



DOI: 10.15199/ELE-2014-025

## Wielokrotny wybór czyli rozważania o pewnych pytaniach testowych

mgr **WOJCIECH PRZYŁUSKI**, Instytut Maszyn Matematycznych, Warszawa

Wśród pytań zamkniętych (czyli takich, w których testowany wybiera odpowiedzi spośród pewnej liczby konkretnie sformułowanych propozycji) najczęściej spotykanymi pytaniami – **obok pytań polegających na wskazaniu jednej poprawnej odpowiedzi** – są tzw. pytania wielokrotnego wyboru. Słowo „wielokrotnego” jest przy tym różnie interpretowane przez autorów publikacji związanych z tematyką testowania; może oznaczać zarówno liczbę proponowanych do wyboru opcji, jak i liczbę możliwych prawidłowych odpowiedzi.

Każdy chyba odpowiadał kiedyś na pytania polegające na wyborze właściwej odpowiedzi. Polega to na tym, że dostaje się kilka, rzadziej kilkanaście możliwości do wyboru i należy zaznaczyć jedną lub kilka odpowiedzi uznanych za prawidłowe. Możliwość zaznaczenia więcej niż jednej odpowiedzi zależy od treści polecenia zawartej w pytaniu lub od formy graficznej pytania, która sugeruje bądź wręcz determinuje (testy komputerowe) zasady udzielania odpowiedzi.

Spśród kilku polskich i angielskich nazw proponowanych w literaturze dla pytań wyboru najtrafniejsze wydają się określenia z pracy Marka Hylia [1]. Wymienia on dwa typy pytań wyboru:

1. pytania jednokrotnego wyboru z wielu możliwych odpowiedzi (Multiple Choice Single Answer, MCSA),
2. pytania wielokrotnego wyboru z wielu możliwych odpowiedzi (Multiple Choice Multiple Answer, MCMA).

Dopiero te dwie polskie i odpowiadające im angielskie nazwy w pełni oddają sens jaki chcemy przypisywać tym typom pytań.

W pierwszej chwili różnica pomiędzy tymi typami pytań wydaje się niewielka. Okazuje się jednak, że wielość możliwych do wskazania odpowiedzi stwarza w sferze weryfikacji wiedzy zupełnie nowe, nieporównywalne z pytaniami jednokrotnego wyboru, możliwości, rodząc jednocześnie wiele dodatkowych problemów. Tymi właśnie problemami chcemy się zająć, przy czym używać będziemy sformułowań: „pytanie wielokrotnego wyboru” i „wielokrotny wybór” w sensie drugiej z podanych wyżej definicji.

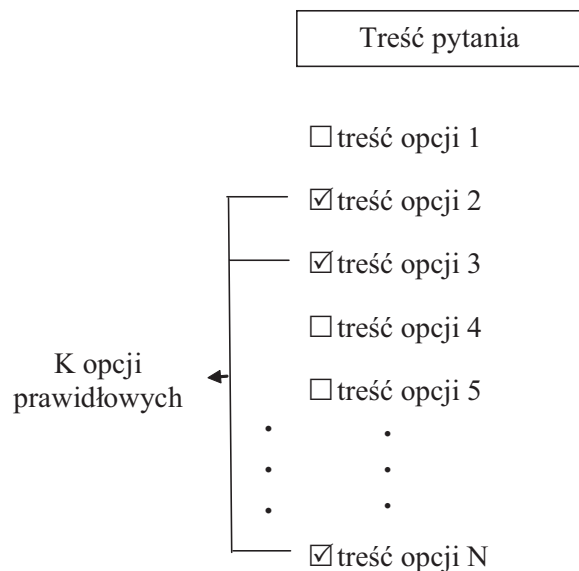
Test tradycyjny lub komputerowy to swoista gra pomiędzy twórcą testu a testowanym. Ten pierwszy formułuje zasady tej gry, a drugi powinien się do nich stosować. Zasady, o których tu mowa obejmują kilka zagadnień. Należy do nich przede wszystkim wyjaśnienie celu przeprowadzania danego testu, ale tę sprawę pozostawimy na boku przyjmując jedynie założenie, że test jest oceniany i że wynik testu jest liczbą. Jednym z najważniejszych zagadnień jest przekazany testowanemu zakres informacji dotyczących sposobu weryfikacji wiedzy, który zastosował twórca testu. O ile w klasycznych testach „kartkowych” ich twórcy dążą do maksymalnej prostoty systemu oceniania albowiem ułatwia to później sprawdzanie

i wystawianie ocen, o tyle w testach komputerowych stopień komplikacji systemu oceniania nie ma znaczenia z punktu widzenia czasochłonności procesu wystawiania ocen. Z tym jednak zastrzeżeniem, że twórca testu komplikując system ocen musi zdawać sobie sprawę, że nie w każdym środowisku komputerowym będzie mógł ten system oceniania zrealizować.

Informacja dotycząca systemu oceniania ma więc dla testowanego szczególne znaczenie, bowiem właśnie na jej podstawie testowany będzie realizował swoją strategię w grze, którą jest zdawanie testu. Oczywiście należy również wyjaśnić testowanemu sposób wybierania odpowiedzi, ewentualnego ich modyfikowania oraz zatwierdzania dokonanego wyboru, czyli wszystkie sprawy „techniczne” związane z testowaniem.

### Pytanie wielokrotnego wyboru

W tym artykule analizujemy wyłącznie te aspekty testowania wiedzy za pomocą pytań wielokrotnego wyboru, które dotyczą problemu ich oceniania przez twórcę testu. Cała sfera zagadnień związanych z dydaktyczną stroną testu jest tutaj celowo pominięta. W dalszych rozważaniach pomoże nam schemat typowego pytania wielokrotnego wyboru:



Rys. 1. Schemat pytania wielokrotnego wyboru  
Fig. 1. Pattern of multiple choice question

gdzie N to liczba wszystkich możliwych odpowiedzi (opcji), a K jest liczbą odpowiedzi (opcji) uznanych za prawidłowe. Przyjmijmy  $N > 0$  i  $K > 0$ .



Przyjmujemy  $K > 0$ , bowiem gdyby dopuścić brak opcji prawidłowych ( $K = 0$ ), to egzaminator nie może odróżnić sytuacji kiedy testowany celowo nie zaznaczył żadnej opcji od sytuacji rezygnacji z odpowiedzi na dane pytanie.

Testowany odpowiadając na pytanie wielokrotnego wyboru może zaznaczyć dowolny podzbiór opcji (w tym pusty, co oznacza, że nie wybrał żadnej z opcji). Każdy taki wybór nazywamy dalej układem odpowiedzi (albo krócej – układem). Poszczególnym układom można przyporządkowywać liczbową ocenę; po wyborze przez testowanego jakiegoś układu jego liczbowa ocena staje się oceną pytania.

Przedstawimy teraz kilka wybranych problemów, które powinien uświadamiać sobie twórca testu wykorzystujący pytania wielokrotnego wyboru. Problemy te dotyczą dwóch sfer: systemu oceniania oraz strategii zdawania testu, którą stosują osoby egzaminowane.

## System oceniania

### Założenie dotyczące systemu oceniania

Analizujemy wyłącznie ocenę w zakresie jednego testowego pytania. Zakładamy, że za każde pytanie testu można otrzymać tę ocenę z ustalonego wspólnego zakresu. Przyjmujemy też, że oceny za poszczególne pytania są kumulowane. Nie określamy zasad tej kumulacji. Czytelnik może jednak dla ułatwienia przyjąć, że oceny za poszczególne pytania są sumowane, a globalną oceną testu jest stosunek sumy ocen zdobytych do sumy ocen możliwej do zdobycia.

Istotnym jest, że nie zakładamy, że poszczególne opcje pytania wielokrotnego wyboru mają jednakowe merytoryczne znaczenie (wagi). Oceniany jest każdy układ odpowiedzi i na tę ocenę wagi poszczególnych opcji mogą mieć wpływ. Warto podkreślić, że ponieważ dopuszczamy w naszych rozważaniach zupełnie dowolne systemy oceniania, to pytanie wielokrotnego wyboru z  $N$  opcjami nie jest w sensie oceny równoważne wiązce  $N$  pytań jednokrotnego wyboru z pojedynczą opcją. Bowiem, podkreśliśmy to raz jeszcze, oceniamy układy opcji, a ocena ta nie musi być prostą kumulacją ocen dla poszczególnych opcji z rozważanego układu (patrz Przykład 1).

### Dwuwartościowa ocena (DO)

Kluczowym problemem twórcy testu jest to, jaki należy przyjąć system oceniania pytań wielokrotnego wyboru. Najprostsze rozwiązanie to przyjęcie dla takich pytań dwuwartościowej oceny. Będzie to liczba  $A$  w przypadku zaznaczenia wyłącznie wszystkich prawidłowych  $K$  opcji oraz liczba  $B$  w każdym innym przypadku. Zbiór wszystkich możliwych ocen jest postaci  $\{A, B\}$ .

Jeśli twórca testu nie zdecydował się na takie proste rozwiązanie, mamy do czynienia z przypadkiem analizowania układów odpowiedzi częściowo prawidłowych bądź nieprawidłowych.

### Odpowiedzi częściowo prawidłowe bez żadnej nieprawidłowej opcji (CO)

W tym przypadku twórca testu chce oceniać wszystkie niepuste układy odpowiedzi, w których nie zaznaczono żadnej opcji nieprawidłowej. Każdemu takiemu  $i$ -temu układowi odpowiada ocena  $A_i$ . Zbiór wszystkich możliwych ocen jest teraz postaci  $\{A_1, \dots, A_P, B\}$ , gdzie  $P = 2^K - 1$  odpowiada liczbie tych układów.

W tym przypadku nabierają znaczenia wagi poszczególnych (uznanych za prawidłowe) opcji. Na każdą ocenę  $A_i$  mogą bowiem mieć wpływ wagi opcji należących do  $i$ -tego układu.

### Odpowiedzi częściowo poprawne i częściowo niepoprawne (TO) (totalna ocena)

Ten ostatni przypadek obejmuje wszystkie systemy oceniania. Każdy układ odpowiedzi może zostać indywidualnie oceniony. Każdemu  $i$ -temu układowi odpowiada ocena  $A_i$ . Zbiór wszystkich możliwych ocen jest teraz postaci  $\{A_1, \dots, A_W\}$ , gdzie  $W = 2^N$  odpowiada liczbie tych układów.

W tym przypadku również nabierają znaczenia wagi wszystkich poszczególnych opcji. Na każdą ocenę  $A_i$  mogą bowiem mieć wpływ wagi opcji należących do  $i$ -tego układu.

### Problem liczby K

Oczywiście liczba naturalna  $K$  zawiera się w przedziale domkniętym  $[1, N]$ . Można założyć, że niezależnie od tego czy analizujemy testy klasyczne czy komputerowe, liczba  $N$  zawiera się pomiędzy 3 a 10. Oczywiście zdarzają się testy z pytaniami wielokrotnego wyboru zawierającymi kilkanaście opcji, ale zdarza się to niezwykle rzadko.

Przypomnijmy, że liczba wszystkich możliwych wyborów wśród wszystkich poprawnych opcji, których dokonać może testowany przy jednoczesnym założeniu, że nie zaznaczył żadnej błędnej opcji wynosi:  $2^K - 1$ .

Jeśli twórca testu decyduje się na system oceniania DO, wtedy wielkość liczby  $K$  nie ma żadnego praktycznego znaczenia.

W przypadku systemów CO i TO sytuacja jest zupełnie inna. Jeśli na przykład w systemie CO liczba  $K=6$ , wtedy oceniać należy  $2^6 - 1$  układów odpowiedzi czyli 63 układy. Pamiętając o tym, że na ocenę każdego układu mogą wpływać wagi opcji do niego należących może się okazać, że w przypadku testów kartkowych wyliczenie ocen dla takiego pytania jest prawie niewykonalne. Natomiast w przypadku testu komputerowego wykonalność tego zadania zależy od użytych narzędzi informatycznych.

W przypadku systemu TO sytuacja wygląda podobnie, bowiem wzrasta jeszcze liczba układów, które należy oceniać.

Reasumując, jeśli wybieramy bardziej rozbudowany system oceniania warto pomyśleć o zmniejszeniu parametrów  $N$  i  $K$  (na przykład do wartości 5 i 3) oraz o maksymalnym uproszczeniu systemu wag dla opcji pytań.

## Strategia zdawania testu

### Rozważania przy obieraniu najlepszej strategii

Wyobraźmy sobie następującą sytuację. Testowany przystępuje do egzaminu, w którym napotyka pytania wielokrotnego wyboru, natomiast nie wie nic na temat systemu oceniania tych pytań. Oto hipotetyczna lista zapytań, które może on sobie stawiać napotykając na takie pytanie, a następnie hipotetyczna lista rozważań i spekulacji wywołanych tymi zapytaniami.

1. Ile jest opcji prawidłowych (żadna- rozważa również taką możliwość, ponieważ nasze założenie, że  $K$  jest większe od zera przyjęliśmy arbitralnie – jedna, kilka, wszystkie)?
2. Czy muszę wskazać wszystkie opcje prawidłowe?



3. Nie jestem pewien co do żadnej z opcji, czy mam ryzykować wskazując wątpliwe odpowiedzi?
4. Wskazałem opcje (których jestem pewien) ale mam wątpliwości co do innych. Czy zaryzykować ich wskazanie?

**Ad 1.** Może to pytanie jest pytaniem z jedną prawidłową opcją ale ma formę pytania wielokrotnego wyboru aby mnie zmylić. W teście były już pytania jednokrotnego wyboru więc to pewnie jest inne. To chyba niemożliwe by zadawano pytanie wielokrotnego wyboru bez żadnej prawidłowej opcji albo ze wszystkimi prawidłowymi opcjami. Pewnie tylko kilka wybranych opcji jest prawidłowych.

**Ad 2 i 4.** Nie wiem czy pozytywnie są oceniane odpowiedzi tylko częściowo prawidłowe. Jeśli nie, to wskażę tylko te których jestem pewien w 100 procentach. Jeśli tak, to chyba powinienem również wskazać te, których jestem pewien więcej niż na 50 procent.

**Ad 3.** Jeśli każde złe wskazanie oznacza ocenę jak w przypadku braku odpowiedzi, to warto spróbować. Jeśli jednak złe wskazanie oznacza ocenę gorszą niż w przypadku braku odpowiedzi, to może nie warto próbować i lepiej nic nie zaznaczać.

### O czym informować testowanego

Warto zauważyć, że powyżej opisane rozważania testowanego sprowadzają się w zasadzie do dwóch pytań:

1. Z jakim systemem oceny (DO, czy innym) mam do czynienia?
2. Jeśli innym, to jakie przyjęto w nim zasady punktacji?

Wydaje się, że egzaminowany powinien znać odpowiedź na pierwsze z wyżej wymienionych pytań. Szczegółowe dane dotyczące testu (np. liczba K, informacje o wagach opcji pytań) nie muszą być mu znane.

Sytuacja komplikuje się w przypadku systemu TO. Można sądzić, że jeśli twórca testu zdecydował się na system TO, to chodzi mu o test różnicujący (a nie sprawdzający) (w teście różnicującym układem odniesienia wyniku każdego testowanego są wyniki innych uczestników testu. W teście sprawdzającym układem odniesienia wyniku każdego testowanego są wymagania programowe [2]). Szczególnego znaczenia nabierają tu pytania podchwytliwe i drobniawo oceniane. Zdradzanie testowanemu wszystkich zasad punktacji jest wtedy równoznaczne z ujawnieniem istnienia w teście takich ewentualnych „pułapek” i „niespodzianek” przygotowanych przez egzaminatora. Z drugiej strony, informacja o istniejących pułapkach wzmaga jedynie czujność testowanego ale przed wpadnięciem w nie nie musi go uchronić.

Przykładami takich pułapek mogą być pytania z opcjami, których wybór w jakiegokolwiek konfiguracji powoduje przyznanie minimalnej oceny.

### Przykłady pytań wielokrotnego wyboru

**Przykład 1.** Wskaż autorów epoki Pozytywizmu:

- A. Bolesław Prus
- B. Aleksander Głowacki
- C. Henryk Sienkiewicz
- D. Zygmunt Krasiński
- E. Adam Mickiewicz
- F. Eliza Orzeszkowa

Opcje: A, B, C oraz F uznajemy za prawidłowe.

Oczywistym jest, że układy odpowiedzi, w których występuje tylko jedna z opcji A i B mogą być dodatkowo niżej ocenione niż wynikałoby ze zwykłej kumulacji ocen za opcje wchodzące w skład układu. Egzaminator może bowiem chcieć uwzględnić przy wystawieniu oceny fakt, że testowany nie wie, że w tych dwóch opcjach jest mowa o tej samej osobie.

Przyjmijmy szkolną skalę ocen {1, 2, 3, 4, 5, 6} – z możliwymi „połówkami” – oraz system oceniania typu CO. Możemy zaproponować egzaminatorowi następujący algorytm oceniania układów odpowiedzi:

Przyznajemy 1,5 punktu za każdą prawidłowo wskazaną opcję przy czym odejmujemy od sumy uzyskanych punktów 0,5 jeśli opcje A i B nie występują łącznie.

Otrzymamy wtedy następujące oceny:

Tab. 1. Tabela ocen (w systemie CO) dla Przykładu 1

Tabl. 1. Valuation table (CO system) for the Example #1

Układ odpowiedzi	Ocena
ABCF	6
ABC	4,5
ABF	4,5
ACF	4
BCF	4
AB	3
AC	2,5
AF	2,5
BC	2,5
BF	2,5
CF	3
A	1
B	1
C	1,5
F	1,5
Inne	1

Można oczywiście dyskutować nad trafnością tego algorytmu, jednak jest on bardzo prosty oraz spełnia postulaty egzaminatora.

Oczywiście w ramach przyjętej szkolnej skali ocen nie wykorzystaliśmy tu jeszcze ocen: 2, 3,5, 5, 5,5 co łatwo umożliwiałoby dalsze zróżnicowanie ocen tych układów, których jednako oceną nas nie zadowalała. Dotyczy to np. układów: AB oraz CF.

**Przykład 2.** Wskaż ssaki:

- A. pies
- B. kot
- C. krowa
- D. bocian

Opcje: A, B, C uznajemy za prawidłowe.

Zastosujemy tym razem skalę ocen {-2, -1, 0, 1, 2, 3, 4, 5, 6}. W przypadku systemów DO lub CO nie napotykamy problemów. Nasuwają się następujące oceny:



Tab. 2. Tabela ocen (w systemie DO) dla Przykładu 2  
Tabl. 2. Valuation table (DO system) for the Example #2

System DO

Układ odpowiedzi	Ocena
ABC	6
Inne	1

Tab. 3. Tabela ocen (w systemie CO) dla Przykładu 2  
Tabl. 3. Valuation table (CO system) for the Example #2

System CO

Układ odpowiedzi	Ocena
ABC	6
AB	4
BC	4
AC	4
A	2
B	2
C	2
Inne	1

W przypadku systemu TO przypisanie ocen każdej z 16 możliwości nie jest już takie proste.

Jeśli zastosujemy przykładowe przeliczenie: dobra opcja: 2 punkty, zła: -2 punkty otrzymamy:

Tab. 4. Tabela ocen (w systemie TO) dla Przykładu 2  
Tabl. 4. Valuation table (TO system) for the Example #2

System TO

Układ odpowiedzi	Ocena
ABC	6
AB	4
BC	4
AC	4
A	2
B	2
C	2
D	-2
ABCD	4
ABD	2
BCD	2
ACD	2
AD	0
BD	0
CD	0
Inne (nic nie wybrane)	0

Rozważmy teraz inną, trudniejszą wersję tego pytania.

Wskaż ssaki:

- A. pies
- B. nietoperz
- C. ichneumon
- D. kulik

Opcje: A, B, C uznajemy za prawidłowe.

Teraz trudność przypisania ocen poszczególnym odpowiedziom (w systemach CO i TO) potęguje fakt, że wagi (ze względu na stopień trudności) poszczególnych opcji, nie są jednakowe. O ile w systemie CO można sobie jeszcze jakoś poradzić, to w systemie TO jest to prawdziwe wyzwanie.

### Jak wybierać system oceniania

W pytaniach wielokrotnego wyboru kryje się duży potencjał. Można oczywiście korzystać z nich w sposób ograniczony, ale można również za ich pomocą prowadzić swoistą grę edukacyjną z testowanym. Ma to sens w przypadku niestandardowych testów, których celem jest bardzo rzetelna, zindywidualizowana ocena testowanego.

Analizując system oceniania w klasycznych testach kartkowych nie można pominąć faktu, że musiał on umożliwić nauczycielowi szybkie i łatwe wyliczenie oceny dla każdej z osób.

Dlatego też najczęściej spotykane testy wielokrotnego wyboru są testami typu DO i CO. W chwili pojawienia się testów elektronicznych podstawowy argument łatwości i szybkości sprawdzania prac uczniów staje się nieaktualny. W teście elektronicznym komplikacja systemu oceniania nie wpływa w zauważalny sposób ani na łatwość, ani na szybkość otrzymania wyników. Tak więc w testach elektronicznych komplikacja systemu oceniania ma znaczenie tylko dla testowanego, a ściślej tylko dla jego strategii rozwiązywania testu.

Okazuje się jednak, że w ogromnej większości testów komputerowych są one nadal typu DO lub CO. Wydaje się, że przede wszystkim jest to wygodne dla twórców testów. Mogą oni łatwo adaptować istniejące już testy kartkowe i stworzyć ich elektroniczne odpowiedniki.

Stosowanie pytań wielokrotnego wyboru w wersji TO jest dość trudne i odpowiada w pewnej mierze egzaminowaniu ustnemu. Tyle tylko, że podczas egzaminu ustnego egzaminator trochę „na wycucie” ocenia takie pytanie, natomiast w teście komputerowym trzeba z góry precyzyjnie określić ocenę dla każdego układu odpowiedzi. Opracowanie testu typu TO wymaga na ogół niełatwego i czasochłonnego zaprojektowania systemu oceniania, a systemy komputerowe realizujące test elektroniczny muszą być przystosowane do takiego typu testów.

### TeleEdu

Jednym z niewielu systemów realizujących omawiany typ testów elektronicznych i uwzględniających przedstawione wyżej techniki oceniania jest system TeleEdu™.

W edytorze TeleEdu™, który m.in. służy do konstruowania testów elektronicznych, możemy sformułować pytanie wielokrotnego wyboru, a następnie przypisać mu zestaw ocen wraz z ewentualnymi komunikatami powiązаныmi z konkretnymi układami odpowiedzi.

Dla omawianego wyżej Przykładu 2 z systemem CO opracowywane w tym edytorze pytanie wygląda w sposób następujący.



Rys. 2. Pytanie z Przykładu 2 (system CO) w oknie edytora TeleEdu™  
Fig. 2. Question from the Example #2 (CO system) presented in TeleEdu™ Editor window

W górnej części okna edytora znajduje się pytanie z zaznaczonym układem odpowiedzi, a poniżej tzw. Panel Ewaluacyjny czyli lista rozważanych przez egzaminatora układów odpowiedzi. Wyróżnionej odpowiedzi (Odpowiedź ABC) przypisana jest ocena o wartości 6 (Aspect1=6). Pod przyciskiem Message ukryty może być komunikat dla testowanego, wyświetlany gdy testowany wybierze tę właśnie odpowiedź. Takie rozwiązanie pozwala w pełni realizować ocenianie w systemach DO, CO i TO.

Rozwiązanie to ma jednak pewien minus; opisać w ten sposób możemy tylko „rozsadną” liczbę układów. Tworząc edytor można było zrezygnować z wizualnej (układowej) edycji pytań na przykład na rzecz definiowania algorytmicznego. Wybierałoby się wtedy pojedynczo prawidłowe opcje i przypisywało im punkty, po czym zliczana byłaby ocena wg wybranej zasady kumulacji (sumowanie, średnia itp.). W takim wypadku wielkość wymienionych wcześniej liczb K i N nie miałyby znaczenia. Nieuniknione modyfikacje takich algorytmów byłyby jednak bardzo kłopotliwe. Sam proces algorytmizacji pewnych aspektów oceniania jest zresztą bardzo trudny, a czasem wręcz niemożliwy.

Zaletą przyjętego rozwiązania jest m.in. łatwość oceniania układów, w których mają znaczenie powiązania między opcjami (patrz Przykład 2). Można po prostu tak jak w egzaminie ustnym, przypisywać oceny intuicyjnie każdemu konkretnemu układowi.

Więcej informacji o systemie TeleEdu™, który umożliwia stosowanie bardzo złożonych strategii testowania, znaleźć można w artykule [3].

## Podsumowanie

Wydaje się, że w sytuacji gdy nauczyciel nie chce bądź nie może poświęcić więcej czasu na przygotowanie testów, powinien ograniczyć się tylko do pytań jednokrotnego wyboru. Unika tym samym wielu opisanych wcześniej problemów. Egzaminowany natomiast nie musi prowadzić żadnej (często stresującej) gry podczas zdawania testu.

Pytania wielokrotnego wyboru warto zatem stosować raczej w różnicujących testach autorskich, pieczołowicie przygotowywanych, których celem jest precyzyjne ustalenie poziomu wiedzy testowanego. Egzaminowany może być wtedy bardziej zestresowany, ale większa waga i rzetelność takiego testu są dla niego pewną rekompensatą.

## Literatura

- [1] Hyla M.: Szkolenia. Przewodnik po e-learningu. Oficyna Ekonomiczna, Kraków 2005.
- [2] Niemierko B.: Pomiar wyników kształcenia. Wydawnictwa Szkolne i Pedagogiczne Spółka Akcyjna, Warszawa 1999.
- [3] Abramowicz A., Przyłuski W.: Złożone strategie testowania w środowisku e-learningowym TeleEdu™. Elektronika nr 2/2011, ss. 145–150.