Problem solving. W K. J. Holyoak i R. G. Morrison (red.), *Cambridge handbook of thinking and reasoning* (s. 321-349). New York, NY: Cambridge University Press.
44. OECD (2013). *PISA 2015 draft collaborative problem solving framework*. Pobrano z <http://www.oecd.org/pisa/pisaproducts/Draft%20PISA%202015%20Collaborative%20Problem%20Solving%20Framework%20.pdf>.
45. Pearl, J. (1988). *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. San Mateo, CA: Morgan Kaufmann.
46. Poddiakov, A. (2012). Reshenie kompleksnykh problem v PISA-2012 i PISA-2015: vzaimodeistvie so slozhnoi real'nost'yu [Complex Problem Solving at PISA 2012 and PISA 2015: Interaction with Complex Reality]. *Obrazovatel'naya Politika*, 6, 34-53.
47. Quellmalz, E. S., Timms, M. J. i Schneider, S. A. (2009). *Assessment of student learning in science simulations and games*. Paper commissioned for the National Research Council Workshop on Gaming and Simulations, Washington, DC.
48. Shute, V. J. (2011). Stealth assessment in computer-based games to support learning. W S. Tobias i J. D. Fletcher (red.), *Computer games and instruction* (s. 503-524). Charlotte, NC: Information Age Publishers.
49. Soller, A. i Stevens, R. (2007). Applications of stochastic analyses for collaborative learning and cognitive assessment. W G. R. Hancock i K. M. Samuelsen (red.), *Advances in latent variable mixture models* (s. 217-253). Charlotte, NC: IAP.
50. Sternberg, R. J. i Williams, W. M. (2002). *Educational psychology*. Boston: Allyn & Bacon.
51. Stevens, R. H., Lopo, A. C. i Wang, P. (1996). Artificial neural networks can distinguish novice and expert strategies during complex problem solving. *Journal of the American Medical Informatics Association*, 3(2), 131-138.
52. von Davier, A. A. i Halpin, P. F. (2013). *Collaborative Problem Solving and the Assessment of Cognitive Skills: Psychometric Considerations*. ETS Research Report Series, 2013(2), 1-36.
53. Voss, J. F., Wolfe, C. R., Lawrence, J. A. i Engle, R. A. (1991). From representation to decision: An analysis of problem solving in international relations. W R. J. Sternberg i P. A. Frensch (red.), *Complex problem solving: Principles and mechanisms* (s. 119-158). Hillsdale, NJ: Lawrence Erlbaum Associates.
54. Wagner, R. K. (1991). Managerial problem solving. W R. J. Sternberg i P. A. Frensch (red.), *Complex problem solving: Principles and mechanisms* (s. 159-184). Hillsdale, NJ: Lawrence Erlbaum Associates.

55. Wenke, D., Frensch, P. A. i Funke, J. (2005). Complex problem solving and intelligence: Empirical relation and causal direction. W R. J. Sternberg i J. E. Pretz (red.), *Cognition and intelligence: Identifying the mechanisms of the mind* (s. 160-187). New York: Cambridge University Press.
56. Williamson, D. M., Mislevy, R. J. i Bejar, I. I. (red.). (2006) *Automated Scoring of complex performances in computer based testing*. Mahwah, NJ: Lawrence Erlbaum Associates.
57. Woolley, A. W., Chabris, C. F., Pentland, A., Hashmi, N. i Malone, T. W. (2010). Evidence for a collective intelligence factor in the performance of human groups. *Science*, 330(6004), 686-688.