

MICHAŁ DASZKIEWICZ

Zakład Diagnostyki Edukacyjnej, Uniwersytet Gdański

## KIERUNEK I PŁASZCZYZNY ROZWOJU SKOMPUTERYZOWANEGO POMIARU DYDAKTYCZNEGO. OD POSŁUSZNEGO NARZĘDZIA DO INTELIGENTNEGO DORADCY

Tekst ma charakter porządkujący. Przedstawia kierunek rozwoju, jaki skomputeryzowanemu pomiarowi dydaktycznemu (SPD) otworzyła i obecnie wytycza probabilistyczna teoria wyniku zadania (IRT). Pokazuje, w jak odmienny sposób komputer wspiera pomiar dydaktyczny dziś i jak wspierać go będzie w przyszłości. Przybliża zasadniczo obraz tego rozwoju nakreślony w rozprawie Victora Bundersona i współautorów, zamieszczonej w 3. edycji *Educational Measurement* z 1989 r.<sup>1</sup>, zatytułowanej *Cztery generacje skomputeryzowanego pomiaru dydaktycznego*. Godny podkreślenia jest fakt, że napisany ponad 15 lat temu, ten ponad 150-stronicowy tekst zachowuje swoją aktualność, a nawet – przynajmniej u nas w kraju – dopiero dziś staje się rzeczywiście aktualny. Odwołując się zasadniczo do tej właśnie rozprawy oraz do paru innych krótkich tekstów, niniejszy artykuł prezentuje rozwój SPD na pięciu płaszczyznach, dotyczących zakresu realizowanych przez komputer **funkcji** oraz badanych w pomiarze **dziedzin**, podstawowego **celu**, **metody** wykorzystanej przy pracy z komputerem oraz zakresu wykorzystania **wyników**. Dodatkowo, artykuł przywołuje wybrane oryginalne angielskojęzyczne określenia, co służyć ma pełniejszemu zobrazowaniu przyjętego sposobu myślenia oraz rozwojowi dziedziny w naszym kraju.

Cztery generacje, o których mowa w przywołanej wyżej rozprawie, zachowują kumulatywny charakter oraz pozwalają uporządkować terminologię związaną z wykorzystaniem komputera w testowaniu. Są to kolejno:

---

<sup>1</sup> Bunderson V.C., Inouye D.K., Olsen J.B., *The Four Generations of Computerized Educational Measurement*, [w:] Linn R. L. (red.), *Educational Measurement. Third Edition*, American Council of Education, Washington 1989.

Generacja 1. **Testowanie skomputeryzowane (TS)**

Generacja 2. **Skomputeryzowane testowanie adaptacyjne (STA)**

Generacja 3. **Pomiar ciągły (PC)**

Generacja 4. **Pomiar inteligentny (PI)**

Warto w tym miejscu zwrócić uwagę na szczególną rolę, jaką w rozwoju komputeryzacji testowania pełni teoria wyniku zadania (IRT). Stworzywszy możliwość adaptowania testu do poziomu umiejętności osoby, która go rozwiązuje, probabilistyczna teoria IRT, ze względu na założenie jednowymiarowości testu (o którym mowa będzie w dalszej części niniejszej rozprawy), przynależy zasadniczo do 2. generacji. Wykroczenie poza generację STA wymaga teorii bardziej rozbudowanych, o większej złożoności warsztatu psychologicznego oraz aparatu matematycznego, aniżeli posiada IRT w jej dzisiejszym kształcie.

Autorzy rozprawy *Cztery generacje...* obrazują znaczenie zmian w zakresie komputeryzacji testowania poprzez porównanie ich do przemian, które niegdyś dokonały się w zakresie mikrobiologii oraz medycyny. Wprowadzenie bowiem do tych dwóch dziedzin odpowiednio: mikroskopu oraz prześwietleń rentgenowskich pozwoliło obserwować to, co wcześniej pozostawało nieobserwowalne. Autorzy rozprawy twierdzą, że to samo rzecz można o skomputeryzowanym testowaniu ludzkich umiejętności na gruncie edukacji – nowoczesna technologia pozwala uchwycić takie informacje na temat osób uczących się, które wcześniej wymykały się poznaniu osób informację te rozpoznających. Ten przyrost ilości informacji, zawdzięczany zwiększonej „mocy obserwacji”<sup>2</sup>, stanowi w pierwszym miejscu o wartości skomputeryzowanego pomiaru dydaktycznego w kształceniu i o wadze jego rozwoju.

## 1. Funkcje

Rozwój skomputeryzowanego testowania zaznacza się najwyraźniej w zakresie funkcji, które „przejmują na siebie” komputer. Kierunek tego rozwoju obserwujemy głównie właśnie na tej płaszczyźnie. Komputeryzacja testowania 1. generacji ogranicza wykorzystanie komputerów jedynie do przeprowadzenia konwencjonalnych testów o stałej długości i uprzednio ustalonej sekwencji zadań oraz do zliczania punktacji i wyprowadzenia klasycznych wskaźników statystycznych, podczas gdy funkcja komputerów w STA staje się znacznie donioślejsza. Komputer „przygląda się” osobie rozwiązującej zadania i, w zależności od zaobserwowanego poziomu umiejętności, dobiera w oparciu o wyliczone parametry kolejne zadania tak, aby ich poziom trudności był dla danej osoby optymalny, a tym samym dostarczał maksimum informacji, zgodnie ze znanym założeniem, iż zadania nazbyt łatwe oraz nazbyt trudne dostarczają niewiele informacji o oso-

---

<sup>2</sup> Ibidem, s. 134 (Numery stron odnoszą się do wersji tekstu dostępnej drogą elektroniczną pod adresem: [www.eric.ed.gov/ERICDocs/data/ericdocs2/content\\_storage\\_01/0000000b/80/26/52/57.pdf](http://www.eric.ed.gov/ERICDocs/data/ericdocs2/content_storage_01/0000000b/80/26/52/57.pdf)).

bie rozwiązującej zadania, bądź nie dostarczają jej wcale. Dobierając zadania, komputery determinują treść, której opanowanie jest przedmiotem testowania.

Funkcje komputera rozrastają się jeszcze bardziej w kolejnych dwóch generacjach. W przypadku pomiaru ciągłego obejmują – obok zadań typowych dla wcześniejszych generacji – oszacowanie zmian zaistniałych na „trajektorii osiągnięć” ucznia (ang. *student's achievement trajectory*) oraz rozpoznanie – za pomocą zaawansowanego aparatu matematycznego i psychologicznego – jego profilu jako osoby uczącej się. W pomiarze inteligentnym, pozostającym do tej pory wyłącznie wyimaginowanym etapem rozwoju komputeryzacji testowania, rozpoznany profil uczącego się zostaje poddany interpretacji, która stanowi podstawę formułowania rad udzielanych uczącemu się przez komputer, wykorzystujący w tym celu tzw. bazy wiadomości (ang. *knowledge bases*) oraz procedury służące wyciągnięciu wniosków (ang. *inferencing procedures*).

Rozwój komputeryzacji testowania oznacza zatem postęp w zakresie indywidualizacji pomiaru osiągnięć uczniów. Pociąga za sobą również włączenie pomiaru w codzienny proces kształcenia oparty na programie, którego dwoma częściami składowymi są: z jednej strony – doświadczenia służące realizacji określonych edukacyjnych celów, a z drugiej – zbiór wyznaczników osiągnięć (ang. *milestones of accomplishment*)<sup>3</sup>, pozwalających określić stopień realizacji celów przyjętych na początku, w toku oraz na końcu danego cyklu kształcenia. Pomiar przestaje tu stanowić odrębną czynność w stosunku do codziennych czynności wykonywanych w klasie, lecz nabiera charakteru nie-odrębnego, nie-inwazyjnego<sup>4</sup> (ang. *unobtrusiveness* – dosł.: „nienarzucanie się, nienatrętność, nienatarczywość”). Autorzy *Czterech generacji...* stwierdzają, że *stając się nieinwazyjnym, pomiar ciągły czyni ogromny krok do przodu pod względem użyteczności*<sup>5</sup>.

## 2. Dziedziny

Jakościowe zmiany zachodzące pomiędzy kolejnymi generacjami testowania komputerowego dotyczą również płaszczyzny zmiennych stanowiących przedmiot zainteresowań i badań. I jak w przypadku dwóch wcześniejszych generacji płaszczyzny te należą zasadniczo do dziedziny poznawczej, tak w zakresie pomiaru ciągłego oraz inteligentnego obraz ucznia tworzony jest w oparciu o wybrane właściwości związane ze sferą emocjonalną. Ponadto informacja pozyskiwana za pośrednictwem komputera w pierwszych dwóch generacjach ma charakter wyraźnie analityczny i dotyczy pojedynczych czynności ucznia, natomiast w obrębie dwóch kolejnych generacji zdradza zdecydowanie holistyczny charakter i – sprawdzając bardziej złożone czynności oraz skomplikowane umiejętności – tworzy całościowy obraz uczniowskiego profilu jednostki.

<sup>3</sup> Ibidem, s. 77.

<sup>4</sup> Ibidem, s. 79.

<sup>5</sup> Ibidem, s. 93.

We wcześniejszych generacjach (TS, STA) komputer dostarcza informacji na temat poziomu umiejętności ucznia, które traktuje i przedstawia jednowymiarowo. I choć procedury stosowane w ramach generacji STA pozwalają pominąć w testowaniu zadania, które ze względu na swój poziom trudności nie są odpowiednie dla osób je rozwiązujących, to dopiero 3. generacja (PC) rzeczywiście uwzględni różnice pomiędzy uczniami. Posługując się złożonymi i wielowymiarowymi skalami<sup>6</sup>, oszacowuje zmiany w zakresie osiągnięć, reaguje na określone preferencje jednostki<sup>7</sup>, a także rozpoznaje strategie rozwiązywania problemów i sposoby radzenia sobie z porażką na wstępnym etapie pracy nad zadaniem testowym<sup>8</sup>.

Za istotny można uznać fakt, iż zaawansowane skomputeryzowane testowanie, które autorzy rozprawy *Cztery generacje...* przyporządkowują do 3. i 4. generacji, służy rozpoznaniu „trajektorii wzrostu” (ang. *growth trajectory*), w czym wyraźnie „celuje wyżej” od 1. i 2. generacji, nie odnosząc się do tradycyjnego, jednowymiarowego przyrostu (ang. *increase*). Kluczową rolę pełni w tym procesie wykorzystanie tzw. map postępu (ang. *mastery map*), czyli zbioru właściwości charakteryzujących osoby uznane w danej dziedzinie za biegłe znające określony przedmiot (ang. *increments in expertise*). Podkreśla się tu zindywidualizowany charakter tak rozumianej biegłości, którą nabywa się *po długim okresie wytrwałości i zaangażowania*<sup>9</sup>, a której ocena musi oprzeć się na „wytworach niepowtarzalnych”, np. na rozwiązaniu złożonych problemów, ustnych prezentacjach, pisemnych analizach bądź teczkach ucznia<sup>10</sup>.

### 3. Cele

Rozwój testowania komputerowego oznacza również zmiany w zakresie celów realizowanych przez pomiar. W ujęciu klasycznym, tradycyjnym, służy on bowiem w głównej mierze instytucjom, które jego wyniki wykorzystują przy podejmowaniu szeregu mniej lub bardziej ważkich decyzji. Realizacja tych instytucjonalnych celów to domena zasadniczo dwóch pierwszych generacji. Komputerowe testowanie przynależne z kolei do późniejszych generacji służy głównie celom indywidualnym. Nie ograniczają się one do zdobycia za pośrednictwem komputera ilościowej bądź jakościowej informacji na temat postępu w nauce, lecz obejmują również inteligentną interpretację wyniku oraz sposobu rozwiązywania zadań, a także poradę dotyczącą treści i metod dalszego uczenia się (ang. *prescriptive advice*).

Indywidualizacja skomputeryzowanego testowania zaznacza się już w zakresie pierwszej generacji TS, gdyż samo zastosowanie komputera przez jednost-

---

<sup>6</sup> Ibidem, s. 80.

<sup>7</sup> Ibidem, s. 106, 109.

<sup>8</sup> Ibidem, s. 108.

<sup>9</sup> Ibidem, s. 96.

<sup>10</sup> Ibidem, s. 98.

kę można uznać za pracę samodzielną, bez udziału grupy, co otwiera drogę do tworzenia testów służących do celów prywatnych, jednostkowych, a tym samym niestandardowych (ang. *customized tests*). W zakresie drugiej generacji, STA, charakter indywidualny ma również sama konstrukcja testu, gdyż dobór i sekwencja zadań zależą od poziomu umiejętności jednostki. Najwyraźniej jednak indywidualizacja skomputeryzowanego testowania zaznacza się w dwóch kolejnych generacjach, czyli w pomiarze ciągłym oraz pomiarze inteligentnym. Obejmuje ona szereg usług, jakie jednostce świadczyć może testowanie skomputeryzowane, w tym m.in.: wstępne umiejscowienie na mapie osiągnięć, jednoczesny pomiar umiejętności składowych, prezentacja osobistego profilu ucznia, uprzedzenie o możliwym traceniu nabytej wiedzy lub umiejętności<sup>11</sup>. Przedstawiając tak dalece posuniętą indywidualizację jako etap, w kierunku którego rozwija się skomputeryzowane testowanie, autorzy przywoływanej tu rozprawy zwracają uwagę na fakt, iż *[takie] programy nauczania, [które] podejmują problem różnic osobowościowych, nie odnajdują obecnie swojego miejsca w rozkładzie zajęć [współczesnej] szkoły*<sup>12</sup>.

#### 4. Metody

Stopień zaawansowania warsztatu psychologiczno-matematycznego wyznacza kolejną płaszczyznę, na której dwie wcześniejsze generacje odróżniają się od dwóch późniejszych. Te pierwsze bowiem realizują pomiar statyczny, określający pozycję osoby, sytuacji bądź zdarzenia w danym momencie procesu kształcenia. Drugie zaś przeprowadzają pomiar dynamiczny, ujmujący zmiany te same pozycje, jakie następują z upływem czasu. Pierwsze posiłkują się prostszym aparatem matematycznym, wyniki przedstawiają na jednowymiarowych skalach (porządkowych, nominalnych oraz przedziałowych)<sup>13</sup>, wykorzystują niewielki zbiór wskaźników, w tym przede wszystkim te, które określają poziom trudności, mocy różnicującej oraz poziom umiejętności. Późniejsze generacje natomiast odwołują się do bardziej zaawansowanego aparatu psychologicznego oraz matematycznego, pozwalających scharakteryzować zmiany obserwowane na określonej „trajektorii wzrostu”, a wyniki przedstawiają na złożonych skalach, opierając się na rozwiniętym zbiorze miar adekwatnych do wielowymiarowych właściwości poddanych testowaniu.

Postępującą złożoność procedur łatwo zauważyć w ramach kalibrowania testu, które w testowaniu komputerowym zajmuje miejsce szczególne i stanowi o możliwości jego zindywidualizowania. Przeprowadzane w ramach drugiej generacji, STA, kalibrowanie zadań, którego celem jest pozyskanie ramy odniesienia dla uzyskiwanych w zadaniach wyników, sprowadza się do określenia we wspólnej metryce wartości parametrów zadania oraz wartości *theta* określających

<sup>11</sup> Ibidem, s. 105.

<sup>12</sup> Ibidem, s. 107.

<sup>13</sup> Ibidem, s. 24.

poziom (leżącej u podstaw) cechy latentnej (ang. *underlying latent trait*)<sup>14</sup>. Metodą często wykorzystywaną w tym celu jest technika zaproponowana przez Alana Birnbauma w 1968 r., oparta na dwóch etapach oszacowania maksymalnego prawdopodobieństwa (ang. *maximum likelihood estimation*).

Kalibrowanie przeprowadzane w ramach trzeciej generacji PC określa się natomiast w wyrażnie mniej jednolity sposób, traktując je jako proces *określenia wskaźników, które znaczą postęp na kontinuum od nowicjusza do eksperta*<sup>15</sup>. Kalibrowanie dotyczy tu tzw. zadań odwoławczych (ang. *reference tasks*), na ogół bardziej złożonych od pojedynczych zadań testowych, wzbogaconych w kontekst i wbudowanych (ang. *embedded*) w program nauczania. Terminy stosowane w ramach trzeciej generacji do opisu postępu przyjmują inne znaczenie, w zależności od tego, czy kalibrowane są zadania z zakresu zdolności motorycznych czy wiadomości akademickich, wstępnego przyswajania zdolności akademickich przez dzieci czy przyswajania nowej wiedzy przez profesjonalistów, nauki przedmiotów ścisłych czy języka obcego<sup>16</sup>.

## 5. Wyniki

W ramach czterech generacji inny jest – jak widać z powyższych rozważań – zakres informacji uzyskiwanych za pośrednictwem komputera. W przypadku pierwszej generacji zachowują one liczbowy charakter, a podstawowa korzyść sprowadza się do ich sprawnego pozyskania, przetworzenia, przekazania do odpowiedniego ośrodka oraz utrwalenia na potrzeby przyszłych zastosowań i użytkowników. Generacja STA, która włącza w testowanie tylko i wyłącznie zadania o odpowiedniej wartości informacyjnej<sup>17</sup>, uzyskuje informacje bardziej adekwatne w stosunku do poziomu umiejętności danej grupy uczniów, a tym samym bardziej szczegółowe. Dzięki zwiększonej wydajności możliwe jest bądź ograniczenie długości testu przy zachowaniu tej samej dokładności co w przypadku testowania konwencjonalnego (ang. *paper-and-paper test*), bądź uzyskanie większej precyzji pomiaru przy zachowaniu długości testu równej długości testu konwencjonalnego, bądź też poszerzenie treści merytorycznej objętej testem, przy pozostawieniu długości oraz precyzji zbliżonych do konwencjonalnych testów<sup>18</sup>.

Najbardziej wyraźny wzrost uzyskiwanych informacji ma miejsce w dwóch kolejnych generacjach, zaś przetworzenie tych danych wykracza poza analizę wyników punktowych i obejmuje umiejscowienie danego ucznia na dynamicznej mapie osiągnięć, rozpoznanie jego preferencji oraz stylu uczenia się, a w przy-

<sup>14</sup> Baker F., *The Basics of Item Response Theory*, ERIC Clearinghouse on Assessment and Evaluation, University of Maryland, College Park, MD 2001, s. 130.

<sup>15</sup> Bunderson V.C., Inouye D.K., Olsen J.B., op. cit., s. 101.

<sup>16</sup> Ibidem, s. 89.

<sup>17</sup> Reckase M.D., *Computerized adaptive testing: a good idea for the right technology*, dokument ERIC nr 294 901, 1988.

<sup>18</sup> McBride J.R., *Computerized Adaptive Testing*, [w:] „Educational Leadership”, October 1985, s. 26.

padku generacji pomiaru inteligentnego – także wysunięcie propozycji dotyczących tego, jakie dalsze kroki w uczeniu się należy podjąć, by osiągnąć wyższy poziom biegłości w określonej dziedzinie. Wskazania te dotyczą tak treści przedmiotowych, jak i strategii, które do ich opanowania wydają się – w świetle zaobserwowanego profilu ucznia – najefektywniejsze.

Najważniejsze płaszczyzny, na których zaznaczają się jakościowe różnice pomiędzy kolejnymi generacjami skomputeryzowanego pomiaru dydaktycznego, zostały przedstawione w poniższej tabeli:

Tab. 1. Zakres wykorzystania komputera w testowaniu dla czterech generacji

| Sposób wykorzystania komputera w testowaniu             | TS | STA* | PC | PI |
|---|----|------|----|----|
| Zliczanie punktacji                                     | √  | √    | √  | √  |
| Obliczanie klasycznych wskaźników statystycznych        | √  | √    | √  | √  |
| Podjęcie decyzji o charakterze instytucjonalnym         | √  | √    | √  | √  |
| Indywidualna praca ucznia                               | √  | √    | √  | √  |
| Rozpoznanie poziomu umiejętności (latentnej) („theta”)  | -  | √    | √  | √  |
| Uniezależnienie parametrów zadań od grupy kalibracyjnej | -  | √    | √  | √  |
| Dobór zadań o optymalnym poziomie trudności             | -  | √    | √  | √  |
| Zróżnicowanie długości testu                            | -  | √    | √  | √  |
| Ograniczenie czasu testowania                           | -  | √    | √  | √  |
| Oszacowanie zmian w zakresie wielowymiarowych cech      | -  | -    | √  | √  |
| Rozpoznanie profilu osiągnięć                           | -  | -    | √  | √  |
| Zastosowanie holistycznych zadań testowych              | -  | -    | √  | √  |
| Rejestracja trajektorii wzrostu, pomiar dynamiczny      | -  | -    | √  | √  |
| Wykorzystanie map osiągnięć                             | -  | -    | √  | √  |
| Diagnoza metaumiejętności                               | -  | -    | √  | √  |
| Wykorzystanie aparatu psychologicznego                  | -  | -    | √  | √  |
| Wzbogacenie zadań w kontekst                            | -  | -    | √  | √  |
| Określenie poznawczych preferencji ucznia               | -  | -    | √  | √  |
| Interpretacja profilu osiągnięć ucznia                  | -  | -    | -  | √  |
| Wskazanie dalszego kierunku i sposobu nauki             | -  | -    | -  | √  |

\* Wyłuszczonego etapu, na wstępie którego skomputeryzowany pomiar dydaktyczny znajduje się obecnie w naszym kraju.

W tabeli na szczególną uwagę zasługują następujące kwestie:

1. Indywidualizacja nauczania zaznacza się już w ramach pierwszej generacji.
2. Wyróżnikiem drugiej generacji jest możliwość dostosowania poziomu trudności zadań, a także długości czasu oraz testu do poziomu umiejętności ucznia.
3. Trzecią generację wyróżnia wielowymiarowość informacji na temat wiedzy i umiejętności ucznia, pozwalająca określić jego profil osiągnięć oraz umiejscowić go na mapie osiągnięć.
4. Czwarta generacja zapowiada możliwość skomputeryzowanej interpretacji profilu osiągnięć ucznia prowadzącej do formułowania rad dotyczących dalszego uczenia się, ugruntowanych informacją na temat stylu uczenia się oraz preferencji ucznia. Ten ostatni rodzaj danych zawiera istotny pierwiastek emocjonalny!

Czas pokaże, czy predykcje autorów rozprawy okażą się trafne i czy wrażliwość obecność komputera w pomiarze dydaktycznym niesie ze sobą więcej korzyści czy zagrożeń. Wydaje się, że jeżeli wdrażaniu komputera w testowanie osiągnięć ucznia towarzyszyć będą: rozwaga, wrażliwość i świadomość wirtualnych ograniczeń, to istnieją duże szanse, że komputer przysłuży się rozwojowi umiejętności oraz wiedzy ludzkiej.